

# Storia della probabilità



2023

*Riccardo Rosso*  
*Università di Pavia*

# Indice

<b>1</b>	<b>Primi passi della probabilità</b>	<b>3</b>
1.1	Probabilità nel mondo greco . . . . .	3
1.2	Spunti dalla letteratura . . . . .	6
1.3	Il lancio dei dadi nel “General Trattato” di Niccolò Fontana . . .	11
1.4	Galileo Galilei e la probabilità . . . . .	14
1.5	<i>Punctaturae e cadentiae</i> . . . . .	16
1.6	Il <i>Liber de ludo aleae</i> di Gerolamo Cardano . . . . .	20
<b>2</b>	<b>Il problema della divisione della posta</b>	<b>31</b>
2.1	Le origini: Manoscritti . . . . .	31
2.2	Origini: testi a stampa . . . . .	34
2.3	Cardano e la divisione della posta . . . . .	39
2.4	Fermat e Pascal . . . . .	42
2.5	Il <i>Traité</i> di Pascal . . . . .	49
<b>3</b>	<b>Huygens e De Moivre</b>	<b>57</b>
3.1	Il <i>De Ratiociniis in Ludo Aleae</i> di Huygens . . . . .	57
3.2	Probabilità: un nome ambivalente . . . . .	66
3.3	Intermezzo. Le tavole di Graunt. . . . .	67
3.4	La corrispondenza tra i fratelli Huygens . . . . .	72
3.5	Ulteriori sviluppi: le tavole di Breslavia . . . . .	76
<b>4</b>	<b>Nuovi giochi e nuovi metodi di soluzione</b>	<b>85</b>
4.1	Il gioco delle concordanze . . . . .	85
4.2	La durata del gioco . . . . .	98
4.3	De Moivre e la ripartizione della posta . . . . .	109
4.4	Un gioco strategico . . . . .	111
<b>5</b>	<b>L’<i>Ars Conjectandi</i> di Jakob Bernoulli</b>	<b>121</b>
5.1	Analisi dell’opera . . . . .	121
5.2	Le <i>Annotationes</i> di Jakob Bernoulli . . . . .	122
5.3	La Sezione IV dell’ <i>Ars Conjectandi</i> . . . . .	124
5.4	Reazioni al lavoro di Bernoulli . . . . .	141
5.5	Appendice 1: la formula di Stirling . . . . .	146

5.6	Appendice 2: i numeri di Bernoulli . . . . .	148
<b>6</b>	<b>Il paradosso di San Pietroburgo</b>	<b>153</b>
6.1	Le origini . . . . .	153
6.2	La prima critica dei fondamenti del calcolo delle probabilità . . .	171
6.3	L'approccio di Condorcet . . . . .	175
6.4	Gestire il rischio in modo razionale: l'inoculazione del vaiolo . . .	179
<b>7</b>	<b>Le origini della probabilità inversa</b>	<b>191</b>
7.1	Il <i>Saggio</i> di Thomas Bayes . . . . .	191
7.2	La <i>Mathématique sociale</i> di Condorcet . . . . .	213
7.3	Laplace e la probabilità . . . . .	222
7.4	Laplace e la probabilità delle cause . . . . .	225
7.5	Un problema demografico . . . . .	231
7.6	Poisson e la probabilità inversa . . . . .	234
7.7	La regola di successione finita . . . . .	239
<b>8</b>	<b>La teoria degli errori</b>	<b>245</b>
8.1	I primi passi della teoria . . . . .	245
8.2	Primi contributi originali di Laplace . . . . .	257
8.3	La nascita del metodo dei minimi quadrati . . . . .	261
8.4	Adrain e la legge degli errori . . . . .	263
8.5	Gauss e la legge degli errori . . . . .	267
8.6	Il teorema centrale del limite . . . . .	271
8.7	Il <i>Theorema insigne</i> di Gauss . . . . .	275
8.8	Il modello lineare di Gauss . . . . .	279
8.9	Appendice . . . . .	283
<b>9</b>	<b>Sviluppi e crisi nel XIX secolo</b>	<b>289</b>
9.1	La legge dei grandi numeri . . . . .	289
9.2	Čebišev e la legge dei grandi numeri . . . . .	291
9.3	Le critiche al calcolo delle probabilità . . . . .	296
9.4	I Paradossi di Bertrand . . . . .	301
9.5	Le istanze della fisica e l'opera di Poincaré . . . . .	305
<b>10</b>	<b>La critica dei fondamenti della probabilità</b>	<b>315</b>
10.1	Primi tentativi di assiomatizzazione . . . . .	315
10.2	Borel e la probabilità . . . . .	321
10.3	La probabilità secondo von Mises . . . . .	327
10.4	Misurare la probabilità soggettiva . . . . .	334
10.5	L'idea di scambiabilità . . . . .	342
10.6	I <i>Grundbegriffe</i> di Kolmogorov . . . . .	347
10.7	La logica del probabile . . . . .	352
10.8	Appendice . . . . .	355

# Capitolo 1

## Primi passi della probabilità

### 1.1 Probabilità nel mondo greco

In questa sezione vogliamo esaminare la *preistoria* della probabilità occupandoci di una serie di *sintomi* che lasciano intravedere la formazione di concetti probabilistici senza però che vi sia alcun tentativo di formalizzazione matematica. Sull'assenza di indagini sistematiche di natura probabilistica sono state proposte diverse congetture. Per esempio, lo scienziato e storico della scienza Samuel Sambursky (1900-1990) annotò [22]:

Nessuna delle due pietre angolari della scienza moderna, la tecnica sperimentale e la teoria della probabilità si poterono sviluppare nell'antica Grecia perché la mentalità dei greci vedeva la precisione ristretta alle sfere celesti e non riusciva a scorgere delle leggi nelle fluttuazioni ed irregolarità proprie delle cose terrestri.<sup>1</sup>

Il giudizio è un po' estremo e ci sono delle tracce di ragionamenti riconducibili alla probabilità e di pratiche nelle scienze sperimentali che lasciano intravedere la consapevolezza della presenza di varie fonti di errore nelle osservazioni astronomiche. Per esempio, molti passi dell'*Almagesto* di Claudio Tolomeo (ca 100-ca 175) testimoniano l'attenzione a garantire l'affidabilità di una serie di osservazioni, indicando anche alcuni procedimenti da evitare [25]. Tolomeo attribuiva al caso alcuni degli errori da cui le osservazioni potevano essere affette, intendendo per caso ciò che può aver luogo o meno, anche se non sempre distinse nettamente tra errori sistematici ed accidentali.

Non è superfluo osservare che una delle tradizionali aree di applicazione del calcolo delle probabilità, cioè il gioco d'azzardo, era piuttosto problematica perché, ad esempio, gli *astragali*, antenati dei dadi—che avevano una forma allungata e presentavano quattro risultati possibili ad ogni lancio—erano però

---

<sup>1</sup>Neither the technique of experimentation nor the theory of probability, two cornerstones of modern science, could develop in ancient Greece, because Greek mentality saw precision restricted to the heavens and failed to see laws in fluctuations and irregularities on earth.

piuttosto irregolari cosicché non tutte le facce si presentavano, in media, con la stessa frequenza, rendendo difficile trarre delle conclusioni dall'analisi di una serie prolungata di lanci [22, 9].<sup>2</sup>

Un passo di Tucidide (460-385 a.C.), portato all'attenzione della comunità scientifica da Raffaele Invrea (1890-1947)<sup>3</sup> è al contrario di un certo interesse come testimonianza antica—i fatti narrati si riferiscono all'inverno tra il 428 ed il 427 a.C.—di un argomento di natura probabilistica.

In quello stesso inverno, i Plateesi, sempre stretti nell'assedio dei Peloponnesi e dei Beoti, poiché la scarsità di vettovaglie infieriva, mentre la speranza di un soccorso ateniese o di qualche altra via di salvezza diveniva via via più fioca, concepirono di propria iniziativa e di concerto con gli Ateniesi bloccati con loro un primo progetto: un'uscita in massa, per tentare il valico delle mura nemiche e di aprirsi un passaggio con la forza. Fattori del piano erano tra loro Teeneto figlio dell'indovino Tolmide, ed Eupompide figlio di Daimaco, che era anche stratego. Ma in seguito, una metà di loro si lasciò vincere dallo sgomento, giudicando l'impresa troppo arrischiata mentre un gruppo di circa duecentoventi uomini, di libera scelta, confermò il proposito di attuare la sortita, in questo modo. Avevano approntato alcune scale, pari in altezza al vallo nemico, misurato, in base agli strati di mattoni sovrapposti, là dove la superficie a loro rivolta del muro non aveva ricevuto l'intonaco. Erano in molti ad eseguire insieme il computo delle file di mattoni: alcuni potevano imbrogliarsi, ma certo la maggior parte non fallì il calcolo. Si consideri che il conto fu ripetuto più volte; anche la distanza era piuttosto breve e qualsiasi punto del muro risultava perfettamente visibile. Si stabilì con questo accorgimento la misura delle scale, congetturandone l'altezza dallo spessore dei mattoni. (Tucidide, *La guerra del Peloponneso*, Libro III, 20)

A ben vedere, non si parla qui di operazioni di media ma si sottolinea che la maggior parte del manipolo di coraggiosi abitanti di Platea *non fallì il calcolo*, lasciando intravedere come risultato più attendibile la *moda* più che la *media aritmetica* dei risultati delle misure, che sono qui numeri interi. Pur con tutte le riserve del caso, appare abbastanza chiaro che fosse ritenuto necessario un esame dei dati per estrarre il valore più attendibile da una serie di misure non tutte coincidenti. Che una qualche forma di media fosse tenuta in considerazione presso i Greci risulta anche da un altro passo in cui Tucidide, commentando la descrizione della flotta proveniente dalla Beozia data da Omero, affermò [26, 2]

<sup>2</sup>Al contrario vi è un'interessante testimonianza dall'India e contenuta nel poema *Nalopakhyanam*: Storia di Nala, databile al IV secolo avanti Cristo. Ad un certo punto della narrazione il protagonista, Rtuparna, stima in 2095 il numero di foglie e di frutti presenti su due grossi rami di un albero frondoso, basandosi sul numero di foglie e frutti presenti su un solo ramoscello, una sorta di ramo medio. Quando, effettuato il conteggio, Váhuka manifestò la propria ammirazione per la precisione della previsione di Rtuparna, questi gli rispose [13]

dei dadi posseggo la scienza e così sono versato nei numeri

affermazione che lascia pensare come, almeno in quel contesto culturale, la matematica fosse applicata ai giochi d'azzardo.

<sup>3</sup>Laureato in Ingegneria e successivamente, nel 1925, in Matematica, vinse l'anno seguente vinse il concorso per assistente negli Osservatori Astronomici iniziando la carriera a Pino Torinese, sotto la guida di Luigi Volta. Per motivi familiari, fra il 1928 e il 1929 lasciò la carriera astronomica e si dedicò alla Matematica attuariale, divenendo condirettore delle Assicurazioni Generali di Trieste.

Infatti, di milleduecento navi, il poeta ha descritto quelle dei Beoti come fornite di centoventi uomini d'equipaggio ciascuna, quelle al comando di Filottete di cinquanta, volendo indicare, a mio avviso, le maggiori e le minori (...). Se si calcola dunque la media tra le navi minori e le più capaci, risulta chiaro che non presero il mare in molti, considerato che erano inviati da tutti i paesi di Grecia.

Più che alla media aritmetica, questo passo fa pensare alla semidisersione, cioè alla semisomma tra il minimo ed il massimo valore ottenuti delle misure di una certa grandezza. Similmente, in un celebre passo del Libro II dell'*Etica Nicomachea*, Aristotele (384-322 a.C.) illustra la dottrina del giusto *mezzo* in questi termini (Aristotele, *Etica a Nicomaco* II, 4. in [28], p. 232):

In ogni quantità è possibile determinare il più, il meno, il giusto mezzo, e ciò tanto nella cosa in sé, quanto per rapporto con noi; il mezzo sta tra l'eccesso ed il difetto. Intendo dire con mezzo d'una cosa in sé ciò che è ugualmente distante dai due estremi, che è uno e lo stesso per tutti; con mezzo per rapporto con noi, ciò che non soverchia, né difetta. Questo non è uno, né il medesimo per tutti. Se, per esempio, dieci sono molti, due pochi, diremo quanto alla cosa in sé, che sei è il mezzo, poiché è di tanto superiore al due, di quanto è inferiore ai dieci, questo adunque è il mezzo secondo la proporzione aritmetica. Ma per il rapporto con noi la cosa non sta così. Se, ad esempio, per uno è molto il mangiar dieci mine,<sup>4</sup> e poco il mangiarne due, non per questo il maestro di ginnastica gli prescriverà di mangiarne sei. Può darsi infatti che anche ciò sia molto o sia poco per chi deve mangiarle (...). Così ogni persona assennata fugge l'eccesso come il difetto e cerca il mezzo e questo sceglie, e non già il mezzo della cosa in sé, ma il mezzo per rapporto con noi.

Sottolineiamo in questo passaggio di Aristotele la distinzione tra un'operazione *oggettiva* di media, che coincide con la semidisersione, ed una media *soggettiva*, per così dire, di cui si prescrive solo l'intervallo di appartenenza. Altri brani di opere di Aristotele sono stati invocati come contenenti tracce di ragionamenti probabilistici; in particolare questo passo dalla *Metafisica*:

Poiché "essere" ha in generale molti significati, uno dei quali è "essere per caso", dobbiamo anzitutto considerare ciò che "è" in questo senso. Chiaramente nessuna delle scienze tradizionali si occupa di ciò che è accidentale (...). Che una scienza di quanto è accidentale non possa essere nemmeno possibile risulta evidente nel momento in cui cerchiamo di capire cosa sia ciò che è accidentale. Accidentale è quanto si verifica ma non sempre né necessariamente, né per la maggior parte dei casi (...) ogni scienza riguarda ciò che è sempre o accade nella maggior parte dei casi ma ciò che è accidentale non rientra in nessuna di queste classi.

Affiora in questo passaggio un'idea di *casualità* che però Aristotele pone al di fuori di ogni possibile discorso scientifico, ed un concetto di probabilità

<sup>4</sup>Unità di peso diffusa nell'oriente ed in Grecia, indicava anche il nome di una moneta, la sessantesima parte del talento.

*epistemica*, legata cioè alla nostra conoscenza limitata e parziale della realtà, che riguarda ciò che accade il più delle volte.

Infine, nel *De generatione animalium* Aristotele collegò la comparsa casuale di individui di un sesso piuttosto che di un altro con una necessità di ordine naturale, quella di ottenere un rapporto ottimale tra maschi e femmine di una certa specie per garantirne la sopravvivenza. È questo un tema che ritornerà negli studi di probabilità del XVII e XVIII secolo, a cui dedicò attenzione anche Laplace [16].

Altre testimonianze di interesse riguardano la medicina antica, segnatamente Ippocrate (460-377 a. C. ca.) e Galeno (130-210) [24]. Per Ippocrate il medico deve essere in grado di prevedere quando e come somministrare un farmaco; non tutti i pazienti sono uguali ma le risposte ad un trattamento medico differiscono. Sono affermazioni che denotano una consapevolezza almeno implicita dell'importanza delle fluttuazioni causali. Nelle opere di Ippocrate compaiono anche esempi di correlazioni qualitative, come l'osservazione che le persone corpulente sono più predisposte a morire prima di quelle magre o che persone anche di una certa età abituate a sostenere degli sforzi, sono più resistenti alla fatica di persone più giovani ma prive di allenamento. In buona sostanza, le basi del ragionamento probabilistico erano presenti come supporto razionale a scelte da compiere in circostanze ordinarie o di emergenza e come premessa per costruire un discorso scientifico.

Vale la pena di ricordare che, tra i filosofi presocratici, fu formulato il *principio di ragione insufficiente* perché, a differenza di quanto accadrà nel XVIII secolo, non permise di individuare un insieme di eventi elementari cui spetta uguale probabilità, grazie ai quali risalire alla probabilità di eventi composti [32]. Al contrario, riconoscere la presenza di più casi egualmente probabili condusse ad escludere che essi si possano realizzare. Per esempio, per Anassimandro (610-540 a.C. ca.), la Terra era collocata al centro dell'universo per ragioni di simmetria: ogni direzione ne ha una opposta e siccome non vi sono più ragioni perché la Terra debba sceglierne una piuttosto che un'altra, si conclude necessariamente che essa non si muove affatto. In epoca successiva, il filosofo scettico Carneade di Cirene (214-129 a.C.) elaborò una teoria che combinava scetticismo e conoscenza in cui propose una scala di vari gradi di attendibilità che una impressione derivante dall'esterno può produrre su un soggetto, a seconda che questa sia credibile, credibile e consistente, credibile, consistente e verificata [32]. Questa graduazione si può vedere come una forma di probabilità soggettiva, nella quale manca però ogni tentativo di quantificare numericamente i vari gradi di convincimento prodotti nel soggetto da impressioni o fenomeni esterni.

## 1.2 Spunti dalla letteratura

Riferimenti a giochi d'azzardo sono sparsi in varie testimonianze letterarie. Ne evidenziamo tre che riteniamo di interesse per i nostri obiettivi, tutti centrati su uno stesso problema: numerare correttamente gli esiti del lancio di tre dadi. Perché proprio *tre* dadi? Per comprenderlo, alteriamo l'ordine cronologico dei

passi che vogliamo analizzare, iniziando dalle prime terzine del Canto VI del Purgatorio della *Divina Commedia* di Dante Alighieri ([1], pp. 58-59)

Quando si parte il gioco della zara,  
 colui che perde si riman dolente,  
 repetendo le volte, e tristo impara:  
 con l'altro se ne va tutta la gente;  
 qual va dinanzi, e qual di dietro il prende,  
 e qual da lato li si reca a mente:  
 el non s'arresta, e questo e quello intende;  
 a cui porge la man, più non fa pressa;  
 e così dalla calca si difende.

Il gioco descritto rapidamente da Dante, la *zara*, si giocava con *tre* dadi, raramente con due, si diffuse in tutta Europa nel Medioevo e divenne una vera e propria *malattia*. Esso consisteva nel lanciare i dadi dopo che i contendenti avevano puntato su quale sarebbe stata la somma dei punteggi ottenuti nel lancio. Non si poteva puntare sui risultati 3, 4, 17, 18 che nascevano da un'unica partizione— $3 = 1 + 1 + 1$ ,  $4 = 1 + 1 + 2$ ,  $17 = 5 + 6 + 6$  e  $18 = 6 + 6 + 6$ —anche se i casi favorevoli sono diversi: uno per 3 e 18, 3 per 4 e 17. Ricostruzioni storiche affermano che quando la somma risultava uno dei numeri “proibiti”, tutti gli astanti urlarono “zara!” per sottolineare l'annullamento della puntata. Se nessuno vinceva, ciascun contendente aggiungeva una moneta al montepremi ed il gioco riprendeva. L'etimologia del termine *zara* sembra essere la stessa di *azzardo* e deriva dall'arabo *al-zahr*, dado.

Il passo di Dante mette in evidenza la presenza di un pubblico ad assistere alla gioco della *zara*, il cui ruolo fu così descritto dal critico letterario Natalino Sapegno ([1], p. 59)

Al gioco assistevano molti sfaccendati e, alla fine, il vincitore veniva circondato da postulanti che approfittavano della sua passeggera euforia per ottenere da lui soccorsi e regalie. Un brano del giurista Odofredo<sup>5</sup> illustra assai bene questo particolare del costume: “sicut videmus in lusoribus ad taxillas vel similem ludum, nam multi stare solent ad videndum ludum, et quando unum lusorum obtinet in ludo, illi instantes solent petere aliquid sibi dari de lucro illo in ludo habito, et illi lusoires dare solent”.<sup>6</sup>

Queste assemblee di nullafacenti potevano facilmente degenerare ed anche per questo motivo le autorità civili ed ecclesiastiche emanarono ripetuti quanto disattesi divieti di giocare alla *zara*. La prima terzina si chiude con l'immagine del perdente che cerca di apprendere come migliorare i propri risultati, simbolo di tutti quegli irriducibili giocatori che ritengono di poter addomesticare il caso:

<sup>5</sup>Giurista nativo di Ostia, morto a Bologna nel 1265, l'anno della nascita di Dante. Esercitò la professione di avvocato in Italia ed in Francia per poi divenire professore presso l'università di Bologna.

<sup>6</sup>come infatti vediamo tra quanti giocano ai dadi o a giochi simili, molti sono soliti fermarsi ad osservare il gioco e, quando uno dei giocatori vince, quegli astanti sono soliti chiedere che dia loro qualcosa di quanto ottenuto nel gioco e i giocatori sono soliti concederglielo.

Se mai si possa imparare qualcosa in un'arte nella quale i grandi maestri sono sempre mendicanti.<sup>7</sup> (Benvenuto Rambaldi da Imola, [1], p. 59)

A margine notiamo che già Leonardo Pisano (Fibonacci, ca 1175-ca 1225) nella sezione dedicata a problemi di vario tipo (*Miscellanea*) nel *Liber Abaci* aveva ambientato un problema di matematica *dilettevole* ricorrendo ai tre dadi della zara, indice che già a quel tempo il gioco aveva una certa popolarità. Uno dei primi commentarii della Divina Commedia, scritto tra il 1324 ed il 1328 da Jacopo Giovanni della Lana, benché non esente da qualche svarione, contiene alcune osservazioni che mostrano almeno un po' di consapevolezza dei rudimenti del calcolo delle probabilità [16]. Spiegando il fatto che i numeri 3, 4, 17 e 18 sono scartati, della Lana affermava che ciò è fatto

per ischivare tale fastidio e per non aspettare troppo,

l'attesa essendo dovuta al basso numero di combinazioni in cui questi punteggi possono presentarsi che Della Lana, erroneamente, ritiene essere pari ad uno per tutti e quattro i punteggi estremi, errore che peraltro sarà commesso molto tempo più tardi anche da Leibniz il quale, in una lettera indirizzata a Louis Bourget nel 1714 riteneva come, nel lancio di due dadi, i punteggi di 11 e 12 si potessero ottenere con uguale facilità ([16], p. 266)

è altrettanto fattibile ottenere dodici od undici punti, perché entrambi i punteggi non si possono ottenere che in un unico modo.<sup>8</sup>

Comunque sia, pur mancando il concetto di permutazione, Della Lana era consapevole che la maggiore o minore possibilità di ottenere un risultato era legata al maggiore o minor numero di casi in cui questo poteva presentarsi. Inoltre gli era pure chiaro come le irregolarità nei risultati ottenuti successivamente fossero da attribuirsi al caso (*fortuna*).

La corretta numerazione dei diversi esiti del lancio di tre dadi, cioè l'esistenza di 56 risultati diversi, si trova già a partire dal X secolo quando, nel 965, Wibold, vescovo di Cambrai ed Arras, nel nord della Francia, ideò un gioco da tavolo, il *Ludus Regularis*, le cui regole sono state tramandate dallo storico Balderico nel *Chronicon Cameracense et Atrebatense*. Il gioco richiedeva il lancio di tre dadi e poteva essere giocato anche dal clero senza che per questo fosse esposto alle tentazioni del gioco d'azzardo. L'idea di Wibold era quella di associare, ad ogni possibile punteggio complessivo, una virtù che era, per così dire, acquisita dal giocatore che aveva effettuato il lancio. Ad esempio, la virtù più importante, la carità, corrispondeva all'esito (1, 1, 1), l'umiltà al punteggio più alto (6,6,6). Wibold elencò gli esiti possibili in un lancio, in ordine lessicografico. Le virtù conseguite potevano essere combinate quando la somma dei loro punteggi era 21, come nel caso di carità ed umiltà, ovvero di castità (1,2,4) ed intelligenza (4,5,5) [31].

<sup>7</sup>Si tamen discere potest aliquid in ea arte, in qua magni magistri sunt sempre mendici.

<sup>8</sup>il est aussi faisable de jeter douze points que d'en jeter onze, car l'un et l'autre ne peut se faire que d'une seule manière

Molto interessante è poi un poema in lingua latina, scritto nel XIII secolo: il *De Vetula*, letteralmente la vecchia signora, il cui autore è stato identificato, non senza riserve, nell'erudito francese Richard de Fournival (1201-1260), cancelliere del capitolo della cattedrale di Notre-Dame ad Amiens [3]. Il poema in tre libri vorrebbe essere una autobiografia di Publio Ovidio Nasone, di cui si narra il passaggio da una giovinezza dedita ai piaceri dei sensi all'età matura in cui egli si sarebbe convertito al Cristianesimo. Tra i passatempi coltivati in giovinezza, l'autore menziona i giochi con i dadi, fornendo ragioni per tenersene lontano. Il passo che vogliamo analizzare si trova nel Libro I e contiene una articolata discussione dei possibili punteggi e delle loro frequenze nel lancio di tre dadi:

Poiché un dado ha sei facce e sei numeri,  
 in tre dadi ve ne saranno diciotto.  
 Se non tre di questi possono sopravvivere nei dadi.  
 Questi variano in modi diversi e ne derivano  
 sedici numeri composti, non però allo stesso modo,  
 dal momento che i numeri più piccoli e i più grandi di questi  
 si presentano di rado, i medi più frequentemente.  
 Gli altri tanto più quanto sono più vicini ai medi.<sup>9</sup>

L'autore notò dunque che i punteggi totali si presentano come diverse combinazioni di punteggi parziali, distribuiti simmetricamente attorno ai punteggi centrali, secondo la tabella seguente:

3	(1, 1, 1)					
4	(2, 1, 1)					
5	(3, 1, 1)	(2, 2, 1)				
6	(4, 1, 1)	(3, 2, 1)	(2, 2, 2)			
7	(5, 1, 1)	(4, 2, 1)	(3, 3, 1)	(2, 2, 3)		
8	(6, 1, 1)	(5, 2, 1)	(4, 3, 1)	(4, 2, 2)	(3, 3, 2)	
9	(6, 2, 1)	(5, 3, 1)	(5, 2, 2)	(4, 4, 1)	(4, 3, 2)	(3, 3, 3)
10	(6, 3, 1)	(6, 2, 2)	(5, 4, 1)	(5, 3, 2)	(4, 4, 2)	(4, 3, 3)
11	(6, 4, 1)	(6, 3, 2)	(5, 5, 1)	(5, 4, 2)	(5, 3, 3)	(4, 4, 3)
12	(6, 5, 1)	(6, 4, 2)	(6, 3, 3)	(5, 5, 2)	(5, 4, 3)	(4, 4, 4)
13	(6, 6, 1)	(6, 5, 2)	(6, 4, 3)	(5, 5, 3)	(4, 4, 5)	
14	(6, 6, 2)	(6, 5, 3)	(6, 4, 4)	(5, 5, 4)		
15	(6, 6, 3)	(6, 5, 4)	(5, 5, 5)			
16	(6, 6, 4)	(6, 5, 5)				
17	(6, 6, 5)					
18	(6, 6, 6)					

<sup>9</sup>Cum decius sit sex laterum, sex etiam numerorum simplicium, tribus in decius sunt octo decemque. Quorum non nisi tres possunt decius superesse. Hi diversimode variantur, et inde bis octo compositi numeri nascuntur, nontamen aequae virtutis, quoniam maiores atque minores ipsorum raro venient, mediique frequenter. Et reliqui quanto mediis quamvis propriores.

Come Wibold, anche l'autore del *De Vetula* numera correttamente i 56 possibili esiti (*punctaturae*) nel lancio dei tre dadi. Egli però, a differenza di Wibold, procede ulteriormente in questa analisi ed osserva come ciascuno dei *sei* punteggi del tipo  $(a, a, a)$  possa presentarsi una sola volta mentre ciascuno dei *trenta* punteggi del tipo  $(a, a, b)$  si può presentare *tre* volte ( $3 = \frac{3!}{2!}$ ). Infine, ognuno dei *venti* punteggi del tipo  $(a, b, c)$  ottenuto da tre addendi tutti diversi tra loro può presentarsi *sei* volte:

Ora, quando i tre numeri che compongono il lancio saranno simili,  
poiché sei si possono comporre,  
anche i punteggi saranno sei, uno per ogni composizione.  
Quando però un numero è differente dagli altri due  
che invece sono simili, i punteggi possono presentarsi in trenta modi diversi  
poiché se si presenta due volte uno qualsiasi dei sei numeri  
e si aggiunge uno degli altri, si ottiene trenta,  
come se avessi moltiplicato sei per cinque.  
Se poi i tre numeri saranno tutti diversi,  
allora conterai venti punteggi.

(...)

Inoltre, vi sono alcune cose più sottili  
da osservare circa i punteggi tra i quali ve ne sono  
con una sola occorrenza, oppure tre o sei, siccome  
quando i tre numeri sono simili non è possibile  
distinguere gli esiti del lancio. Se uno di loro è distinto  
e gli altri sono simili, si hanno tre esiti del lancio  
avendoli sovrapposti a ciascuno dei dadi diversi.  
Se però sono tutti diversi, troverai che un punteggio  
si può ottenere in sei modi perché,  
fissato il posto ad uno di questi, gli altri possono scambiarsi;  
così come insegna l'alternanza delle partizioni.  
E pertanto le varie partizioni si differenziano  
in cinquantasei modi.<sup>10</sup>

---

<sup>10</sup>Nam quando similes fuerint sibi tres numeri, qui  
iactum componunt, quia sex componibiles sunt  
et punctaturae sunt sex, pro qualibet una.  
Sed cum dissimilis aliis est unus eorum  
atque duos similes, triginta potest variari  
punctatura modis, quia si duplicaveris ex sex  
quemlibet, adiuncto reliquorum quolibet, inde  
producens triginta quasi sex quintuplicatis.  
Quod si dissimiles fuerint omnino sibi tres,  
Tunc punctaturas viginti connumerabis.

(...)

Rursum sunt quaedam subtilius inspicienti  
de punctaturis, quibus una cadentia tantum est  
suntque quibus sunt tres aut sex, qua schema cadendi  
Tunc differre nequit, quando similes fuerint tres  
praedicti numeri. Si vero si unus eorum  
dissimilis, similesque duo, tria schemata surgunt

### 1.3. IL LANCIO DEI DADI NEL “GENERAL TRATTATO” DI NICCOLÒ FONTANA11

È chiara la distinzione per ciascun punteggio tra le partizioni (*punctaturae*) associate ad un punteggio totale e l’occorrenza (*cadentia*) di questo punteggio, differenza chiarita da un’altra tabella

3	18	Punctatura	1	Cadentia	1
4	17	Punctatura	1	Cadentiae	3
5	16	Puncturae	2	Cadentiae	6
6	15	Puncturae	3	Cadentiae	10
7	14	Puncturae	4	Cadentiae	15
8	13	Puncturae	5	Cadentiae	21
9	12	Puncturae	6	Cadentiae	25
10	11	Puncturae	6	Cadentiae	27

L’autore del *De Vetula* aveva pieno possesso, diremmo oggi, della natura dello spazio degli eventi relativo al lancio di tre dadi, anche se non ne sviluppò conseguenze probabilistiche. Lo spazio degli eventi relativo al lancio di tre dadi compare anche nel *De Ludo Aleae* di Cardano, ai Capitoli XII e XIII e nell’opuscolo di Galileo *Sopra le scoperte dei dadi*, pubblicato postumo nel 1718 [11].<sup>11</sup> In [8], Bellhouse suggerì la possibilità che Cardano fosse a conoscenza del *De Vetula*, poema che certamente dovette essere molto conosciuto, dal momento che esistono almeno sessanta manoscritti sparpagliati in tutta Europa [3]. Vi sono però anche motivi che possono lasciar intendere una indipendenza delle conclusioni di Cardano da una conoscenza del poemetto pseudo-ovidiano visto che egli non lo citò mai, a differenza della sua abitudine a rendere il giusto riconoscimento per il lavoro altrui.

### 1.3 Il lancio dei dadi nel “General Trattato” di Niccolò Fontana

Concentriamoci sul contributo di Niccolò Fontana (detto il Tartaglia, 1499-1557) allo studio del lancio di tre dadi. Tartaglia pubblicò nel 1556 una voluminosa opera in sei parti, il *General Trattato de’ numeri et misure* [33], in cui sono formulati e risolti diversi problemi di calcolo combinatorio e probabilità. Noi siamo interessati alla chiusura del Libro I della II parte del *Trattato*, in cui Tartaglia espone la

Regola generale dal presente autore ritrovata il primo giorno di quaresima l’anno 1523, in Verona, di saper trovare in quanti modi può variar il getto di che quantità di dati si voglia nel tirar quelli.

---

dissimili cuicunque; superposito deciorum. Sed si dissimiles sunt omnes, invenies sex Verti posse modis, quia, quemlibet ex,tribus uni Cum dederis, reliqui duo permutant loca; sicut Punctaturarum docet alternatio. Sicque Quinquaginta modis et sex diversificantur In punctaturis, punctaturaeque ducentis

<sup>11</sup>Osservo a margine come il titolo dell’opera sia stato tradotto erroneamente in [14]: le scoperte sono i punteggi realizzati nei singoli lanci, non delle scoperte (*discoveries*) nel senso attuale del vocabolo.

Seguiamo il racconto che Tartaglia dà della sua scoperta

Stanziando l'anno 1523 in Verona, e il giorno di Carnevale una comitiva di giovinetti, e di altri di matura età traevano con 3 dadi sul libro (detto) della ventura di Lorenzo Spirto,<sup>12</sup> cercando ciascuno di loro di intendere quello che tal libro gli determinava circa alle materie, che tal libro prepone da notificarli. Et vedendo che in ogni carta li detti 3 dati con la isperienza havea il detto autore trovato poter variar in 56 modi, la qual cosa considerando deliberai di voler trovare, come che con regola generale tal cosa si potesse determinare, e non solamente in detti 3 dati, ma in ogni altra maggior quantità di dati, e così tutta la notte sopra tal materia andai tanto freneticando che il giorno seguente (che fu il primo di Quaresima) trovai tali ordini, over regole formarsi da strane sorte di progressioni, come intenderai.

Prima egli è manifesto che un sol dato può variare in 6 modi per esser di 6 fазze, over di 6 basi, nelle quali sono 6 ordini di numeri, cioè 1. 2. 3. 4. 5. e 6. come in figura vedi.

Ma per trovar in quanti modi può variar il getto di due dati trovai che raccogliendo tutte le unità, che sono da 1 per fino in 6 nella sopra notata progression continua, che fanno 21 e così in 21 modo trovai poter variar il getto di duoi dati.

Tre dati poi ponno variar il loro getto nella summa di questi 6 termini di progression 1, 3, 6, 10, 15, 21 la qual summa sarà 56.

Li 4 dati ponno variar il lor getto nella summa di questi altri 6 termini di progression 1, 4, 10, 20, 35, 56, la qual somma sarà 126.

Li 5 dati ponno variar il lor getto nella summa di questi altri 6 termini di progression 1, 5, 15, 35, 70, 126, la qual somma sarà 252.

Li 6 dati ponno variar il lor getto nella summa di questi altri 6 termini di progression 1, 6, 21, 56, 126, 252, la qual somma sarà 462.

Li 7 dati ponno variar il lor getto nella summa di questi altri 6 termini di progression 1, 7, 28, 84, 210, 462, la qual somma sarà 792.

Li 8 dati ponno variar il lor getto nella summa di questi altri 6 termini di progression 1, 8, 36, 120, 330, 792, la qual somma sarà 1287.

Ma a volerti mo dichiarare minutamente in scrittura l'origine di tutti li sopra notati 6 termini di progressioni bisognaria formarvi sopra un libro, ma accioché in parte resti satisfatto, sappi che ogni una di dette progressioni si forma dalla progression anciana, e la prima progression viene a esser di 6 termini di una unità per termine in questa forma 1, 1, 1, 1, 1, 1, et così la summa di questi 6 termini di progression può variar il getto di un dato solo, come vedi in figura. Et nota che l'ultimo termine di ciascuna di dette progressioni viene a esser la summa della anciana progression,

<sup>12</sup>Nel *Libro delle Sorti*, o libro della ventura, volume pubblicato nel 1482 da Lorenzo Gualtieri (detto Spirito) (1426(?)-1496), sono riportati correttamente i punteggi possibili, a meno di permutazioni, nel lancio di tre dadi. Il *Libro delle Sorti* era un libro di divinazioni utilizzato con grande successo per giochi di società. Fu ristampato fino al '600 ma, con altri libri simili, finì all'indice dopo il Concilio di Trento.

### 1.3. IL LANCIO DEI DADI NEL “GENERAL TRATTATO” DI NICCOLÒ FONTANA<sup>13</sup>

come nella figura puoi vedere, e con tal ordine potrai saper li 10000 dati in quanti modi ponno variar il loro getto.

per 1 dato	1	1	1	1	1	1
per 2 dati	1	2	3	4	5	6
per 3 dati	1	3	6	10	15	21
per 4 dati	1	4	10	20	35	56
per 5 dati	1	5	15	35	70	126
per 6 dati	1	6	21	56	126	252
per 7 dati	1	7	28	84	210	462
per 8 dati	1	8	36	120	330	792

Tartaglia non fornisce alcuna indicazione sul metodo seguito per giungere a questo risultato ed occorre affidarsi alle ricostruzioni storiche. In [10], Edwards propose uno schema ricorsivo deducibile dalla tabella riportata qui sopra. Il metodo proposto poggia sull’ordinamento opportuno dei punteggi ottenuti su  $n$  dadi tramite  $n$ -uple ordinate. Nel caso di due dadi, consideriamo le coppie  $(p_1, p_2)$ , con  $p_1 \geq p_2$ : si combina in questo modo l’esito del lancio del primo dado con i punteggi ottenuti dal secondo dado che *non* superano il primo punteggio. Così raggrupperemo i punteggi diversi secondo lo schema

(1, 1)  
 (2, 2), (2, 1)  
 (3, 3), (3, 2), (3, 1)  
 (4, 4), (4, 3), (4, 2), (4, 1)  
 (5, 5), (5, 4), (5, 3), (5, 2), (5, 1)  
 (6, 6), (6, 5), (6, 4), (6, 3), (6, 2), (6, 1)

che non solo conduce ai 21 punteggi dichiarati da Tartaglia ma mostra anche il cammino per ottenere il risultato in generale. Infatti, una coppia del tipo  $(1, p_2)$  si ottiene solo quando  $p_2 = 1$ ; una coppia  $(2, p_2)$  si ottiene quando  $p_2 = 1, 2$  e così via fino alle coppie del tipo  $(6, p_2)$  che non limitano in alcun modo i sei valori possibili per  $p_2$ . Indichiamo con  $K_j(\ell)$  il numero di risultati ottenuti con  $\ell$  dadi e che hanno  $j$  come punteggio più alto, ovvero come primo elemento della  $n$ -upla associata. Avremo

$$K_j(1) = 1 \quad \forall j = 1, \dots, 6$$

e lo schema descritto in precedenza mostra che

$$\begin{aligned} K_1(2) &= K_1(1) = 1 & K_2(2) &= K_2(1) + K_1(1) = 2 \\ K_3(2) &= K_3(1) + K_2(1) + K_1(1) = 3 \\ K_4(2) &= K_4(1) + K_3(1) + K_2(1) + K_1(1) = 4 \\ K_5(2) &= K_5(1) + K_4(1) + K_3(1) + K_2(1) + K_1(1) = 5 \\ K_6(2) &= K_6(1) + K_5(1) + K_4(1) + K_3(1) + K_2(1) + K_1(1) = 6. \end{aligned}$$

Questo ragionamento può essere ripetuto quanto si vuole e si può riassumere nell’espressione

$$K_j(\ell + 1) = \sum_{i=1}^j K_i(\ell) \quad \forall \ell, \quad \forall j = 1, 2, \dots, 6$$

che consente di riprodurre la tabella di Tartaglia. Quale sia stato il metodo seguito per ottenere questo risultato, il passo appena riportato è istruttivo per mostrare la posizione di un matematico di rango di fronte ad un problema: non solo lo traduce in formule ma ne cerca la generalizzazione, anche se questa non è di per sé necessaria alla comprensione del problema che l'ha generata.

Poco tempo dopo la pubblicazione del *General Trattato*, il matematico francese Jean Borrel (latinizzato in Johannes Buteo, 1492-1572) trattò il problema dei punteggi nel lancio di più dadi, fermandosi al caso di quattro dadi, nel Cap. IV della *Logistica* [6]. Non vi è una dimostrazione formale ma le tabelle contenute nella *Quaestio 90*, mostrano come egli seguisse un ordinamento dei risultati secondo lo schema della ricostruzione precedente: a differenza di Tartaglia, Buteo non dà l'idea di possedere la formula per trattare il caso generale. Osserviamo poi che Tartaglia non considerò le *cadentiae* dei punteggi ma solo le *punctaturae*.

## 1.4 Galileo Galilei e la probabilità

Il gioco della zara è uno dei giochi menzionati da Galileo Galilei (1564-1642) in un breve scritto: *Sopra le scoperte dei dadi*, di cui non si conosce la data di composizione. Lo stile è lontano da quello del *De Vetula* ma l'analisi di Galileo è corretta nel valutare come le permutazioni distinte con cui si presenta un dato punteggio concorrano a determinare il numero di casi favorevoli, insieme al numero delle partizioni del punteggio stesso. Ciò che un ignoto committente (*per servire a chi m'ha comandato*), forse il Granduca di Toscana, si attendeva da Galileo era la spiegazione di una osservazione maturata con la frequentazione dei tavoli da gioco: nel lancio di tre dadi, il 9 ed il 10 si possono comporre in un numero uguale di modi diversi: il 9 con le *sei* terne (1, 2, 6), (1, 3, 5), (1, 4, 4), (2, 2, 5), (2, 3, 4), (3, 3, 3); il 10 con le *sei* terne (1, 3, 6), (1, 4, 5), (2, 2, 6), (2, 3, 5), (2, 4, 4), (3, 3, 4). A dispetto di questo,

la lunga osservazione ha fatto dai giocatori stimarsi più vantaggioso il 10...  
che il 9...

Galileo rispose sostanzialmente riproponendo in altri termini la distinzione tra *cadentia* e *punctatura* già presente nel *De Vetula*. Riportiamo alcuni stralci del testo originale ([11], pp. 56-57).

Che nel gioco dei dadi alcuni punti sieno più vantaggiosi di altri, vi ha la sua ragione assai manifesta, la quale è, il poter quelli più facilmente e più frequentemente scoprirsi, che questi, il che dipende dal potersi formare con più sorte di numeri (...). Ora io per servire a chi m'ha comandato, che io debba produr ciò, che sopra tal difficoltà mi sovviene, esporrò il mio pensiero, con isperanza (...) di aprire la strada a poter puntualmente scorgere le ragioni, per le quali tutte le particolarità del giuoco sono state con grande avvedimento e giudizio compartite ed aggiustate. E per condurmi colla maggior chiarezza che io possa al mio fine, comincio

a considerare come essendo un dado, terminato da 6 facce, sopra ciascuna delle quali gettato, egli può indifferentemente fermarsi; sei vengono ad essere le sue scoperte, e non più, l'una differente dall'altra. Ma se noi insieme col primo getteremo il secondo dado, che pure ha altre sei facce, potremo fare 36 scoperte tra di loro differenti, poiché ogni faccia del primo dado può accoppiarsi con ciascuna del secondo, ed in conseguenza fare 6 scoperte diverse [per ogni faccia scoperta del primo]; onde è manifesto tali combinazioni esser sei volte 6 cioè 36. E se noi aggiungeremo il terzo dado, perché ciascuna delle sue facce, che pur son sei, può accoppiarsi con ciascuna delle 36 scoperte degli altri due dadi, avremo le scoperte di tre dadi esser 6 volte 36 cioè 216 tutte tra di loro differenti. Ma perché i punti dei tiri di tre dadi non sono se non 16, cioè 3, 4, 5 sino a 18, tra i quali si hanno a compartire le dette 216 scoperte, è necessario, che ad alcuni di essi ne tocchino molte; e se noi ritroveremo quante ne toccano per ciascheduno, avremo aperta la strada di scoprire quanto cerchiamo, e basterà fare tale investigazione dal 3 sino al 10 perché quello che converrà a uno di questi numeri, converrà ancora al suo sossopra. Tre particolarità si debbono notare per chiara intelligenza di quel che resta: la prima è, che quel punto dei tre dadi, la cui composizione risulta da tre numeri eguali, non si può produrre, se non da una sola scoperta, ovvero tiro di dadi, e così il 3 non si può formare se non dalle tre facce dell'asso, ed il 6, quando si dovesse comporre con tre dui, non si farebbe se non da una sola scoperta. Seconda: il punto, dai tre numeri, due dei quali sieno i medesimi, e i terzo diverso, si può produrre da tre scoperte, come v.g. il 4 che nasce dal 2 e dalli due assi, può farsi con tre cadute diverse, cioè quando il primo dado sopra 2 e il secondo e il terzo scuoprano asso; o scuoprendo il secondo dado 2, e il primo e il terzo asso; o scuoprendo il terzo 2, ed il primo e secondo asso. E così v.g. l'8 in quanto risulta da 3.3.2. può prodursi parimenti in tre modi; cioè scuoprendo il primo dado 2 e gli altri 3 per uno, o scuoprendo il secondo dado 2 ed il primo e terzo 3 o finalmente scuoprendo il terzo dado 2 ed il primo e secondo 3. Terza: quel numero di punti, che si compone di tre numeri differenti, può prodursi in 6 maniere, come per esempio, l'8 mentre si compone da 1 3 4 si può fare con 6 scoperte differenti; prima, quando il primo dado faccia 1 il secondo 3 e il terzo 4; seconda, quando il primo dado faccia pur 1 ma il secondo 4 e il terzo 3; terza, quando il secondo dado faccia 1 e il primo 3 e il terzo 4; quarta, facendo il secondo pur 1 e il primo 4 e il terzo 3; quinta, quando facendo il terzo dado 1, il primo faccia 3 e il secondo 4; sesta, quando sopra l'1 del terzo dado, il primo farà 4 e il secondo 3. Abbiamo dunque sin qui dichiarati questi tre fondamenti, primo, che le triplicità, cioè il numero delle scoperte dei tre dadi, che si compongono da tre numeri eguali, non si producono se non in un modo solo; secondo, che le triplicità che nascono da due numeri uguali, e dal terzo differente, si producono in tre maniere; terzo, che quelle che nascono da tre numeri tutti differenti, si formano in sei maniere. Da questi fondamenti facilmente raccorremo in quanti modi, o vogliam dire, in quante scoperte differenti

si possono formare tutti i numeri [o punti] dei tre dadi (...) Si vede, come il punto 10 può farsi da 27 scoperte di dadi differenti ma il punto 9 [come il punto 12] da 25 solamente.

Dunque la frequenza teorica con cui può presentarsi il punteggio complessivo di 9 è di  $\frac{25}{216}$  mentre quella con cui si presenta il 10 è pari a  $\frac{27}{216}$ : il committente dello studio di Galileo, come altri giocatori incalliti, ha ritenuto significativa una differenza di meno dell'uno per cento, il che fa nascere spontaneamente la domanda su quanti debbano essere i lanci necessari perché questa differenza sia significativa dal punto di vista statistico [14]. Vedremo più avanti i contributi di Galileo alla teoria degli errori.

## 1.5 *Punctaturae e cadentiae* calcolate con altri metodi

Abbiamo visto come la determinazione delle *punctaturae*, il numero di partizioni distinte in interi del punteggio complessivo nel lancio di più dadi, e delle *cadentiae*, cioè il numero di casi favorevoli al verificarsi di quel punteggio, fossero stati determinati in diversi modi. In questa sezione vediamo due approcci al problema contenuti in importanti pubblicazioni dell'inizio del XVIII secolo: l'*Essay d'Analyse sur les jeux d'hazard* pubblicato dal matematico francese Pierre Rémond de Montmort (1678-1719) in due edizioni—nel 1708 e nel 1713—e l'*Ars conjectandi* di Jakob Bernoulli (1654-1705). Avremo modo di riornare in seguito con maggiori particolari su queste opere; qui ci limitiamo a dar conto di come sono state determinate *punctaturae* e *cadentiae*. Montmort determinò le *punctaturae* nella proposizione XI del *Traité des combinaisons*, capitolo iniziale dell'*Essay*, nel caso di  $p$  dadi, ciascuno con  $f$  facce, senza restringersi al caso  $f = 6$ :

Trovare quanti esiti diversi si possono ottenere con un numero  $p$  di dadi, aventi ognuno un numero  $f$  di facce.<sup>13</sup>

Il metodo proposto da Montmort è in sostanza analogo a quello che doveva aver seguito da Tartaglia ma produce una formula diretta che risponde al problema in funzione di  $p$  ed  $f$ : il numero cercato è

$$\frac{f \cdot (f + 1) \cdot \cdots \cdot (f + p - 1)}{p!} = \binom{f + p - 1}{p}. \quad (1.1)$$

Montmort ottiene questo risultato attraverso uno studio approfondito del triangolo aritmetico, già utilizzato da Pascal con successo in problemi di probabilità, grazie al quale egli determinò anche le combinazioni di  $n$  oggetti presi a gruppi

<sup>13</sup>Trouver combien on peut amener de coups differens avec un nombre quelconque  $p$  de dés, dont le nombre des faces  $f$  soit aussi quelconque.

di  $k$ . Il triangolo era stato scritto in questa forma

1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	2	3	4	5	6	7	8	9	10	11	12	13
		1	3	6	10	15	21	28	36	45	55	66	78
			1	4	10	20	35	56	84	120	165	220	286
				1	5	15	35	70	126	210	330	495	715
					1	6	21	56	126	252	462	792	1287
						1	7	28	84	210	462	924	1716
							1	8	36	120	330	792	1716
								1	9	45	165	495	1287
									1	10	55	220	715
										1	11	66	286
											1	12	78
												1	13
													1

e la formazione di una riga, per esempio la terza, a partire dalla precedente, era spiegata in questi termini:

La terza riga orizzontale si ottiene in questo modo a partire dalla seconda; 1°: mi sposto di una cella da sinistra a destra: 2° per formare il numero da inserire in ogni cella di questa riga, sommo tutti i numeri che lo precedono a sinistra nella riga orizzontale superiore. In questo modo il numero sei, terzo numero della terza riga orizzontale, è uguale alla somma del primo, del secondo e del terzo numero della seconda riga orizzontale.<sup>14</sup> ([19], pp. 2-3)

Montmort spiegò nella Proposizione II come ottenere i termini di una colonna del triangolo. Ogni colonna inizia con 1 e la  $p$ -esima colonna<sup>15</sup> avrà, sotto l'unità, proprio il numero  $p$ , come segue dalla regola di formazione. Montmort vuole mostrare che, detti  $B$ ,  $C$ ,  $D$ ,... i termini successivi che stanno sotto  $p$  nella  $p$ -esima colonna, allora

$$B = p \frac{p-1}{2} \quad C = p \frac{p-1}{2} \frac{p-2}{3} \quad D = p \frac{p-1}{2} \frac{p-2}{3} \frac{p-3}{4}. \quad (1.2)$$

La dimostrazione è condotta per induzione ([19], p. 10):

Occorre ora mostrare che, se questa regola vale per caso lungo una colonna, deve valere necessariamente per tutte le altre.<sup>16</sup>

<sup>14</sup>La troisième bande horizontale est formée sur la seconde en cette manière; 1° je rétrograde de gauche à droite d'une celule: 2°, pour former le chiffre de chaque celule de cette bande, j'ajoute tous les chiffres qui les précèdent à gauche dans la bande supérieure horizontale. Ainsi le nombre six, troisième chiffre de la troisième bande horizontale, est égal à la somme du premier, du second et du troisième chiffre de la seconde bande horizontale.

<sup>15</sup>La colonna che contiene solo una unità viene etichettata come colonna 0.

<sup>16</sup>Il faut maintenant prouver que cette règle ayant lieu comme par hazard pour cette colonne, elle a lieu par nécessité à l'égard des autres.

La curiosa espressione *se questa regola vale per caso lungo una colonna* esprime il passo iniziale dell'induzione che consta di una verifica diretta. Messo in questa forma, Montmort rende l'idea dell'imprevedibilità di una osservazione—qui l'esistenza di un legame tra i valori degli elementi successivi di una colonna del triangolo aritmetico—che permette di far partire il ragionamento induttivo. Il lemma è dimostrato facendo tesoro di un'osservazione, che Montmort riportò nell'articolo iniziale, e che si ottiene immediatamente dalla costruzione del triangolo aritmetico: detto  $C_{n,p}$  il numero della  $n$ -esima riga che occupa la  $p$ -esima colonna del triangolo aritmetico, allora

$$C_{n,p} = C_{n,p-1} + C_{n-1,p-1}. \quad (1.3)$$

Se supponiamo valide le (1.2), applicando (1.3) al secondo e terzo elemento della  $p$ -esima colonna,  $p$  e  $p\frac{p-1}{2}$ , si ottiene il valore del terzo elemento della  $(p+1)$ -esima colonna:

$$p\frac{p-1}{2} + p = p\frac{p+1}{2} = (p+1)\frac{p}{2},$$

confermando la bontà della regola. Montmort verifica che lo stesso vale per un altro paio di elementi e ciò gli basta per convincersi della validità del lemma. Tornando al lancio dei dadi, Montmort liquida rapidamente il caso in cui vi sia un solo dado e costruisce l'insieme dei risultati possibili per due dadi, depurandoli dalle ripetizioni.

Per sapere quanti esiti differenti vi sono con due dadi, occorre sommare i primi sei numeri della seconda riga orizzontale; è infatti evidente che unendo l'asse del secondo dado alle sei facce del primo, si ottengono 6[ risultati], ed unendo il 2 del secondo dado con le cinque facce del primo da cui si esclude l'asse, se ne ottengono 5; e quando si unisce il 3 del secondo dado con le quattro facce del primo su cui non figurano né l'asse né il 2, se ne ottengono 4, ecc. Si vede dunque che ci sono in tutto  $6 + 5 + 4 + 3 + 2 + 1 = 21$  esiti differenti in cui ciascuna delle facce può presentarsi.<sup>17</sup> ([19], p. 37)

Anche se Montmort espose i dettagli solo per il caso  $f = 6$ , il caso generale enunciato si ottiene ponendosi nella colonna  $f - 1$  della prima riga del triangolo aritmetico e scendendo verso destra in diagonale di  $p$  righe. L'equazione (1.2), applicata all'elemento della colonna  $f + p - 1$  e della riga  $p$ , permette di ottenere l'espressione (1.1).

Passando al calcolo delle *cadentiae*, abbiamo visto che l'autore del *De Vetula*, Cardano e Galilei limitarono l'attenzione ai tre dadi, dove il peso di una elencazione minuta di tutti i casi possibili era ancora sopportabile. All'aumentare

<sup>17</sup>pour sçavoir combien il y a de coups differens avec deux dés, il faut ajouter en une somme les six premiers nombres de la seconde bande horizontale; puisq'il est évident qu'en joignant l'as du deuxième dé avec les six faces du premier dé on a 6, et le 2 du deuxième dè avec les cinq faces du premier où l'as n'est point, on a 5; et qu'en joignant le 3 du deuxième dé avec les quatre faces du premier où l'as ni le 2 se trouvent point, on a 4, etc. En sorte qu'il y a en tout  $6 + 5 + 4 + 3 + 2 + 1 = 21$  coups differens où chacune des faces peut entrer.

del numero di dadi il metodo di elencazione è però impraticabile. Nell'*Ars Conjectandi* Jakob Bernoulli trovò la probabilità di ottenere un certo punteggio nel lancio di un numero arbitrario di dadi ordinari a sei facce. Bernoulli considerò equivalenti tutti gli esiti grazie alla simmetria perfetta dei dadi:

Anzitutto per il primo dado vi sono sei risultati diversi, ciascuno dei quali ha uguale facilità di presentarsi. Suppongo infatti che il dado abbia la forma perfetta di un cubo. Così vi sono 36 diversi esiti di due dadi, ciascuno dei quali, anche in questo caso, può presentarsi con uguale facilità<sup>18</sup>

Siamo già nel cuore del problema di definire la probabilità attraverso l'individuazione di eventi elementari *equiprobabili* (*aeque facili*), evidenziando il problema della circolarità nella definizione classica di probabilità che, a dispetto di questa difficoltà, sopravvisse a lungo. Bernoulli invocò la simmetria di un dado come garanzia della uguale facilità a realizzarsi dei punteggi possibili. Egli procedette ad elencare i possibili casi nel lancio di tre e quattro dadi con un metodo che, in sostanza, non differiva da quanto aveva fatto Buteo nella logistica. Bernoulli concluse però che

Certamente questo metodo di calcolare il numero di punteggi ottenibili con più dadi è oltre modo noioso e prolisso<sup>19</sup> ([4], p. 23)

e propose allora di usare una speciale tavola

che può essere costruita molto rapidamente e mostra meglio la natura e la progressione che mostrano gli esiti dei lanci.<sup>20</sup> ([4], p. 24)

La costruzione si compone di vari passaggi.

1. si riportino su righe successive, in ordine crescente, i punteggi che si possono ottenere lanciando 1, 2, 3, ... dadi;
2. sotto ciascuno dei primi sei numeri si riporti una unità e si ripeta questa successione su altre cinque righe, con l'accortezza di iniziare ogni volta una posizione più a destra rispetto alla riga precedente;
3. si sommino le unità presenti su ciascuna colonna: i risultati rappresentano i numeri di modi in cui possono presentarsi i punteggi nel lancio di due dadi;

<sup>18</sup>Primo unius tesserae sex esse iactus diversos, quorum quivis aequae facile eveniat. Sumo enim tesseram habere figuram cubi perfectam. Porro duarum tesserarum 36 esse diversos iactus, quorum similiter quivis aequae facile obtingere potest. ([4], p. 20)

<sup>19</sup>Verum enimvero quoniam haec methodus supputandi numerum jactuum in pluribus tesseris, supra modum taediosa et prolixa est.

<sup>20</sup>quae et expedite admodum construi potest, et naturam progressionemque, quam numeri jactuum inter se servant, apertius ob oculis ponit.

4. si riportino i punteggi appena ottenuti su altre cinque righe, spostando ogni volta a destra di una posizione e si ripeta il procedimento precedente, ottenendo così i casi favorevoli ad un certo punteggio nel lancio di tre dadi. La procedura può essere ripetuta fino a raggiungere il numero  $k$  di dadi richiesto.

I	1	2	3	4	5	6														
II	2	3	4	5	6	7	8	9	10	11	12									
III	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18				
I	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>														
		1	1	1	1	1	1													
			1	1	1	1	1	1												
				1	1	1	1	1	1											
					1	1	1	1	1	1										
II	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>									
		1	2	3	4	5	6	5	4	3	2	1								
			1	2	3	4	5	6	5	4	3	2	1							
				1	2	3	4	5	6	5	4	3	2	1						
					1	2	3	4	5	6	5	4	3	2	1					
III	<b>1</b>	<b>3</b>	<b>6</b>	<b>10</b>	<b>15</b>	<b>21</b>	<b>25</b>	<b>27</b>	<b>27</b>	<b>25</b>	<b>21</b>	<b>15</b>	<b>10</b>	<b>6</b>	<b>3</b>	<b>1</b>				

Bernoulli non si dilungò su questo punto ma la serie di precetti rappresenta una traduzione schematica della legge di formazione dei coefficienti dello sviluppo di un multinomio ([8], pp. 65-66). Infatti considerando il polinomio

$$\frac{1}{6} (x^1 + x^2 + x^3 + x^4 + x^5 + x^6) :$$

un attimo di riflessione permette di concludere come il coefficiente di  $x^s$  nello sviluppo di

$$\left[ \frac{1}{6} (x^1 + x^2 + x^3 + x^4 + x^5 + x^6) \right]^k$$

fornisca la probabilità di ottenere  $s$  punti lanciando  $k$  dadi: la frazione  $\frac{1}{6^k}$  fornisce semplicemente i casi possibili ed i numeratori dei coefficienti riproducono la tabella di Bernoulli [15].

## 1.6 Il *Liber de ludo aleae* di Gerolamo Cardano

Dedichiamo questa sezione ad esporre la struttura del *Liber de Ludo Aleae* [7] di Gerolamo Cardano (1501-1576) che rappresenta un primo abbozzo di una teoria delle probabilità relativamente ai giochi d'azzardo e che permette di considerare Cardano come un precursore del calcolo delle probabilità. Questo breve opuscolo venne pubblicato solo nel 1663, nel primo volume dell'*Opera Omnia* di Cardano: in quel momento però, gli sviluppi del calcolo delle probabilità, pur

essendo ancora nella sua infanzia come disciplina scientifica, erano già sufficienti a far apparire superate le considerazioni che vi erano contenute. Il testo ha una struttura molto composita che si snoda in 32 capitoli, dove sono mescolati ragionamenti di indole matematica ad argomenti di natura disparata che risentono del fatto che Cardano era un accanito giocatore d'azzardo. L'opera si apre con una distinzione dei giochi (*De ludorum generibus*) in giochi di agilità, come il gioco della palla, di forza, come la lotta, di ingegno, come gli scacchi, per arrivare ai giochi propriamente d'azzardo, come il gioco dei dadi (*Ludus talorum*); compare anche il gioco delle carte, di introduzione relativamente recente, in cui abilità e azzardo si combinano: tra questi spicca il *Primerio*, una sorta di antenato del poker. Questa distinzione tra i giochi, basata sulle abilità coinvolte non è una novità di Cardano ma risale almeno al XIII secolo quando i profondi cambiamenti nella vita economica dell'Italia comunale avevano portato, tra l'altro, ad un diverso modo di guardare il variegato mondo del gioco [21]. In altri punti del *De ludo aleae*, Cardano dispensò consigli, non si sa quanto praticati, sull'opportunità di giocare in particolari condizioni (Cap. II):

deve essere tuttavia imposto un limite alla quantità di denaro.<sup>21</sup>

Si tratta di una misura cautelare per evitare perdite eccessive che già prendeva l'imperatore Augusto con i suoi famigliari. Inoltre, Cardano elencò i pericolosi annessi alla pratica del gioco d'azzardo, un triste comitato di vizi che erano in effetti il vero bersaglio della legislazione che cercava di limitare la pratica del gioco d'azzardo proprio per le derive pericolose che questo comportava. Cardano ne appuntò alcune

Il gioco muove all'ira, turba la mente e talora si giunge a litigare per i soldi, ciò che è esecrabile, pericoloso e vietato dalle leggi.<sup>22</sup> (*De ludo aleae*, Cap. II)

Il gioco non si addice a principi o prelati; occorre evitare di giocare contro i professionisti, i *lusores*; inoltre è bene giocare, prosegue Cardano, in casa propria o di amici, per evitare pubblico scandalo. Il gioco può però avere effetti positivi, a patto che che esso sia ben gestito:

tuttavia i giochi bene amministrati sono utili a rilassarsi dagli affanni ed il piacere che ne deriva ci permette di affrontare le cose serie con prontezza e alacrità.<sup>23</sup> (*De ludo aleae*, Cap. IV)

Qui Cardano sembra estendere al gioco d'azzardo un antico precetto del poeta Fedro:

Ludus animo debet aliquando dari ad cogitandum melior ut redeat tibi.

<sup>21</sup>Impositus est tamen modus, circa pecuniae quantitatem.

<sup>22</sup>Ludus iram movet, turbat mentem, et quandoque homo erumpit ad certamen pecuniae, quod turpissimum est, et periculosum, et legibus prohibitus.

<sup>23</sup>Ludi autem bene administrati, utilitates sunt relaxatio curarum, et voluptas ex quibus ad seria prompti ed alacriores surgimus.

Inoltre, se si riesce a non cadere nell'ira, soprattutto quando si perde, il gioco educa al dominio di sé:

[il gioco] è un'ottima prova per la continenza.<sup>24</sup> (*De ludo aleae*, Cap. IV)

In definitiva però, il consiglio che Cardano lascia filtrare tra le righe è quello di non praticare il gioco d'azzardo:

il più grande vantaggio dei dadi consiste nel non giocarci affatto.<sup>25</sup> (*De ludo aleae*, Cap. IV)

La presenza di aspetti positivi nel gioco d'azzardo fu il motivo che spinse Cardano a scriverne ma egli non voleva che il suo trattato potesse essere in alcun modo interpretato come una lode del gioco d'azzardo.

Nel capitolo VI iniziamo ad intravedere con chiarezza il principio guida che anima tutta l'analisi dei giochi d'azzardo di Cardano, l'*equità*. Si tratta di un concetto comprensivo che deve abbracciare tutte le componenti del gioco: non solo i dadi debbono essere *onesti* (equilibrati) ma il lancio deve essere equo grazie all'uso di un corretto supporto (il *fritillus*); occorre che vi sia equità anche tra chi assiste alla partita e tra i giocatori, intendendosi in questo caso l'*equità* come l'assenza di condizioni impari tra di essi. Infine deve esserci equità quanto al denaro in palio, il punto questo più delicato in tutti gli approcci matematici iniziali ai giochi d'azzardo:

Nei giochi d'azzardo è estremamente importante l'*equità*: dei giocatori come di chi osserva, del denaro come del luogo, degli strumenti per lanciare i dadi come dei dadi stessi. Nella misura in cui ti allontani dall'*equità* contro il tuo interesse sei stolto, a tuo vantaggio sei ingiusto.<sup>26</sup> (*De ludo aleae*, Cap. VI)

L'insistenza sull'*equità* del gioco e la necessità di trovare un punto di equilibrio rispettoso delle aspettative di tutti i giocatori rimanda alla dottrina aristotelica del giusto mezzo e non sorprende allora di trovare nel Cap. X, inframmezzata all'analisi dello spazio degli eventi relativo al gioco dei dadi, l'esposizione delle ragioni che Aristotele aveva addotto nell'*Etica Nicomachea* contro il gioco d'azzardo, da lui visto come un modo per ottenere denaro peggiore di quello dei ladri che debbono esporsi a rischi notevoli per derubare chi, forse, nemmeno conoscono personalmente, mentre i giocatori di professione rubano a persone che conoscono, quando non ad amici, e non corrono pericoli in ciò.

È nel capitolo IX che Cardano definì il concetto di *revolutio*—altre volte indicata con *circuitus*—come il numero di lanci di un dado che, teoricamente, occorre effettuare perché sia almeno possibile che ciascun risultato si presenti

<sup>24</sup>Maximum autem commodum est experimentum continentiae.

<sup>25</sup>Optima autem aleae commoditas est non ludere.

<sup>26</sup>Est autem, omnium in alea principalissimum, aequalitas, ut pote collusoris, astantium, pecuniarum, loci, fritilli, aleae ipsius. Et quantumcumque declinaveris ab ea aequalitate adversum te, stultus es, et pro te iniustus.

una volta. Si tratta dell'insieme dei casi possibili che egli determina correttamente anche per il caso di due o tre dadi, affrontati rispettivamente nei Cap. XI e XII. In particolare, ecco come spiega punteggi e casi possibili per il lancio di tre dadi:

I lanci in cui tutti i numeri sono simili coincidono, salvo per un aspetto, con quelli di prima [in cui vi erano due dadi]; pertanto sono sei; i lanci in cui due numeri sono simili e il terzo distinto sono trenta e siccome ciascun lancio può avvenire in tre modi, ve ne sono novanta; i lanci in cui tutti i numeri sono distinti sono venti e possono variare ciascuno in sei modi, per cui sono centoventi.<sup>27</sup>

Accanto a questo concetto, Cardano tornò a parlare dell'*aequalitas*, definita come la metà del numero di lanci che formano una *revolutio*. L'*aequalitas* è il perno attorno al quale ruota il *contratto* che i giocatori stringono quando fissano le regole per dividere la posta in palio:

I patti hanno luogo intorno a questa equità, se il dado è onesto; e tanto di più o di meno, quanto più ci si allontana dalla vera equità.<sup>28</sup>

L'*aequalitas* rappresenta una situazione ideale di cui Cardano stesso denunciò i limiti già al Capitolo XI, dedicato al caso in cui i dadi da lanciare siano due. Qui la *revolutio* è pari a 36 ma la *aequalitas* è diversa da punteggio a punteggio perché per le combinazioni del tipo  $(n, n)$  che si presentano una sola volta nella *revolutio*, la *aequalitas* è 18 mentre per punteggi come  $(1, 2)$ , che si presentano due volte nella *revolutio*, la *aequalitas* scende a 9. Detto questo Cardano aggiunse un inciso molto importante, a mio parere:

se più frequentemente o più raramente, dipende dalla fortuna.<sup>29</sup>

Nel capitolo IX, dove l'attenzione era basata al lancio di un solo dado, Cardano aveva introdotto la *revolutio* come il numero di lanci entro i quali tutti i numeri dovrebbero (*deberent*) comparire: il fatto che in sei lanci qualche faccia compaia più di una volta e qualche altra non compaia affatto è frutto, ancora, del caso. Cardano sembra voler dire che, attribuire una probabilità  $\frac{m}{n}$  ad un evento non garantisce che, su  $n$  tentativi, esso si presenterà *esattamente*  $m$  volte. Il valore di  $m$  ci permette però di congetturare quale sarà l'esito di una serie di lanci di dadi: a dispetto dei capricci del caso, l'*aequalitas* non è sempre lontana dal vero:

<sup>27</sup>Terna puncta similia sunt, nisi uno modo, ut in precedenti; ideoque sunt sex. Puncta vero bina similia, et tertium dispar sunt triginta; et unumquodque contingit tribus modis, erunt nonaginta. Puncta vero ex tribus dissimilibus sunt viginti, et variantur sex modis, erunt igitur cum iactu viginti, et circuitus ex omnibus ducenti sexdecim, et aequalitas in centum octo, et ponam simplices, et varios terminatos pro exemplo. Simples, ergo sex geminati cuius puncti quinque modi. Cum ergo sint puncta sex, erunt modi triginta, seu iactuum varietates. Proponitur, et variatio, triplex, ut sint novaginta; Sed viginti, qui sunt omnes dissimiles, cum variantur modis sex, erunt centum et viginti.

<sup>28</sup>Iuxta ergo hanc aequalitatem pacta constant, si alea sit justa; et tanto plus, aut minus, quanto a vera aequalitate longius distiterit.

<sup>29</sup>si frequentius, aut rarius fortunae est.

Succede però che in molti circuiti i fatti seguono la congettura da vicino.<sup>30</sup>  
(*De Ludo Aleae*, Cap. XI, p. 265)

Vediamo come Cardano applichi il concetto di *aequalitas* per decidere il corretto valore della posta da mettere in palio. Egli considerò in quanti casi compare almeno una volta il numero 1, lanciando due dadi: il risultato, peraltro già menzionato nel cap. XI, è corretto: 11. Procedette quindi al calcolo del numero di volte in cui l'1 o il 2 possono presentarsi nell'esito di un lancio di due dadi, ottenendo  $20=11+9$ ; il numero di volte in cui uno dei numeri 1, 2, 3 si presenta è pari a  $27=11+9+7$ ; e così via, ottenendo  $32 = 11 + 9 + 7 + 5$ ,  $35 = 11 + 9 + 7 + 5 + 3$  e  $36 = 11 + 9 + 7 + 5 + 3 + 1$ , se siamo interessati alle volte in cui si può presentare uno dei numeri degli insiemi  $\{1, 2, 3, 4\}$ ,  $\{1, 2, 3, 4, 5\}$  e  $\{1, 2, 3, 4, 5, 6\}$ . Ora, se si vuole scommettere sul fatto che il punteggio contenga uno dei numeri  $\{1, 2, 3\}$ , ciò che avviene in 27 dei 36 casi possibili, si può scommettere 3 contro 1: in generale, se un evento che si presenta in  $m$  degli  $n$  casi possibili, la regola da seguire per ottenere un gioco equo è basata sul rapporto tra casi favorevoli e casi contrari

$$\frac{m}{n - m} : \quad (1.4)$$

Se pertanto qualcuno dicesse: voglio uno, due o tre punti, tu sai che questi sono 27 e siccome il circuito è 36, i lanci nei quali questi punteggi non compaiono sono 9 e pertanto saranno in proporzione tripla. Pertanto, su quattro lanci, tre volte compaiono l'uno, il due od il tre e solo una volta uno degli altri, con una fortuna equilibrata; se pertanto qualcuno puntasse tre ducati sulla comparsa di uno di quei numeri e l'altro giocatore uno, il primo vincerebbe tre volte guadagnando tre ducati, il secondo vincerebbe una volta guadagnando tre ducati cosicché, in un circuito di quattro lanci si fanno sempre equilibrio. Questa è dunque la condizione per giocare equamente e se qualcuno di loro scommette di più, gareggerà in condizioni non eque e con danno, se [scommette] di meno con lucro.<sup>31</sup> (*De Ludo Aleae*, Cap. XIV)

L'equità della sorte rappresenta una situazione ideale in cui *non* solo in media ma *sempre* si hanno tre casi favorevoli su quattro. Nell'introdurre (1.4) per valutare la cifra da scommettere, Cardano sostanzialmente si serve dell'idea di probabilità come rapporto tra casi favorevoli e casi possibili, anche se non la formulò esplicitamente. Il punto matematicamente più elevato del *De Ludo Aleae* si trova nel Cap. XV, laddove Cardano vuole calcolare il rapporto (1.4)

<sup>30</sup>Attamen contingit, quod in multis circuitibus res succedit proxima conjecturae.

<sup>31</sup>Si ergo quis dicat, vellem unum; aut duo; aut tria puncta, tu scis, quod sunt haec viginti septem, et cum circuitus sit triginta sex, erunt reliqui iactus, in quibus haec puncta non evenient, novem erit igitur proportio tripla. In quatuor igitur iactibus ter eveniet aequata fortuna, punctum unum, duo; aut tria, et non nisi semel sine ullo illorum; si igitur poneret ille, qui expectat punctum unum e tribus, tres asses, alter unum vinceret prior ter, et lucraretur tres asses: alter semel et vinceret tres asses, igitur in circuitu quatuor iactuum semper aequantur. Igitur haec est ratio certandi aequali conditione, si ergo alter eorum plus ponat, certabit iniqua conditione, et cum iactura, si minus cum lucro.

se si vuole ottenere un 1 in due o tre lanci successivi di tre dadi: siccome vi sono 91 casi<sup>32</sup> in cui 1 può uscire lanciando 3 dadi, e  $125=216-91$  contrari, il rapporto (1.4) è di  $8281 = (91)^2$  contro  $15625 = (125)^2$ . Similmente, se le serie di lanci sono tre, il pronostico divente  $91^3 = 753571$  contro  $125^3 = 1953125$ . Per ottenere questi numeri Cardano ha utilizzato, di fatto, la legge della probabilità composta per eventi indipendenti: se un evento si presenta  $m$  volte su  $n$ , le probabilità che esso si presenti due o tre volte in prove indipendenti saranno  $\frac{m^2}{n^2}$  o  $\frac{m^3}{n^3}$ .

---

<sup>32</sup>Abbiamo infatti un punteggio (1, 1, 1); 15 punteggi (1, 1,  $a$ ), con  $a \neq 1$ ; 15 punteggi del tipo (1,  $a$ ,  $a$ ), sempre con  $a \neq 1$  e 60 punteggi del tipo (1,  $a$ ,  $b$ ) con  $a \neq b$ , entrambi diversi da 1.



# Bibliografia

- [1] D. Alighieri: *La Divina Commedia. Purgatorio*. Nuova Italia, Firenze, (1982).
- [2] A. Bakker, K.P.E. Gravemeijer: An historical phenomenology of mean and median. *Educational Studies in Mathematics*, **62**, (2006), 149–168.
- [3] D.R. Bellhouse: De Vetula: a Medieval manuscript containing probability calculations. *International Statistical Review*, **68**, (2000), 123–136.
- [4] D.R. Bellhouse: Decoding Cardano’s *de Ludo Aleae*. *Historia Mathematica*, **32**, (2005), 180–202.
- [5] J. Bernoulli: *Ars Conjectandi, opus posthumum*, Basel, (1713).
- [6] J. Borrel: *Logistica, quae et arithmetica a vulgo dicitur*. Guglielmum Rovillum, Lugduni, (1559).
- [7] G. Cardano: *Liber de ludo aleae*. In Hieronymi Cardani Mediolanensis Opera Omnia, vol. I., Huguetan & Ravaud, Lugduni, (1663), 262-276.
- [8] E. Czuber: *Wahrscheinlichkeitsrechnung*. I Band. Teubner, Leipzig und Berlin, (1908).
- [9] F.N. David: Studies in the history of probability and statistics. I. Dicing and gaming (A note on the history of probability). *Biometrika*, **42**, (1955), 1-15.
- [10] A.W.F. Edwards: *Pascal’s arithmetical triangle*. The Johns Hopkins University Press, Baltimore and London, (1987).
- [11] J. Franklin: *The science of conjecture: evidence and probability before Pascal*. The Johns Hopkins University Press, Baltimore and London, (2001).
- [12] G. Galilei: *Considerazione di Galileo Galilei sopra il giuoco dei dadi*. In: *Opere*, vol. I, Bettoni, Milano, 1832.
- [13] I. Hacking: *The emergence of probability*, Cambridge University Press, Cambridge (U.K.), 1975. Trad. it.: *L’emergenza della probabilità*, a cura di M. Piccone. Milano, Mondadori, 1987.

- [14] A. Hald: *History of probability and Statistics and their applications before 1750*. Wiley-Interscience, Hoboken, New-Jersey, U.S.A. (1988).
- [15] R. Ineichen: Das Problem der drei Würfel in der Vorgeschichte der Stochastik. *Elemente der Mathematik*, **42**, (1987), 69-75.
- [16] R. Ineichen: Dante-Kommentare und die Vorgeschichte der Stochastik. *Historia Mathematica*, **15**, (1988), 264-269.
- [17] R. Invrea: La legge dei grandi numeri era nota a Tucidide? *Giornale dell'Istituto Italiano degli Attuari*, **7**, (1936), 229-230.
- [18] M.G. Kendall: Studies in the history of probability and statistics. X. Where shall the history of statistics begin? *Biometrika*, **47**, (1960), 447-449.
- [19] P.-S. de Laplace: *Théorie analytique des probabilités*, II Édition. Courcier, Paris, (1814).
- [20] P. Rémond de Montmort: *Essay d'Analyse sur les jeux d'hazard*. II édition Revûe et augmentée de plusieurs Lettres. Quillau, Paris, (1713).
- [21] G. Ortalli: *Barattieri. Gioco d'azzardo fra economia ed etica nei secoli XIII-XIV*. Il Mulino, Bologna, (2012).
- [22] S. Sambursky: On the possible and the probable in ancient Greece. *Osiris*, **12**, (1956), 35-48.
- [23] O. Sheynin: Finite random sums. (A historical essay). *Archive for History of Exact Sciences*, **9**, (1973), 275-305.
- [24] O. Sheynin: On the prehistory of the theory of probability. *Archive for History of Exact Sciences*, **12**, (1974), 97-141.
- [25] O. Sheynin: The treatment of observations in early astronomy. *Archive for History of Exact Sciences*, **46**, (1993), 153-192.
- [26] E. Rubin: Quantitative commentary in Thucydides. In: *Questions & answers, The American Statistician*, **25**, (1971), 52-54.
- [27] N. Tartaglia: *General Trattato de' numeri et misure*. Curtio Trojano dei Navò, Venezia, (1556).
- [28] G. Vacca: Sul concetto di probabilità presso i Greci. *Giornale dell'Istituto Italiano degli Attuari*, **7**, (1936), 231-234.
- [29] J. van Brakel: Some remarks on the prehistory of the concept of statistical probability. *Archive for History of Exact Sciences*, **16**, (1976), 119-136.
- [30] J. von Plato: *Creating modern probability*, Cambridge University Press, Cambridge (U.K.), 1994.

- [31] R. Wilson, J.J. Watkins: *Combinatorics: ancient and modern*. Oxford University Press, Oxford (U.K.), (2013).
- [32] S.L. Zabell: Symmetry and its discontents. In: B. Skyrms, W.L. Harper, editori: *Causation, Chance, and Credence*, **1**, 155-190, (1988).



## Capitolo 2

# Il problema della divisione della posta

### 2.1 Le origini del problema della divisione della posta. I. Testi manoscritti.

Il problema della divisione della posta, o problema dei punti, è tradizionalmente associato alla nascita del calcolo delle probabilità. Nel caso in cui siano coinvolti due giocatori, esso può essere formulato in questi termini:

Due giocatori  $A$  e  $B$  si accordano nel mettere in palio una certa posta da destinare a chi per primo raggiunga  $N$  punti in un gioco. Il gioco viene però interrotto quando  $A$  ha ottenuto  $n$  punti e  $B$   $m$  punti, con  $n$  ed  $m$  entrambi inferiori ad  $N$ . Si domanda come occorra ripartire la posta in questo caso.

Per molto tempo si è ritenuto che la prima formulazione del problema fosse quella proposta da Luca Pacioli (1447-1517) nella *Summa de Arithmetica, proportioni et proportionalità* pubblicata nel 1494. In realtà, nel 1985 il ritrovamento di alcuni manoscritti ha costretto a predatare l'origine di questo problema.

Nel primo manoscritto anonimo, dal titolo *Regole del'Alzibra*, risalente alla fine del XIV secolo, l'autore tratta due problemi basati sul gioco degli scacchi. Nel primo problema, si gioca al meglio delle tre vittorie e la partita viene sospesa sul 2 a 0 per un giocatore, che diremo  $A$ , sull'altro giocatore,  $B$ . Lo schema di soluzione è molto articolato e parte dal presupposto che, quando un giocatore vince una partita, egli conquista parte del denaro dell'avversario. Così  $A$  guadagna  $c$  ducati a  $B$  vincendo la prima partita per cui, al termine di questa egli possiederà  $1 + c$  ducati, mentre  $B$  ne possiederà solo  $1 - c$ . L'incognita  $c$  rappresenta il valore della prima partita. Sul punteggio di 1 a 0 per  $A$ , l'autore del manoscritto afferma che, *per ragione*, anche la seconda partita avrà valore  $c' = c$ . Il ragionamento, ricostruito in [13], sembra basarsi su due considerazioni di simmetria. La prima, ovvia, è che il guadagno di un giocatore coincide con

quanto l'altro perde. L'altra è che il guadagno di  $A$  deve essere lo stesso del guadagno che otterrebbe  $B$ , vincendo. Ora, se  $A$  vincessesse la seconda partita, il suo guadagno arriverebbe a  $c + c'$  portandosi sul 2 a 0. Se  $B$  vincessesse, il punteggio sarebbe di 1 a 1 ed il guadagno complessivo di  $B$  sarebbe  $c' - c$ , visto che ha perduto la prima partita.<sup>1</sup> Tuttavia, sul punteggio di 1 a 1 il guadagno di  $B$  deve essere nullo e quindi  $c' = c$ . Detto questo, dopo che  $A$  ha vinto la seconda partita e si è portato sul punteggio di 2 a 0, il suo capitale totale è  $1 + 2c$ , mentre quello di  $B$  è  $1 - 2c$ . Per procedere, l'autore osserva che, se  $A$  vincessesse ancora,  $B$  perderebbe l'intera somma residua  $1 - 2c$  che rappresenta il valore di questo gioco. Se invece  $B$  vincessesse, portando il punteggio sul 2 a 1 per  $A$ , quest'ultimo possiederebbe  $4c$ , mentre  $B$  si troverebbe con  $2 - 4c$  ducati che è anche il valore della partita successiva, in quanto, se  $B$  la perdesse, perderebbe tutto. Vincendo  $B$ ,  $A$  avrebbe  $8c - 2$  ducati, mentre  $B$  avrebbe  $4 - 8c$  ducati e le due cifre debbono essere uguali perché il punteggio è di 2 a 2. Pertanto

$$4 - 8c = 8c - 2$$

da cui si ottiene  $c = \frac{3}{8}$ . Tornando al punteggio al quale si è effettivamente interrotta la partita, ad  $A$  spettano  $1 + 2c = \frac{7}{4} = 1$  ducato e  $\frac{3}{4}$ , mentre a  $B$  spetta il restante quarto di ducato. L'anonimo autore di questo manoscritto ha avuto il merito di determinare la suddivisione della posta con un metodo che procede "in avanti", partendo cioè dal punteggio iniziale di 0-0. All'anonimo autore mancò però la capacità di generalizzare il procedimento al caso di altri punteggi o al caso di più di due giocatori.

Ancor più interessante è l'altro manoscritto, datato all'inizio del XV secolo, soprattutto perché l'autore mostra di essere in possesso di un metodo *generale* che viene applicato al caso difficile in cui vi siano tre giocatori che si sfidano al meglio di tre punti di un gioco non specificato che viene sospeso quando il punteggio è di 2-1-0. Come farà Pascal, l'autore considera una soluzione "all'indietro", partendo cioè dal punteggio punteggio più semplice di 2-2-1 per discendere per gradi alla soluzione relativa al punteggio proposto. Se  $A$  e  $B$  hanno 2 punti mentre il giocatore  $C$  ne ha solo 1, i punteggi possibili dopo una nuova partita sono 3-2-1, 2-3-1 e 2-2-2. Ora, il giocatore  $C$  non vincerebbe nulla nei primi due casi mentre nell'ultimo caso gli spetterebbe  $\frac{1}{3}$  della posta, visto che tutti i giocatori sono allo stesso livello. In definitiva, pesando ugualmente i tre scenari, bisogna attribuire a  $C$   $\frac{1}{9}$  della posta. Ad  $A$  e  $B$ , che sono in condizioni simmetriche, spettano i  $\frac{4}{9}$  della posta ciascuno. L'autore passa a considerare il caso del punteggio 2-1-1. Se  $A$  vincessesse, egli otterrebbe l'intera posta mentre nei due casi restanti, che porterebbero ad uno dei punteggi 2-2-1 o 2-1-2,  $A$  si troverebbe nelle condizioni del caso precedente e dunque gli spetterebbero i  $\frac{4}{9}$  della posta. L'autore non trae le conclusioni ma esamina la situazione di  $B$  e  $C$  che non otterrebbero nulla nel primo scenario,  $\frac{4}{9}$  o  $\frac{1}{9}$  della posta negli altri due scenari, a seconda che abbiano vinto o meno la partita in oggetto. Facendo la

---

<sup>1</sup>L'autore ha commesso un errore di algebra elementare per cui il risultato finale non è corretto. La procedura è però coerente.

media delle tre frazioni egli ottiene che a  $B$  o  $C$  spettano i  $\frac{5}{27}$  della posta e, per differenza, ad  $A$  ne spettano i  $\frac{17}{27}$ .

Se invece si partisse dal punteggio di 2-2-0, un'ulteriore partita porterebbe ad uno dei risultati: 3-2-0, 2-3-0, 2-2-1: ragionando come prima, ad  $A$  e  $B$  spetterebbe la frazione  $\frac{1}{3} \left(1 + 0 + \frac{4}{9}\right) = \frac{13}{27}$  della posta, mentre a  $C$  spetterebbe il restante  $\frac{1}{27}$ . Infine, nel caso proposto all'inizio di un punteggio di 2-1-0, gli scenari possibili sono: 3-1-0, 2-2-0 e 2-1-1: ad  $A$  spetta allora la frazione

$$\frac{1}{3} \left(1 + \frac{13}{27} + \frac{17}{27}\right) = \frac{19}{27} \quad \text{della posta,}$$

a  $B$  la frazione

$$\frac{1}{3} \left(0 + \frac{13}{27} + \frac{5}{27}\right) = \frac{2}{9} \quad \text{della posta,}$$

e a  $C$  la frazione

$$\frac{1}{3} \left(0 + \frac{1}{27} + \frac{5}{27}\right) = \frac{2}{27} \quad \text{della posta.}$$

L'autore procede con l'esame di altri casi particolari che mostrano come egli fosse in possesso di un metodo generale di soluzione. Per formalizzare meglio la soluzione, indichiamo con  $e(a, b, c)$  la frazione di posta attesa da uno dei tre giocatori quando *mancano* ai giocatori  $A$ ,  $B$ ,  $C$ , rispettivamente,  $a$ ,  $b$ ,  $c$  punti per vincere la partita, lo schema esaminato equivale a risolvere l'equazione

$$e(a, b, c) = \frac{1}{3} [e(a-1, b, c) + e(a, b-1, c) + e(a, b, c-1)]$$

Sostanzialmente, il metodo coincide con quello di Pascal ed Huygens ma, in matematica come nella scienza in generale, non sempre il progresso evolve in modo uniforme nel tempo.

Ad esempio Filippo Calandri, un mestro d'abaco senese nato verso il 1467, fu autore di un trattato di Aritmetica, pubblicato nel 1491 e dedicato a Giuliano de' Medici, figlio di Lorenzo il Magnifico [13]. In un manoscritto, Calandri propose un problema in cui due giocatori si affrontano ad un gioco della *palla grossa*, un progenitore del tennis, al meglio delle sei partite. Sul punteggio di 4 a 3 per un giocatore, che diremo  $A$ , sull'avversario  $B$ , la palla per caso si buca e rende impossibile continuare il gioco. Ogni giocatore aveva scommesso 3 lire e si domanda come occorre suddividere il montepremi. Calandri espone due possibili tracce di soluzione. La prima consiste nel considerare il punteggio al quale si è arrestato il gioco e di ripartire la posta sulla base del punteggio ottenuto. La seconda soluzione, di cui Calandri fornisce i dettagli, è basata su quanto manca ai due giocatori per raggiungere l'obiettivo prefissato dei 6 punti: poiché ad  $A$  mancano due punti e a  $B$  tre, occorre che il rapporto delle somme da riscuotere sia  $\frac{3}{2}$  per cui, dal momento che una lira è formata da 20 soldi, ad  $A$  spetteranno<sup>2</sup>  $120 \times \frac{3}{5} = 72$  soldi mentre i restanti 48 spettano a  $B$ . Il commento

<sup>2</sup>Se ciascun giocatore scommette 3 lire, i soldi sono  $20 \times 3 \times 2$ .

alla soluzione tradisce però un certo disagio nel ritenerla assolutamente valida, segno delle difficoltà percepite nel fornire risposte univoche a problemi che, per loro natura, trattano di eventi aleatori:

ma perché e gl'è giuoco di fortuna non si risponde assolutamente che questo sia la verità apunto. ([2], p. 349)

In un successivo problema, Calandri considerò tre persone, che chiameremo  $A$ ,  $B$  e  $C$ , che si sfidano con la balestra per aggiudicarsi 3 denari che saranno attribuiti a chi realizzerà per primo tre centri. Quando  $A$  ha fatto due centri,  $B$  uno e  $C$  nessuno, la balestra si rompe ed è impossibile continuare il gioco. Come occorre ripartire il montepremi tra i giocatori? La soluzione di Calandri è in effetti una terza soluzione, differente da quelle introdotte prima. Calandri osserva che  $A$  ha conseguito  $\frac{2}{7}$  e  $B$   $\frac{1}{7}$  del gioco. Il numero 7 è il numero *massimo* di partite che si possono giocare e corrisponde alla situazione in cui tutti e tre i giocatori riescono ad ottenere due centri e si decide la vittoria al settimo turno. Osservato questo, la proposta di Calandri è di attribuire i  $\frac{2}{7}$  del montepremi ad  $A$ ,  $\frac{1}{7}$  a  $B$ , ripartendo tra loro che, soli, hanno conseguito dei centri, la frazione di montepremi che si ritiene già aggiudicata; i  $\frac{4}{7}$  restanti del montepremi vengono ripartiti *ugualmente* tra i tre giocatori. In definitiva, ad  $A$  spetteranno  $\frac{2}{7} \times 3 + \frac{1}{3} \times \frac{4}{7} \times 3 = \frac{10}{7}$  di ducati, ovvero un ducato e  $\frac{3}{7}$ ; a  $B$  spetteranno  $\frac{1}{7} \times 3 + \frac{1}{3} \times \frac{4}{7} \times 3 = 1$  ducato ed a  $C$  i  $\frac{4}{7}$  di un ducato.

Come si vede, tra le soluzioni esatte di fine XIV e inizio XV secolo e quelle di Calandri vi è una certa involuzione, a testimonianza che, anche in matematica, se le idee non circolano sufficientemente ed ottengono un certo consenso, la ricerca riparte sostanzialmente daccapo. Anche con l'introduzione della stampa, attorno al 1450, la situazione non migliorò subito.

## 2.2 Le origini del problema della divisione della posta. II. Testi a stampa.

Nella *Summa*, Pacioli fu molto meno attento degli scrittori anonimi che avevano trattato il problema un secolo prima e, a ben vedere, anche di Calandri. Nella *Distinctio Nona*, (X Trattato, *De extraordinariis*) Pacioli propose tre soluzioni simili tra loro e tutte scorrette, che verranno molto criticate sia da Tartaglia che da Cardano. Il primo problema considerato riguardava due squadre che si affrontano in un gioco con la palla che consta di più prove: alla squadra vincitrice di una prova sono assegnati 10 punti, nessuno alla perdente. La partita richiederebbe il raggiungimento di 60 punti ma essa è interrotta quando una squadra—diciamola squadra  $A$ —ha raggiunto 50 punti mentre l'altra—diciamola  $B$ —ne ha soltanto 20. La posta in palio è di 10 ducati e Pacioli adottò questo procedimento per deciderne la ripartizione:

- Come Calandri, Pacioli considerò il numero *massimo* di partite che le squadre potrebbero effettuare per aggiudicarsi la posta: questo numero è pari ad 11 e corrisponde alla vittoria per 60 a 50 di una squadra;

• La partita in esame è stata interrotta sul punteggio di 50 a 20, quando sono stati giocati solo 7 partite e la posta di 10 ducati deve corrispondere alla frazione  $\frac{7}{11}$  di partite già disputate. Le squadre  $A$  e  $B$  hanno vinto, rispettivamente, un numero di partite pari ai  $\frac{5}{11}$  ed ai  $\frac{2}{11}$  del numero massimo di partite disputabili. La parte  $x_A$  della posta da assegnare alla squadra  $A$  risolve la proporzione

$$10 : x_A = \frac{7}{11} : \frac{5}{11} \quad (2.1)$$

cioè  $x_A = 7$  ducati ed  $\frac{1}{7}$ , mentre alla squadra  $B$  spetta una parte  $x_B$  della posta tale che

$$10 : x_B = \frac{7}{11} : \frac{2}{11} \quad (2.2)$$

ed è dunque pari a  $x_B = 2$  ducati e  $\frac{6}{7}$ . Il numero *massimo* di partite che  $A$  e  $B$  possono giocare per aggiudicarsi l'intera posta è un dato importante nella soluzione ma nella proposta di Pacioli, se guardiamo alle proporzioni (2.1) e (2.2), esso non svolge alcuna funzione, visto che si semplifica nel corso della soluzione. Pacioli fotografa il punteggio al momento dell'interruzione del gioco e non tiene conto nella distribuzione della posta di eventuali possibili capovolgimenti di fronte che possano realizzarsi in seguito. Proprio questo aspetto fu duramente criticato dai matematici che nel XVI secolo si occuparono del problema. Prima però di presentare queste critiche e le altre soluzioni proposte, consideriamo un altro caso considerato da Pacioli, in cui tre giocatori, che diremo  $A$ ,  $B$  e  $C$ , si affrontano in un gara di tiro con la balestra. Chi ottiene il punteggio migliore in un turno, acquisisce un punto. La posta in gioco, 10 ducati, sarà accordata a chi per primo otterrà 6 punti. La competizione viene interrotta quando  $A$  ha 4 punti,  $B$  ha 3 punti e  $C$  ne ha 2. Pacioli anche qui osserva che il numero massimo di partite possibili è 16—corrispondente ad un risultato finale in cui un giocatore ha 6 punti e gli altri 5 ciascuno—per cui occorre anzitutto ridistribuire  $\frac{4}{16} = \frac{1}{4}$  della posta ad  $A$ ,  $\frac{3}{16}$  a  $B$  e  $\frac{2}{16} = \frac{1}{8}$  a  $C$ . In questo modo si assegnano 5 ducati e  $\frac{5}{8}$ ; i rimanenti 4 ducati e  $\frac{3}{8}$  si distribuiscono secondo le proporzioni  $\frac{4}{9}$ ,  $\frac{3}{9}$  e  $\frac{2}{9}$ , ovvero secondo le frazioni di successo riferite al numero di giochi effettuati, pari a 9. Pertanto  $A$  si aggiudica anche i  $\frac{4}{9}$  di 4 ducati e  $\frac{3}{8}$ , cioè 1 ducato e  $\frac{17}{18}$ ;  $B$  si aggiudica la frazione  $\frac{3}{9} = \frac{1}{3}$  di 4 ducati e  $\frac{3}{8}$ , cioè 1 ducato ed  $\frac{11}{24}$ ; infine  $C$  si aggiudica i  $\frac{2}{9}$  di 4 ducati e  $\frac{3}{8}$ , ovvero  $\frac{35}{36}$  di ducato. Pacioli in questo caso nota esplicitamente come la regola seguita non sia altro che una *regola di compagnia*, un modo cioè di ripartire la frazione di utili, o di perdite, di una attività commerciale tra i soci che contribuirono alla sua costituzione versando quote di capitale [4]:

Se deve dividere commo compagnia. ([22], a carte 197)

Vi erano in realtà diversi tipi di regole di compagnia, a seconda che tutti i soci fossero entrati nella società allo stesso momento oppure che qualcuno si fosse inserito in un momento successivo: nel primo caso si applicava la regola di compagnia semplice, nel secondo la regola di compagnia con tempo. Infine, vi erano regole con patti che servivano a regolare altri fattori che potevano

intervenire [20]. Nella regola semplice, ad esempio, dette  $q_i$  le quote versate inizialmente dagli  $n$  soci ( $i = 1, \dots, n$ ), la ripartizione del guadagno totale  $g$ , seguiva la proporzione

$$\frac{g_1}{q_1} = \frac{g_2}{q_2} = \dots = \frac{g_n}{q_n} \quad \sum_{i=1}^n g_i = g.$$

L'applicazione di questa regola al contesto della ripartizione della posta richiede un cambio di prospettiva perché le quote  $q_i$  non corrispondono a quelle versate all'inizio del gioco ma ai punteggi raggiunti dai contendenti al momento della sospensione del gioco.

Come detto in precedenza, la soluzione di Pacioli fu criticata da Tartaglia che ne mostrò l'inconsistenza nel *General trattato*. Egli osservò che, in base alla regola di Pacioli, se si fosse sospeso il gioco quando un giocatore avesse vinto una partita e l'altro nessuna, la posta sarebbe stata assegnata totalmente al primo:

La qual regola a me non pare, né bella, né buona, perché se per sorte una delle parti havesse 10 e l'altra havesse nulla, procedendo per tal sua regola seguiria, che quella parte, che havesse il detto 10 doveria tirar il tutto, e l'altra non doveria tirar cosa alcuna, che saria in tutto fuori di ragione, che per aver 10 dovesse tirar il tutto. ([33], Libro XVI, Sez. 206)

Anche se Tartaglia era piuttosto scettico sul valore matematico del problema, la cui soluzione

è più presto giudiciale, che per ragione, tal che in qual si voglia modo la sarà risolta vi si troverà da litigare ([33], Libro XVI, Sez. 206)

ne propose tuttavia una, illustrata ancora sul primo esempio numerico di Pacioli:

nondimeno il men litigioso, a me par che sia questo: prima si debba vedere, che parte ha ciascuno di tutto il gioco, che se per sorte uno avesse 10, e l'altro 0 adunque colui, che ha 10 haveria il sesto di tutto il giuoco, e per tanto dico, che in questo caso, doveria haver la sesta parte delli denari, che metteno per uno, cioè si mettono 22 ducati per parte, lui doveria haver la sesta parte di detti ducati che faria ducati 3 e doi 3 che gionti co li suoi ducati 22 fariano ducati 25 e doi 3 e l'altra parte doveva tirar il resto, il qual resto faria 18 e un 3. Et se una parte havesse 50 e l'altra 30 cava 30 di 50 resterà il venti il qual 20 vien a essere il terzo di tutto il gioco, e però doveva tirar, oltre li suoi, la terza parte delli danari dell'altra parte, la qual terza parte faria ducati 7 e un 3 che coi suoi faria ducati 29 e un 3 e l'altra parte doveria tirar il resto, che faria ducati 14 e doi 3 e così procedendo no si troverà seguir cosa non conveniente, come fece in quella di fra Luca. ([33], Libro XVI, Sez. 206)

Tartaglia non concepisce la posta come un tutto indistinto ma come un fondo in cui restano ben visibili i contributi di ciascun giocatore. Come nelle *Regole*

*del'Alzibra*, chi è in vantaggio al momento della sospensione mantiene la propria quota ed esercita il diritto di erodere una porzione della quota dell'altro giocatore. Il criterio di ripartizione è ancora basato sulla proporzionalità ma ora a contare è la differenza tra i punteggi  $s_1$  ed  $s_2 < s_1$  al momento in cui il gioco viene interrotto perché il vincitore ha diritto, per Tartaglia, alla frazione  $\frac{s_1-s_2}{s}$  dei ducati di chi è in svantaggio, essendo  $s$  il punteggio richiesto per vincere, in assenza di interruzioni. Dunque anche la proposta di Tartaglia cristallizza il risultato al momento dell'interruzione, disinteressandosi completamente degli scenari che si sarebbero potuti verificare in seguito. In effetti, la divisione della posta, a parità del valore di  $s$ —per esempio  $s = 6$  nell'esempio considerato—è la stessa sia che  $s_1 = 3$  ed  $s_2 = 0$ , sia che  $s_1 = 5$  ed  $s_2 = 2$ , che lascia ancora insoddisfatti. L'importante tuttavia per Tartaglia è fornire una regola che sia più *equa* di quella di Pacioli; per il resto egli non sembra essere molto preso dal problema, tanto che non tratta altri esempi di Pacioli:

Due altre quasi simili conseguentemente mette il detto fra Luca, le quali per esser materie di poco sugo, e di letigo assai, mi è parso di non parlarne, abenché molti hanno da caro simili facezie per aver occasione di poter contrastare.

Ribadiamo che il contesto dei problemi di Calandri, Pacioli e Tartaglia *non* è quello del gioco d'azzardo: né il gioco della palla, né quello del tiro con la balestra sono etichettabili come tali ma rientrano nei giochi di *agilità*, per usare il termine del *De Ludo Aleae*. Il solo elemento aleatorio è l'evento accidentale che impedisce di portare a termine una partita. Una soluzione simile alla terza proposta da Calandri si trova nel quinto libro della *Pratica d'arithmeticca e geometria* del padre Lorenzo Forestani da Pescia (1585-1660?), che ripropone il problema dandogli una colorita ambientazione bucolica ([7], p. 364):

Un Gentilhuomo già vecchio, ritrouandosi a vna sua Villa, e diletlandosi grandemente del giuoco di palla, chiamò due giouani Contadini, e disse, eccoui 4 ducati, giuocateli qui in mia presenza alla palla, e chi di voi prima vince 8 giuochi, voglio, che habbia vinto li 4 ducati, e così cominciarono a giocare e quando un di loro hebbe vinto 5 giuochi,<sup>3</sup> e l'altro 3. si perse la palla, e non poterono finire, e il Gentilhuomo disse, eccovi i denari, divideteli tra voi, si domanda quanti ne toccara per uno.

Forestani osserva come vi fossero *diverse opinioni* sulla risoluzione corretta del problema che egli ritiene essere la seguente: si parte ancora dal numero massimo, pari a 15, di partite necessarie perché uno dei contendenti raggiunga 8 punti. Al momento dell'interruzione si può dire che il primo giocatore abbia conseguito  $\frac{5}{15} = \frac{1}{3}$  della posta mentre il secondo giocatore ne ha conseguito  $\frac{3}{15} = \frac{1}{5}$ . Pertanto, quando il gioco viene sospeso, per Forestani è stata assegnata la frazione  $\frac{1}{3} + \frac{1}{5} = \frac{8}{15}$  della posta. Per decidere come ripartire i  $\frac{7}{15}$  ancora non assegnati, Forestani adotta la proposta salomonica già adoperata da Calandri: siccome questi

---

<sup>3</sup>Nell'originale i giochi vinti dal primo giocatore sono 6 ma deve trattarsi di un refuso tipografico.

non sono affaticati, né giocati, né vinti da nessuno di loro, (...) perciò bisogna dividerli per metà.

Al primo giocatore spettano

$$4 \times \left( \frac{1}{3} + \frac{7}{30} \right) = 4 \frac{17}{30} = 2 + \frac{4}{15} \text{ ducati}$$

mentre all'altro spettano i restanti  $\frac{13}{30}$  della posta, pari a 1 ducato e  $\frac{11}{15}$  di ducato. Come Pacioli, anche Forestani parla di divisione *per modum societatis*. Il secondo problema proposto riguarda un gioco con tre giocatori.

Tre soldati essendo dentro ad una Fortezza per la quale andando a spasso, trovarono uno scudo, e ciascun di loro lo voleva, pur alla fine s'accordarono che si dovesse giocare alle pallottole, con patto, che chi di loro vincerà prima 14 giochi habbia vinto lo scudo, cioè lire 7. Accadde che quando il primo hebbe vinto 10 giuochi il secondo 8 ed il terzo 5 gli conuenne andare in guardia; si domanda in che modo sarà dovere che dividino il detto scudo, e che parte ne toccherà a ciascuno.

Lo schema di soluzione è lo stesso di prima: si calcola il massimo numero di giochi disputabili:  $40 = 14 + 13 + 13$  e si attribuiscono in prima istanza le porzioni  $\frac{1}{4}$ ,  $\frac{1}{5}$  ed  $\frac{1}{8}$  della posta ai tre giocatori, nell'ordine. Resta così assegnata la frazione  $\frac{23}{40}$  dell'intera posta e ancora una volta, visto che la sorte può cambiare in modo imprevedibile, i restanti  $\frac{17}{40}$  sono divisi equamente fra i tre giocatori.

Criticare le soluzioni proposte da altri, Forestani conclude la sezione dedicata ai giochi con un'osservazione che ricorda lo scetticismo di Calandri e Tartaglia circa la possibilità di ottenere una soluzione ragionevole del problema

Ma perché le solutioni di simil proposte consistono nell'opinioni, e l'opinioni, e i pareri essendo varj, però lasceremo tal giudicio a più savj, e intendenti, perciò che a noi basta haver detto il parer nostro, e dimostrato questa sua contraditione. La ragione che alcuni adducono in contrario è questa, cioè, dicono, che chi ha più giuochi, è più vicino al poter finire, e conseguire il tutto, e perciò gli si convien tirare di quei danari per rata de' giuochi vinti, e noi diciamo che la Fortuna si può rivoltar presto, e favorir quell'altro a vincer il tutto; si come infinite volte s'è visto, e vedesi, tanto nel giuoco di palla, come in ogn altro, ma molto piu nelle cose di guerra, li come dottamente ne dimostra l'Ariosto<sup>4</sup> in persona di Carlo con questi due versi.

Così Fortuna ad Agramante arrise,

Ch'un'altra volta a Carlo assedio mise.

il qual'havendo assediato Agramente, si rivoltò talmente la fortuna, che Agramarite in un attimo ruppe l'esercito di Carlo, e nuovamente l'assedio in Parigi.

<sup>4</sup>La citazione che segue si trova nel Canto XXVII dell'*Orlando Furioso*.

Forestani ha considerato esempi in cui la posta in gioco è o offerta da un mecenate oppure è trovata per caso dal gruppo dei giocatori: anche quando riferisce l'esempio di giocatori che *giocano al Tavoliere*, precisa subito che i denari in palio erano stati *trovati a caso in una borsa*, quasi a voler moralizzare—Forestani era un sacerdote—gli esempi proposti, mostrando come la competizione fosse per aggiudicarsi qualcosa che avrebbe arricchito qualcuno senza defraudare altri, a differenza dei giochi d'azzardo che portavano spesso sul lastrico i giocatori più accaniti.

## 2.3 Cardano e la divisione della posta

Come già fatto in precedenza, riserviamo uno spazio a sé stante per Cardano che si occupò del problema della divisione della posta nella *Practica Arithmeticae et misurandi singularis* [3], pubblicata nel 1539 e dunque prima sia del *General Trattato* di Tartaglia che della *Pratica* di Forestani. A dispetto del loro interesse, le considerazioni di Cardano sono però rimaste piuttosto nell'ombra, come vedremo in chiusura di questa sezione. L'ultimo dei 68 capitoli della *Practica Arithmeticae* contiene un lungo elenco di errori commessi da Pacioli. Tra questi spicca l'errore 5:

Nella determinazione dei giochi [Pacioli] commise un errore molto evidente, che anche un bambino riconoscerebbe, mentre egli accusa gli altri e loda la sua eccellente opinione in base alla quale a due giocatori che giocano per arrivare a 6 punti, conferisce, dopo molte considerazioni superflue, a chi ne ha 5 e a chi ne ha 2 le parti 5 e 2, dividendo la somma totale in 7. Supponiamo pertanto che due giochino per arrivare a 19 punti e che uno ne abbia 18 mentre l'altro solo nove; il primo ottiene  $\frac{2}{3}$  della somma totale ed il secondo  $\frac{1}{3}$  per cui se il deposito di ciascuno è 12 ducati, la somma di entrambi è 24, 16 dei quali toccheranno al primo ed 8 al secondo: pertanto chi è giunto a 18 punti avrà guadagnato all'altro solo 4 ducati, che sono la terza parte del deposito, anche se non gli manca che un punto a raggiungere il traguardo mentre all'altro ne mancano 10, ciò che è completamente assurdo, soprattutto perché ognuno deve prendere quella parte che potrebbe scommettere in quella condizione. Trovandosi sul punteggio di 18 a 9 e dovendo arrivare a 19, può scommettere 10 contro 1 o addirittura 20 contro 1; nella divisione pertanto egli deve avere 20 parti e l'altro una sola. In terzo luogo, se si gioca a 19 [punti] ed uno ne ha 2, l'altro zero, dal suo [di Pacioli] calcolo, chi ha due punti deve ottenere tutto il deposito e l'altro nulla, la qual cosa non c'è da dubitare sia sconveniente, che cioè pur trovandosi ancora molto distante dalla fine, debba ottenere tanto quanto otterrebbe se avesse 19 punti, pur sopravanzando l'altro giocatore di così poco.<sup>5</sup>

<sup>5</sup>Et errauit ludorum determinatione errore, manifestissimo, et a puero etiam cognoscibilis, dum alios arguit et suam laudat exquisitam opinionem; unde ludentibus ad 6. et habenti 5. et alteri 2. dat post multas superfluas supputationes partes 5 et 2 ita quod totam summam

Al di là di questa veemente *pars destruens*, nel capitolo 61 della *Practica—De Extraordinariis et ludis*—Cardano si era dilungato su quali dovessero essere i principii cui mantenere fede nella divisione della posta in un gioco interrotto:

Quanto ai fondamenti dei giochi bisogna sapere che non occorre considerare altro se non il fine verso cui si tende e ciò dividendo il tutto in progressione per le loro parti[.] esempio [:] due giocano a dieci [punti:] uno ne ha 7 l'altro 9[.] Si domanda, qualora occorra dividere la posta perché il gioco non può terminare, quanto ciascuno debba avere dividendo l'intero deposito[.] Sottrai 7 da 10: resta 3. Sottrai 9 da 10: resta 1. La progressione di 3 è 6, quella di 1 è 1. Dunque, diviso il deposito in 7 parti darai 6 parti a chi ha 9 punti ed 1 a chi ne ha 7. Supponiamo dunque che ciascuno avesse scommesso 7 denari, cosicché il deposito complessivo ammonti a 14 denari: 12 di questi toccheranno a chi ha 9 punti e due a chi ha 7 punti, per cui chi ha 7 punti perde i  $\frac{5}{7}$  del suo capitale. Un altro esempio. Supponiamo di giocare al meglio dei 10 punti e che uno abbia 3 punti, l'altro 6. Sottraendo [da 10] si hanno i resti 7 e 4[:] la progressione del 7 è 28, quella del 4 è 10. Pertanto darò 28 parti della somma totale a chi ha 6 punti e 10 parti a chi ne ha 3 dividendo così il deposito in 38 parti, e chi ha 3 punti perderà i  $\frac{9}{19}$  del proprio capitale.<sup>6</sup> ([3], p. 112)

Cardano ha colto il punto cruciale del problema: solo i punti *mancanti* al conseguimento del successo sono essenziali nella suddivisione della posta. Il testo diventa però ellittico proprio quando passa agli esempi numerici ed il precetto di costruire la *progressione* di 3 o di 1, di 7 o di 4, lascia perplessi ad una prima lettura. Ora, dicendo che la progressione di 3 è 6, di 1 è 1, di 7 è 28 e di 4 è 10, Cardano associa al numero di punti che mancano ad un giocatore per vincere

---

dividit in 7. ponamus igitur quod duos ludant at 19 et unus habeat 18 alius tantum 9 habet igitur primo  $\frac{2}{3}$  totius summae et secundo  $\frac{1}{3}$ , sit igitur depositum aurei 12 summae amborum erit 24. quibus 16 primo et 8 secundo contingent: non igitur ille qui habet 18 ludos lucratus est nisi aureos 4, et ex adversario, qui sunt tertia pars depositi, et tam ad complendum non deest nisi unus ludus, secundo autem desunt 10 hoc autem est absurdissimum praeterea illam partem quisque debet assumere, quam aequa ratione deponere posset ea conditione, sed habens 18 cum habente 9 potest eundo 19 deponere 10 contra 1 imo 20 contra unum: igitur in divisione debet habere partes 20 et ille tantum unam, tertio si ludimus ad 19 et unus habeat 2 alter nullum, per suam rationem qui habet 2 debet acquirere totum depositum, patet ex suo computo, hoc autem quale sit inconueniens non est dubitandum, cum ex tam modica superatione, cum tanta remotione a fine debeat acquirere tantum, quantum si lucratus fuisset 19 ludos.

<sup>6</sup>Quantum ad rationem ludorum sciendum est quod in ludis non habet considerari nisi terminus ad quem et hoc in progressione dividendo totum per easdem partes[.] exemplum[:] duo ludunt ad decem unus habet 7 alius 9 quaeritur in casu divisionis non finiendo ludum quantum quisque debet habere subtrahe 7 a 10 remanet 3. Subtrahe 9 a 10 remanet 1. Progressio 3 est 6, progressio 1 est 1. dabis igitur dividendo totum depositum in 7 parte 6 partes habenti 9 et 1 partem habenti 7. ponamus igitur quod posuissent aureos 7 singuli, tunc totum depositum esset 14. ex quibus 12 contingunt habenti 9 et 2 habenti 7 ludos, quare qui habet 7 perdit  $\frac{5}{7}$  capitalis. Aliud exemplum ponamus quod ludus sit ad 10 et unus habeat 3 alius 6. subtrahe fiunt residua 7 et 4 progressio 7 est 28 progressio 4 est 10. igitur totius summae dabo habenti 6 ludos 28 partes, et habenti 3 dabo partes 10 et ita dividam totum depositum in 38 partes, et ille qui habet 3 perdit  $\frac{9}{19}$  sui capitalis.

la *somma* di tutti gli interi da uno al numero di punti mancanti. Perché questa scelta? Procediamo ancora nella lettura della *Practica*:

La motivazione che sta dietro questa [regola] è che se, effettuata la divisione, si volesse ricominciare il gioco, le parti dovrebbero giocarsi quanto riceverebbero nel momento dell'interruzione e ponendoci nel primo esempio chi dice di voler giocare ponga la condizione che tu non possa vincere se non vinci 3 partite di seguito mentre a lui ne basta vincere una sola. Supponiamo che chi vuol vincere tre partite scommetta 2 denari: dico che l'altro deve scommettere 12 denari. Il motivo è che, se giocassero al meglio di un punto basterebbe scommettere 2 denari e se al meglio dei due punti dovrebbe scommettere il triplo perché vincendo semplicemente una partita vincerebbe 4 denari e questo va bilanciato dal pericolo di perdere la seconda partita, vinta la prima, pertanto deve guadagnare il primo e se si gioca al meglio dei tre punti [deve scommettere] il sestuplo, perché raddoppia la difficoltà. Pertanto dovrà giocare 12 denari e ora se accetta 12 denari e l'altro 2, la divisione sarà fatta in modo conveniente.<sup>7</sup> ([3], p. 112)

La chiave di lettura sta nel riferimento alla *ripresa* del gioco dopo l'interruzione, senza che i punteggi siano azzerati. Come occorre regolare le scommesse per convincere un giocatore a riprendere il gioco dal punto in cui è stato interrotto? Abbiamo visto la critica a Pacioli: se, occorrendo 19 punti per vincere, il gioco fosse interrotto sul punteggio di 18 a 9, il giocatore che volesse riprendere a giocare dalla posizione di vantaggio dovrebbe puntare 10 ad 1 o anche fino a 20 ad 1, tale è il vantaggio. Coumet [4] ha notato una incoerenza tra la prescrizione generale di Cardano del Cap. 61 e la discussione legata all'errore di Pacioli, probabilmente dovuta alla *vis* polemica nei suoi confronti. Nel passo appena riportato si assiste [4] ad una trasformazione della regola di divisione della posta (*règle du partis*) in una regola sulle scommesse (*règle du paris*) da stabilire avendo di mira sempre la condizione di equità del gioco. Cardano in effetti, per giustificare la ripartizione di 6 ad 1 proposta nel primo esempio del Cap. 61, analizza un altro problema, quello di due giocatori che *iniziano* a giocare sotto la condizione che il primo si aggiudicherà la posta se vincerà consecutivamente tre partite, l'altro se ne vincerà una sola. Il principio assunto da Cardano non è lontano da quello che animerà Fermat e Pascal nel risolvere correttamente il problema. Egli ritiene che occorra *compensare* le minori possibilità di vittoria di un giocatore costringendo l'altro a scommettere una somma superiore, che

<sup>7</sup>§ 14. Ratio autem demonstrativa super hoc est quod si facta divisione iterum ludus esset inchoandus, partes haberent deponere idem quod receperunt stante conditione, et sit in exemplo primo quod quis dicat volo ludere, hac conditione ut tu non possis vincere nisi vincas 3 sine intermissione, et si ego vinco unum volo vincere, et deponat ille qui vult vincere 3 ludos aureos 2 quantum habet deponere alius dico quod deponet 12 ratio nam si ad unum ludum haberent luderet sufficeret ponere 2 et si duos, habere ponere triplum, ratio quia vincendo simpliciter 1 ludos vinceret 4 sed hic stat cum periculo perdendi secundum victo primo, igitur lucrari debet primum, et si ad 3 sextuplam, quia duplicatur difficultas, igitur haberet ponere 12 et iam accepto 12 et ille 2 igitur divisio fuit convenienter facta.

ristabilisca l'equità, bilanciando il rischio che il primo giocatore affronta di veder vanificati i successi parziali [4]. Così, se volessi rimpiazzare, nel primo esempio numerico, il giocatore cui mancano 3 punti per vincere, poiché al mio avversario manca un punto solo, la difficoltà crescente a vincere ininterrottamente tre partite richiede che, se io punto una moneta, anche il mio avversario ne punti una, che basterebbe se fossimo in condizioni di parità; più 2 monete, che basterebbe, insieme al precedente, se a me mancassero 2 punti; più 3 monete. Nel caso di sospensione sul punteggio di 18 a 9, quando il punteggio richiesto per vincere è di 19 punti, la regola di Cardano imporrebbe che il giocatore in svantaggio puntasse 1 moneta quando l'altro ne puntasse 55, la somma dei primi  $10=19-9$  interi.

Il testo di Cardano fu preso a modello dal matematico cuneese Giovanni Francesco Peverone (1509-1559) in un opuscolo [16] pubblicato nel 1558 e dedicato alla risoluzione di problemi di aritmetica e geometria. Gli esempi discussi da Peverone nel libro terzo, in un breve paragrafo intitolato *De giuochi*, coincidono con quelli discussi da Cardano nella *Practica arithmeticae*, anche nei valori numerici. Studiando questa parte della storia della probabilità, Maurice Kendall (1907-1983) attribuì a Peverone i risultati di Cardano [10], il cui contributo fu invece valorizzato da Coumet che osservò anche come Peverone non avesse inserito la spiegazione della regola fornita da Cardano, possibile sintomo di una comprensione parziale del metodo.

## 2.4 La corrispondenza tra Fermat e Pascal

Pierre de Fermat (1601-1665) e Blaise Pascal (1623-1662) furono due tra i più grandi matematici francesi del XVII secolo. La loro corrispondenza dell'estate 1654 è tradizionalmente presa come data di nascita del calcolo delle probabilità come disciplina scientifica. La corrispondenza è rimasta inedita fino alla pubblicazione della prima edizione delle *Opere Matematiche* di Fermat, nel 1679. Fortunatamente però, il contenuto della corrispondenza fu noto subito allo scienziato olandese Christiaan Huygens (1629-1695) durante un viaggio di studio a Parigi ed egli seppe trarre profitto delle indicazioni ricevute sul contenuto dello scambio epistolare, ancorché incomplete. Esaminiamo i passaggi relativi al problema della ripartizione della posta, partendo dalla lettera di Pascal del 29 luglio 1654 dove egli espose il proprio modo di risolvere il problema che differisce da quello di Fermat, basato totalmente sulle *combinazioni* e dunque da intendersi come l'elencazione di tutti i casi che possono presentarsi. Pascal considerò il caso in cui due giocatori si affrontano al meglio di tre partite, scommettendo ciascuno 32 pistole<sup>8</sup> ed introdusse una serie di esempi dalla struttura ricorsiva, costruita a partire dal caso più semplice, quello in cui uno dei due giocatori— $G_1$ —ha conquistato due punti e l'altro— $G_2$ —uno. Pascal esaminò gli esiti possibili della

---

<sup>8</sup>Il termine *pistola*, derivato da *piastola*, piccola piastra, era usato per indicare una moneta d'oro coniatata per la prima volta in Spagna verso la metà del XVI secolo. La parola fu poi impiegata in senso estensivo per indicare anche altre monete d'oro europee, di valore simile a quella spagnola. Equivalenza a due escudi e per questo era detta anche *doblone*.

partita successiva: se  $G_1$  vincessesse, si aggiudicherebbe l'intera posta di 64 pistole; se vincessesse  $G_2$  ci sarebbe parità e dunque dovrebbero riprendersi ciascuno 32 pistole. Qualunque cosa succeda,  $G_1$  si è garantito 32 pistole ma, poiché Pascal ipotizzò che i due giocatori abbiano le stesse possibilità di vincere, una volta in parità, ad  $A$  spetterà anche la metà delle 32 pistole di  $G_2$  e dunque potrà chiedere 48 pistole.

Ecco all'incirca come faccio a trovare il valore di ciascuna delle parti quando, per esempio, due giocatori giocano, al meglio delle tre partite e ciascuno ha scommesso 32 pistole: supponiamo che il primo abbia due punti e l'altro uno; ora giocano un'altra partita il cui esito sarà che: se vince il primo, otterrà tutta la posta, cioè 64 pistole; se vince l'altro, avranno vinto ciascuno due partite e di conseguenza, se volessero separarsi, occorre che ciascuno si riprenda la somma scommessa: 32 pistole.

Considerate dunque, signore, che, se il primo vince, otterrà 64, se perde otterrà 32. Dunque se non vogliono azzardare questa partita e separarsi senza giocare, il primo deve dire; "sono sicuro di avere 32 pistole, perché anche la sconfitta me le lascia; per le altre 32, potrò averle o meno, le possibilità sono uguali. Dividiamo dunque queste 32 pistole a metà e mi darete oltre a queste, le 32 che mi sono assicurato." Egli avrà dunque 48 pistole e l'altro 16.<sup>9</sup> ([5], p. 291)

Pascal procedette ad esaminare il caso in cui la sospensione della partita avvenisse quando  $G_1$  ha due punti e  $G_2$  zero, sempre a parità di montepremi totale: 64 pistole. Il ragionamento è questo: se  $G_1$  vincessesse ancora, si aggiudicherebbe il terzo punto decisivo e dunque 64 pistole; se  $G_2$  vincessesse il punteggio sarebbe di due punti contro uno e dunque, per quanto appena visto, a  $G_1$  spetterebbero 48 pistole. Per decidere quanto assegnare a  $G_1$ , Pascal parte dalle 48 che gli sono garantite e vi aggiunge la metà della differenza tra le due possibili vincite di 64 e 48 pistole, cioè 8 pistole, portando la posta di  $G_1$  a 56 pistole.<sup>10</sup> Per concludere, Pascal analizzò il caso in cui la sospensione avvenga quando  $G_1$

<sup>9</sup>Voici à peu près comme je fais pour savoir la valeur de chacune des parties, quand deux joueurs jouent, par exemple, en trois parties, et chacun a mis 32 pistoles au jeu: Posons que le premier en ait deux et l'autre une; ils jouent maintenant une partie, dont le sort est tel que, si le premier la gagne, il gagne tout l'argent qui est au jeu, savoir 64 pistoles; si l'autre la gagne, ils sont deux parties à deux parties, et par conséquent, s'ils veulent se séparer, il faut qu'ils retirent chacun leur mise, savoir chacun 32 pistoles.

Considérez donc, Monsieur, que, si le premier gagne, il lui appartient 64; s'il perd, il lui appartient 32. Donc, s'ils veulent ne point hasarder cette partie et se séparer sans la jouer, le premier doit dire: "Je suis sur d'avoir 32 pistoles, car la perte même me les donne; mais pour les 32 autres, peut-être je les aurai, peut-être vous les aurez, le hasard est égal. Partageons donc ces 32 pistoles par la moitié et me donnez, outre cela, mes 32 qui me sont sûres." Il aura donc 48 pistoles et l'autre 16.

<sup>10</sup>Ecco l'originale di Pascal:

Posons maintenant que le premier ait *deux* parties et l'autre *point*, et ils commencent à jouer une partie. Le sort de cette partie est tel que, si le premier la gagne, il tire tout l'argent, 64 pistoles; si l'autre la gagne, les voilà revenus au cas précédent, auquel le premier aura deux parties et l'autre une. Or, nous avons déjà montré qu'en ce cas il appartient, à celui qui a les deux parties, 48 pistoles: donc, s'ils veulent ne point jouer cette partie, il doit dire ainsi: "Si je la

ha un punto soltanto e  $G_2$  nessun punto, ancora a parità di montepremi. Se  $G_1$  vincessero l'ipotetica partita successiva, si porterebbe a 2 punti contro 0 e dunque gli spetterebbero 56 pistole mentre, se  $G_2$  vincessero, si porterebbe in parità e ad entrambi i giocatori spetterebbero 32 pistole. Dunque, in entrambi gli scenari 32 pistole sono garantite a  $G_1$  mentre egli può giungere o meno a 56: se l'esito della partita è incerto, allora è lecito che  $G_1$  chieda altre  $\frac{56-32}{2} = 12$  pistole e reclamarne per sé 44 in tutto.<sup>11</sup>

Come sempre, trattandosi di un procedimento ricorsivo, occorre un metodo che consenta di arrivare alla conclusione senza dover ripetere tutti i passaggi e Pascal espone la propria regola, senza fornirne in questa sede la dimostrazione. Indicando con  $P_1(a, b)$  la frazione di posta che spetta a  $G_1$  quando gli mancano  $a$  punti per conseguire la vittoria, mentre a  $G_2$  ne mancano  $b$ , la soluzione di Pascal si formalizza [5] con l'equazione alle differenze parziali

$$P_1(a, b) = \frac{1}{2} [P_1(a-1, b) + P_1(a, b-1)] . \quad (2.3)$$

Inoltre  $P_1$  assume alcuni valori particolari

$$P_1(a, 0) = 0 \quad (2.4)$$

e

$$P_1(a, a) = \frac{1}{2}. \quad (2.5)$$

Da (2.3) e (2.4) si può ricavare, come utile esercizio,

$$P_1(a, 1) = \frac{1}{2^a}.$$

Fermat scrisse il successivo 9 agosto a Pierre de Carcavi che, dopo la morte di padre Marin Mersenne, svolgeva il delicato compito di intermediario nella corrispondenza tra i principali studiosi dell'epoca, manifestando tutta la propria stima per il genio di Pascal, in grado di portare a termine felicemente ogni impresa scientifica cui intendesse applicarsi. Verso la fine del mese di agosto, il giorno 24, Pascal scrisse nuovamente a Fermat ribadendo come il metodo

---

gagne, je gagnerai tout, qui est 64; si je la perds, il m'appartiendra légitimement 48: donc donnez-moi les 48 qui me sont certaines, au cas même que je perde, et partageons les 16 autres par la moitié, puisqu'il y a autant de hasard que vous les gagniez comme moi." Ainsi il aura 48 et 8, qui sont 56 pistoles.

<sup>11</sup>Questo è l'originale di Pascal:

Posons enfin que le premier n'ait que *une* partie et l'autre *point*. Vous voyez, Monsieur, que, s'ils commencent une partie nouvelle, le sort en est tel que, si le premier la gagne, il aura *deux* parties à *point*, et partant, par le cas précédent, il lui appartient 56; s'il la perd, ils sont partie à partie: donc il lui appartient 32 pistoles. Donc il doit dire: "Si vous voulez ne la pas jouer, donnez-moi 32 pistoles qui me sont sûres, et partageons le reste de 56 par la moitié. De 56 ôtez 32, reste 24; partagez donc 24 par la moitié, prenez-en 12 et moi 12, qui, avec 32, font 44". ([5], p. 291)

delle combinazioni fosse non solo malagevole ma offrì risposte ambigue nel caso in cui fossero coinvolti tre giocatori. La lettera di Pascal è preziosa perché permette di ricostruire gli argomenti utilizzati da Fermat. Si parte ancora da un caso particolare, quello in cui a  $G_1$  mancano due punti per vincere e a  $G_2$  ne mancano 3. Il numero massimo di partite per decidere della vittoria di uno dei contendenti è 4. Pascal considerò allora i  $16 = 2^4$  esiti possibili, ottenuti enumerando tutte le ipotetiche successioni di quattro partite. Indicando con  $a$  e  $b$  rispettivamente un successo di  $G_1$  o  $G_2$ , i 16 schemi sono i seguenti

$a$	$a$	$a$	$a$	$a$	$a$	$a$	$a$	$a$	$b$	$b$	$b$	$b$	$b$	$b$	$b$	$b$
$a$	$a$	$a$	$a$	$b$	$b$	$b$	$b$	$a$	$a$	$a$	$a$	$b$	$b$	$b$	$b$	$b$
$a$	$a$	$b$	$b$	$a$	$a$	$b$	$b$	$a$	$a$	$b$	$b$	$a$	$a$	$b$	$b$	$b$
$a$	$b$	$a$	$b$	$a$	$b$	$a$	$b$	$a$	$b$	$a$	$b$	$a$	$b$	$a$	$b$	$b$
1	1	1	1	1	1	1	2	1	1	1	2	1	2	2	2	2

dove i numeri 1 e 2 indicano che il successo finale spetta a  $G_1$  o a  $G_2$ , rispettivamente. Poiché 11 schemi sono favorevoli a  $G_1$  e 5 a  $G_2$ , la posta in palio va ripartita nel rapporto di 11 : 5. Il metodo di Fermat è diverso da quello di Pascal ma si può verificare che entrambi danno le stesse risposte, almeno nel caso in cui ci sono due giocatori. Pascal riferì anche delle obiezioni sollevate contro il metodo di Fermat da Gilles Personne de Roberval (1602-1675) che contestava l'uso di partite fittizie per determinare la vera ripartizione della posta. Sostanzialmente, osservava Roberval, non è possibile fingere che tutte le partite abbiano ugual durata perché, ad esempio, nei primi quattro schemi, partendo da sinistra, ci si arresta dopo due sole partite. Pascal, osservando che l'obiezione riguardava solo il metodo di Fermat, rispose mostrando come l'artificio di introdurre delle partite fittizie non alterasse affatto l'esito della suddivisione della posta ([5], pp. 302-303).

Ho comunicato il vostro metodo ai nostri Signori e su di esso il Sig. de Roberval ha mosso questa obiezione:

È sbagliato basare la suddivisione della posta sull'ipotesi che si giochi in *quattro* partite, visto che, quando mancano *due* partite ad un giocatore e tre all'altro, non è affatto necessario giocare *quattro*, potendo succedere che non se ne giochi che *due* o *tre* e forse *quattro*.

E così per questo non vede perché si pretenda di fare la suddivisione corretta partendo dalla condizione fittizia che si giochino quattro partite, visto che la condizione naturale del gioco è che si smetterà di giocare dopo che uno dei giocatori avrà vinto e che [il metodo], se non è falso, almeno non è dimostrato, per cui egli sospetta che abbiamo commesso qualche paralogismo.

Gli risposi che io non mi fondavo tanto su questo metodo di combinazioni, che in verità è fuori luogo in questa occasione, quanto su un altro mio metodo universale, che non ammette eccezioni, che porta con sé la dimostrazione e che trova esattamente la stessa suddivisione del metodo delle combinazioni; inoltre gli ho mostrato la correttezza della suddivisione tramite combinazioni in questo modo:

Non è forse vero che due giocatori, trovandosi nella situazione ipotizzata in cui manchino due punti ad uno e tre all'altro, convengano di comune accordo che giocheranno quattro partite complete (...) non è forse vero, dissi, che deliberando di giocare quattro partite, la suddivisione della posta deve essere quella che abbiamo detto, secondo il numero di schemi favorevoli a ciascuno?

Rimase d'accordo e questo è in effetti conclusivo; negò tuttavia che la stessa cosa sussisterebbe non costringendoli a giocare le *quattro* partite. Gli dissi allora:

Non è chiaro che gli stessi giocatori, non essendo costretti a giocare quattro partite, ma volendo terminare il gioco dopo che uno ha raggiunto la vittoria, potranno vincolarsi, senza danno o lucro, a giocare tutte le quattro partite e che questa convenzione non cambia in alcun modo la loro condizione?<sup>12</sup>

Sono stati espressi giudizi molto severi su Roberval per questo errore ma dobbiamo essere consapevoli che anche lui, come Fermat e Pascal, si stava muovendo in terra incognita e che il metodo delle combinazioni, di cui Pascal sembra in certi punti fare una difesa d'ufficio, fornisce risposte corrette solo nel caso di due giocatori mentre, passando a tre giocatori, offre il fianco a possibili ambiguità, che Pascal mette in evidenza su un caso particolare in cui tre giocatori— $G_1$ ,  $G_2$ ,  $G_3$ — smettono di giocare quando mancano loro, rispettivamente 1, 2 e 2 punti. La partita sarebbe decisa in al più tre giochi ma lo schema dei  $27 = 3^3$

---

<sup>12</sup>Je communiquai votre méthode à nos Messieurs, sur quoi M. de Roberval me fit cette objection:

Que c'est à tort que l'on prend l'art de faire le parti sur la supposition qu'on joue en *quatre* parties, vu que, quand il manque *deux* parties à l'un et *trois* à l'autre, il n'est pas de nécessité que l'on joue *quatre* parties, pouvant arriver qu'on n'en jouera que *deux* ou *trois*, ou à la vérité peut-être *quatre*;

Et ainsi qu'il ne voyoit pas pourquoi on prétendoit de faire le parti juste sur une condition feinte qu'on jouera quatre parties, vu que la condition naturelle du jeu est qu'on ne jouera plus dès que l'un des joueurs aura gagné, et qu'au moins, si cela n'étoit faux, cela n'étoit pas démontré, de sorte qu'il avoit quelque soupçon que nous avions fait un paralogisme.

Je lui répondis que je ne me fondois pas tant sur cette méthode des combinaisons, la quelle véritablement n'est pas en son lieu en cette occasion, comme sur mon autre méthode universelle, à qui rien n'échappe et qui porte sa démonstration avec soi, qui trouve le même parti précisément que celle des combinaisons; et de plus je lui démontrai la vérité du parti entre deux joueurs par les combinaisons en cette sorte:

N'est-il pas vrai que, si deux joueurs, se trouvant en cet état de l'hypothèse qu'il manque deux parties à l'un et trois à l'autre, conviennent maintenant de gré à gré qu'on joue quatre parties complètes, c'est-à-dire qu'on jette les quatre dés à deux faces tous à la fois, n'est il pas vrai, dis-je, que, s'ils ont délibéré de jouer les quatre parties, le parti doit être, tel que nous avons dit, suivant la multitude des assiettes favorables à chacun?

Il en demeura d'accord et cela en effet est démonstratif; mais il noit que la même chose subsistât en ne s'astreignant pas à jouer les *quatre* parties. Je lui dis donc ainsi:

N'est-il pas clair que les mêmes joueurs, n'étant pas astreints à jouer [les] quatre parties, mais voulant quitter le jeu dès que l'un auroit atteint son nombre, peuvent sans dommage ni avantage s'astreindre à jouer les quatre parties entières et que cette convention ne change en aucune manière leur condition?

risultati possibili richiede qualche cautela.

a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	c	c	c	c	c	c	c	c	c	c
a	a	a	b	b	b	c	c	c	a	a	a	b	b	b	c	c	c	a	a	a	b	b	b	c	c	c	c	c
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
				2						2		2	2	2		2			3		3		2		3	3	3	3

Infatti, mentre alcuni schemi forniscono univocamente un vincitore, in altri casi i risultati delle tre partite fittizie mostrano che due giocatori raggiungono il punteggio che li renderebbe vincitori. Si potrebbero prospettare due strategie: attribuire lo schema corrispondente metà a ciascuno dei vincitori ipotetici, oppure al giocatore che raggiunge per primo il successo. Nel primo caso, ad esempio,  $G_1$ , cui corrisponde il numero 1, si aggiudica senza dubbio 13 dei 27 schemi nel prospetto precedente mentre in altri 6 schemi è vincitore con uno degli altri giocatori ed negli 8 schemi restanti è certamente perdente. Se ogni schema assegnato *solo* a  $G_1$  vale 1, quelli in cui vince in coabitazione con  $G_2$  o  $G_3$  gli conta  $\frac{1}{2}$  e quelli in cui è perdente non fruttano nulla, gli schemi attribuiti a  $G_1$  sono

$$1 \cdot 13 + \frac{1}{2} \cdot 6 + 0 \cdot 8 = 16$$

e la frazione di posta che gli spetta è  $\frac{16}{27}$ . Ragionando similmente, ad ognuno dei giocatori  $G_2$  e  $G_3$  bisogna assegnare

$$1 \cdot 4 + \frac{1}{2} \cdot 3 + 0 \cdot 20 = \frac{11}{2}$$

schemi e quindi la frazione  $\frac{11}{54}$  della posta. Pascal è però insoddisfatto del risultato ottenuto:

Ecco, mi sembra, il modo in cui si dovrebbe suddividere la posta tramite le combinazioni secondo il vostro metodo, a meno che voi non abbiate qualcos'altro su questo metodo che io ignoro. Se non mi sbaglio, questa suddivisione è ingiusta. Il motivo è che si suppone il falso sostenendo che si giocheranno sicuramente tre partite, mentre la condizione naturale del gioco è che si prosegue a giocare finché uno dei giocatori non abbia ottenuto il numero di punti che gli mancano, nel qual caso il gioco cessa.<sup>13</sup>

L'obiezione di Roberval, ininfluenza nel caso di due giocatori, è più calzante in questa situazione più complessa e l'analisi di Pascal mostra che l'utilizzo del metodo delle combinazioni richieda ora uno studio dell'*ordine* con cui si susseguono le vittorie per decidere come ripartire la posta. Il gioco fittizio (*feinte*) che impone di disputare sempre il numero massimo di partite, non si accorda con quello vero (*véritable*) che nel caso di due giocatori. La risposta

<sup>13</sup>Voilà, ce me semble, de quelle manières il faudroit faire les partis par les combinaisons suivant votre méthode, si ce n'est que vous ayez quelque autre chose sur ce sujet que je ne puis savoir. Mais, si je ne me trompe, ce parti est mal juste. La raison en est qu'on suppose une chose fausse, qui est qu'on joue en trois parties infailliblement, au lieu que la condition naturelle de ce jeu-là est qu'on ne joue que jusques à ce qu'un des joueurs ait atteint le nombre de parties qui lui manque, auquel cas le jeu cesse.

di Fermat fu scritta il successivo 29 agosto: egli riconobbe come corrette le limitazioni messe in luce da Pascal ma difese il proprio metodo, purché sia bene interpretato, cosa che Pascal non aveva fatto nella lettera precedente.

Oltre al problema della ripartizione della posta, nella corrispondenza tra Pascal e Fermat si parla di alcuni questiti che il cavaliere de Méré pose a Pascal. Scrivendo a Fermat nel luglio del 1654, Pascal si espresse in questi termini:

Non ho tempo di inviarvi la dimostrazione che supera una difficoltà che sorprese molto il Sig. [de Méré], dotato di buona intelligenza ma che non è un matematico (ciò che, come sapete, è un grande difetto). (...) Se uno vuole ottenere un *sei* con un dado avrà un vantaggio lanciando quattro volte, come 671 sta a 625.

Se uno vuole ottenere un *doppio sei* con due dadi, sarà svantaggioso in 24 lanci.

E tuttavia 24 sta a 36 (numero delle facce su due dadi) come 4 sta a sei (numero delle facce su un dado). Ecco ciò che costituisce per lui un grande scandalo che gli fa esclamare che le proposizioni non sono costanti e che l'aritmetica smentisce se stessa.<sup>14</sup>

Il rapporto 671:625 tra la probabilità di successo e quella di insuccesso si ottiene osservando che, preso  $\frac{1}{6}$  come probabilità di ottenere 6 con un dado, la probabilità di non ottenere mai 6 in quattro lanci—evento complementare di quello vincente—è

$$\left(1 - \frac{1}{6}\right)^4 = \left(\frac{5}{6}\right)^4 = \frac{625}{1296}$$

per cui la probabilità di ottenere almeno un 6 in quattro lanci è  $1 - \frac{625}{1296} = \frac{671}{1296} > \frac{1}{2}$  e si ha dunque vantaggio a scommettere di riuscire ad ottenere almeno un 6 in 4 lanci. Per risolvere il secondo quesito, l'evento complementare del doppio 6 lanciando una coppia di dadi ha probabilità  $\frac{35}{36}$  e dunque il numero  $n$  di lanci a partire dal quale è *svantaggioso* scommettere sul realizzarsi almeno una volta di un doppio 6 è il più grande intero  $n$  tale che

$$\left(\frac{35}{36}\right)^n > \frac{1}{2} :$$

passando ai logaritmi, si ottiene  $n = 24$ .

---

<sup>14</sup>Je n'ai pas le temps de vous envoyer la démonstration d'une difficulté qui étonnoit fort M. [de Méré], car il a bon esprit, mais il n'est pas géomètre (c'est, comme vous savez, un grand défaut) (...)

Si on entreprend de faire un *six* avec un dé, il y a avantage de l'entreprendre en 4, comme 671 à 625.

Si on entreprend de faire *sonnés* avec deux dés, il y a désavantage de l'entreprendre en 24.

Et néanmoins 24 est à 36 (qui est le nombre des faces de deux dés) comme 4 à 6 (qui est le nombre des faces d'un dé). Voilà quel étoit son grand scandale qui lui faisoit dire hautement que les propositions n'étoient pas constantes et que l'Arithmétique se démentoit.

## 2.5 La divisione della posta nel *Traité* di Pascal

La corrispondenza tra Pascal e Fermat non mette in evidenza il metodo generale utilizzato da Pascal per risolvere il problema della ripartizione della posta tra due giocatori che fu invece esposto nel *Traité du Triangle arithmétique et de son application*, pubblicato postumo 1665.

1	1	1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	8	9	
1	3	6	10	15	21	28	36		
1	4	10	20	35	56	84			
1	5	15	35	70	126				
1	6	21	56	126					
1	7	28	84						
1	8	36							
1	9								
1									

I numeri nel triangolo aritmetico sono assegnati a delle celle ed i triangoli che considera Pascal sono ottenuti tracciando delle diagonali a 45°: le celle che sono attraversate da una stessa diagonale formano una base di un triangolo. Pascal pose l'unità in alto a sinistra come elemento *generatore* del triangolo il cui elemento  $a_{n,k}$  all'incrocio tra la  $n$ -esima riga e la  $k$ -esima colonna è definito dalla relazione

$$a_{n,k} = a_{n-1,k} + a_{n,k-1} \quad (2.6)$$

con la convenzione che, quando uno solo dei due indici è nullo, anche il corrispondente elemento lo è. Le proprietà principali del triangolo aritmetico discendono dalla costruzione stessa e per i nostri scopi sono interessanti le Conseguenze VII ed XI. La Conseguenza VII afferma che

in ogni triangolo aritmetico, la somma delle celle di ogni base è doppia di quelle della base precedente.<sup>15</sup> ([15], p. 247)

Questa conseguenza è dedotta dalla proprietà di simmetria (Conseguenza V)

$$a_{n,k} = a_{k,n}$$

<sup>15</sup>En tout triangle arithmétique, la somme des cellules de chaque base est double de celles de la base précédente.

che si ottiene per induzione sulle basi del triangolo: infatti  $a_{1,2} = a_{2,1}$  per costruzione; invocando la (2.6) e supponendo verificata la simmetria per le basi precedenti

$$a_{n,k} = a_{n-1,k} + a_{n,k-1} = a_{k,n-1} + a_{k-1,n} = a_{k,n}$$

che è appunto la tesi. La Conseguenza VII si ottiene osservando che

$$a_{n,1} = a_{n-1,1} \quad a_{1,n} = a_{1,n-1}$$

e che, muovendosi in diagonale e servendosi della (2.6), tutti gli elementi della base  $n-1$  compaiono due volte nella somma degli elementi della base  $n$ . Poiché la somma degli elementi sulla base 2 è 2, la somma degli elementi della base  $n+1$  è  $2^n$ .

La Conseguenza X è espressa in questi termini:

In ogni triangolo aritmetico, la somma di un certo numero a piacere di celle contigue di una base, a cominciare da un'estremità, è uguale al doppio dello stesso numero di celle della base precedente, tranne uno.<sup>16</sup> ([15], pp. 247-248)

Infatti, la somma di un certo numero  $h$  di elementi della base  $n$  si può porre nella forma

$$\sum_{k=1}^h a_{n-k+1,k}$$

ed è sufficiente riflettere sulla (2.6) per concludere che *tutti* i termini della base precedente compaiono due volte fuorché  $a_{n-h+1,h}$ , che compare una volta sola. Per simmetria, la proprietà vale anche partendo da  $a_{1,n}$  e procedendo in basso lungo la base  $n$ -esima.

Pascal illustrò altre proprietà del triangolo aritmetico tra cui spiccano quelle combinatorie che servono da premessa all'applicazione alla ripartizione della posta. Nella parte del *Traité* dedicata alla soluzione di questo problema, Pascal ribadì i principii guida seguiti, specificando anzitutto la natura del contratto che i giocatori stipulano accettando di sfidarsi:

Per comprendere le regole della suddivisione occorre per prima cosa considerare che il denaro che i giocatori hanno messo in palio non appartiene più a loro; a fronte di ciò essi ricevono il diritto di attendersi quanto il caso può regalar loro, secondo le condizioni su cui sono convenuti all'inizio.<sup>17</sup> ([15], p. 257)

<sup>16</sup>En tout triangle arithmétique, la somme de tant de cellules continues qu'on voudra de sa base, à commencer par une extrémité, est égale à autant de cellules de la base précédente, plus encore à autant, hormis une.

<sup>17</sup>Pour entendre les règles des partis, la première chose qu'il faut considérer, est que l'argent que les joueurs ont mis au jeu ne leur appartient plus, car ils en ont quitté la propriété; mais ils ont reçu en revanche le droit d'attendre ce que le hasard peut leur en donner, suivant les conditions dont ils sont convenus d'abord.

La natura *volontaria* del contratto lascia spazio alla possibilità di derogare dalle condizioni previste inizialmente, interrompendo il gioco. Il problema della ripartizione della posta è formulato in questi termini:

Trattandosi di una legge volontaria, possono derogarvi di comune accordo; e così, trovandosi ad un punto qualunque del gioco, possono interromperlo e, contrariamente a quanto fatto all'inizio, rinunciare ad attendersi qualcosa dal caso e tornare ciascuno in possesso di qualcosa; in questo caso, la regola che stabilisce quanto deve loro appartenere deve essere talmente proporzionata a quanto avevano diritto di attendersi dalla fortuna, che per ciascuno di loro è del tutto indifferente accettare quanto si propone loro o continuare l'avventura del gioco: e questa giusta distribuzione è detta *la suddivisione*.<sup>18</sup>

La ripartizione della posta poggia per Pascal su due principii:

a) ciò che spetta ad un giocatore, *qualunque* sarebbe l'esito delle partite non disputate, *deve* essergli garantito;

b) inoltre bisogna garantire ad ogni giocatore la *metà* della differenza  $\Delta$  tra quanto otterrebbe vincendo e quanto otterrebbe perdendo, cioè la "somma esposta al rischio". Il coefficiente  $\frac{1}{2}$  è dovuto al fatto che le probabilità di successo dei contendenti sono uguali. Se così non fosse, occorrerebbe moltiplicare  $\Delta$  per le possibilità che il giocatore ha di vincere. Questa regola, che Pascal sottolinea applicarsi a giochi di *puro azzardo* (*pure hazard*), garantisce l'equità della ripartizione.

Il primo principio per stabilire il modo con cui effettuare la divisione della posta è questo. Se uno dei giocatori si trova in una condizione tale che, qualunque cosa succeda, una certa somma gli deve appartenere sia che vinca sia che perda, senza che l'azzardo gliela possa togliere, questa non deve essere suddivisa ma deve prenderla tutta come certa perché, dovendo essere la suddivisione proporzionata al rischio e dal momento che non vi è rischio di perderla, la deve prendere indivisa.

Il secondo principio è questo. Se i giocatori si trovano in una condizione tale che, se uno vince gli apparterrà una certa somma e se perde questa apparterrà all'altro, se vogliono lasciarsi senza giocare e prendere ciò che appartiene loro legittimamente, debbono suddividere la somma esposta al rischio a metà e ciascuno ne prenderà una.<sup>19</sup>

<sup>18</sup>Mais comme cest une loi volontaire, ils peuvent la rompre de gré à gré; et ainsi en quelque terme que le jeu se trouve, ils peuvent le quitter; et au contraire de ce qu'ils ont fait en y entrant, renoncer à l'attente du hasard, et rentrer chacun en la propriété de quelque chose; et en ce cas, le règlement de ce qui doit leur appartenir doit être tellement proportionné à ce qu'ils avoient droit d'espérer de la fortune, que chacun d'eux trouve entièrement égal de prendre ce qu'on lui assigne, ou de continuer l'aventure du jeu: et cette juste distribution s'appelle *le parti*.

<sup>19</sup>Le premier principe qui fait connoître de quelle sorte on doit faire les partis, est celui-ci. Si un des joueurs se trouve en telle condition, que, quoi qu'il arrive, une certaine somme doit lui appartenir en cas de perte et de gain, sans que le hasard puisse la lui ôter; il ne doit en faire aucun parti, mais la prendre entière comme assurée, parce que le parti devant être proportionné au hasard, puisqu'il n'y a nul hasard de perdre, il doit tout retirer sans parti.

Dal punto di vista operativo, la regola da seguire è esposta in questo corollario al secondo principio appena enunciato e che in effetti ne è una riformulazione:

Se due giocatori giocano ad un gioco di puro azzardo con la condizione che, se il primo vince otterrà una certa somma, mentre perdendo ne otterrà una minore, e se essi vogliono smettere di giocare e prendere ciò che spetta loro, la divisione da fare è che il primo prenda ciò che gli spetta in caso di sconfitta oltre alla metà dell'eccesso tra quanto otterrà in caso di vittoria ed in caso di sconfitta.<sup>20</sup>

Pascal illustrò le regole da seguire, a seconda del numero di punti mancanti a ciascun giocatore per conseguire la vittoria. Il procedimento è in effetti una esposizione più dettagliata dei metodi illustrati a Fermat nella corrispondenza e non ci soffermiamo sui vari casi considerati per passare invece alla parte più originale, in cui il triangolo aritmetico dispiega tutta la sua utilità. Siano  $P_1(a, b)$  e  $P_2(a, b)$  le poste che spettano ai giocatori  $G_1$  e  $G_2$  in funzione del numero di punti  $a$  e  $b$  che mancano loro per conseguire la vittoria. Nell'appendice al *Traité* Pascal considerò la *base* (cioè la diagonale) del triangolo aritmetico corrispondente ad  $a + b$ . A partire dall'elemento posto sulla prima colonna e procedendo lungo la base, si sommino i primi  $a$  numeri: sia  $S_1$  la loro somma. Si considerino i  $b$  numeri restanti e sia  $S_2$  la loro somma:

queste somme stanno tra loro in ragione inversa dei vantaggi dei giocatori.<sup>21</sup> ([15], p. 262)

Formalmente

$$P_1(a, b) = \frac{S_2}{S_1 + S_2} \quad P_2(a, b) = \frac{S_1}{S_1 + S_2} \quad \text{e quindi} \quad \frac{P_1(a, b)}{P_2(a, b)} = \frac{S_2}{S_1}.$$

La dimostrazione è condotta per induzione sulle basi del triangolo. La seconda<sup>22</sup> base contiene il numero 1 ripetuto due volte per cui la regola di Pascal fornisce i risultati corretti quando ad entrambi i giocatori manchi un punto. Pascal suppose che una base del triangolo contenesse tutte le ripartizioni della posta tra due giocatori cui manchino complessivamente  $n$  punti per ottenere la vittoria

---

Le second est celui-ci. Si deux joueurs se trouvent en telle condition, que si l'un gagne, il lui appartiendra une certaine somme, et s'il perd, elle appartiendra à l'autre; si le jeu est de pur hasard, et qu'il y ait autant de hasards pour l'un que pour l'autre, et par conséquent non plus de raison de gagner pour l'un que pour l'autre, s'ils veulent se séparer sans jouer, et prendre ce qui leur appartient légitimement, le parti est qu'ils séparent la somme qui est au hasard par la moitié, et que chacun prenne la sienne.

<sup>20</sup>Si deux joueurs jouent à un jeu de pur hasard, à condition que si le premier gagne, il lui reviendra une certaine somme, et s'il perd, il lui en reviendra une moindre; s'ils veulent se séparer sans jouer, et prendre chacun ce qui leur appartient, le parti est, que le premier prenne ce qui lui revient en cas de perte, et de plus la moitié de l'excès, dont ce qui lui reviendrait en cas de gain, surpasse ce qui lui revient en cas de perte.

<sup>21</sup>ces sommes sont l'une à l'autre comme les avantages des joueurs réciproquement.

<sup>22</sup>La prima base è costituita dal solo numero 1 e rappresenta certamente la divisione della posta quando ad un giocatore manchi un punto e all'altro nessuno: tutta la posta spetta a chi ha vinto!

e si propose di dimostrare che la stessa proprietà vale quando i punti mancanti sono  $n + 1$ . Per questo egli si chiede quanto otterrebbe  $G_1$  se, nel momento in cui  $a + b = n + 1$ , si giocasse un'altra partita. Diciamo per semplicità  $d_{n,h}$  l'elemento  $h$  della  $n$ -esima diagonale, contato a partire dal lato verticale del triangolo aritmetico. Sia che  $G_1$  vinca o perda la nuova partita, dopo averla disputata, ai due contendenti mancheranno  $n$  punti in tutto. Per l'ipotesi di induzione quindi, se  $G_1$  vincessesse,  $G_2$  avrebbe diritto alla frazione

$$P_2(a - 1, b) = \frac{d_{n,1} + d_{n,2} + \cdots + d_{n,a-1}}{\Sigma_n}, \quad (2.7)$$

dove  $\Sigma_n$  è la somma di tutti gli elementi della diagonale  $n$ -esima. Se invece vincessesse  $G_2$ , gli spetterebbe la frazione

$$P_2(a, b - 1) = \frac{d_{n,1} + d_{n,2} + \cdots + d_{n,a-1} + d_{n,a}}{\Sigma_n} \quad (2.8)$$

della posta. In base alle regole per stabilire la ripartizione, a  $G_2$  spetterà la frazione della posta pari a

$$\frac{d_{n,1} + d_{n,2} + \cdots + d_{n,a-1}}{\Sigma_n} + \frac{1}{2} \frac{d_{n,a}}{\Sigma_n} = \frac{2(d_{n,1} + d_{n,2} + \cdots + d_{n,a-1}) + d_{n,a}}{2\Sigma_n}.$$

Ora, per la Conseguenza VII vista prima si ha  $\Sigma_{n+1} = 2\Sigma_n$  mentre, per la Conseguenza XI,

$$2(d_{n,1} + d_{n,2} + \cdots + d_{n,a-1}) + d_{n,a} = d_{n+1,1} + d_{n+1,2} + \cdots + d_{n+1,a}.$$

La somma che spetta a  $G_2$  è allora

$$\frac{d_{n+1,1} + d_{n+1,2} + \cdots + d_{n+1,a}}{\Sigma_{n+1}},$$

che dimostra l'asserto di Pascal. La disposizione adottata da Pascal per il triangolo aritmetico è particolarmente felice perché la diagonale  $m + 1$ -esima contiene i coefficienti binomiali

$$\binom{m}{h} \quad h = 0, \dots, m.$$

La soluzione del problema della ripartizione della posta è allora chiara dal punto di vista combinatorio perché se a  $G_1$  mancano  $a$  punti e a  $G_2$   $b$  punti per vincere, la base  $(a + b)$ -esima contiene i coefficienti del tipo

$$\binom{a + b - 1}{h} \quad h = 0, 1, \dots, a + b - 1$$

pertanto, attribuire la frazione

$$\frac{1}{2^{a+b-1}} \sum_{h=0}^{a-1} \binom{a + b - 1}{h} \quad (2.9)$$

a  $G_2$  e

$$\frac{1}{2^{a+b-1}} \sum_{k=a}^{a+b-1} \binom{a+b-1}{k} \quad (2.10)$$

a  $G_1$  significa isolare, tra i  $2^{a+b-1}$  risultati possibili delle partite, quelli che permettono a  $G_2$  o a  $G_1$ , rispettivamente, di prevalere. Rilassando l'ipotesi che i giocatori abbiano uguale possibilità di vittoria, ma che  $G_1$  abbia  $p$  possibilità di prevalere e  $G_2$  ne abbia  $q$ , la formula precedente può essere modificata con l'intervento delle *probabilità* di successo ad ogni turno

$$p_1 := \frac{p}{p+q} \quad \text{e} \quad p_2 := \frac{q}{p+q} :$$

la frazione di posta da attribuire a  $G_2$  è, posto  $N := a + b - 1$ ,

$$\sum_{h=0}^{a-1} \binom{N}{h} p_1^h q_1^{N-h}$$

mentre quella da attribuire a  $G_1$  è

$$\sum_{h=a}^N \binom{N}{h} p_1^h q_1^{N-h} :$$

questa generalizzazione fu pubblicata nell'*Essay* di Rémond de Montmort ([19], pp. 244-248).

# Bibliografia

- [1] J. Bernoulli: *Ars Conjectandi, opus posthumum*, Basel, (1713).
- [2] U. Bottazzini, P. Freguglia, L. Toti Rigatelli: *Fonti per la storia della matematica*. Sansoni, Firenze, (1992).
- [3] G. Cardano: *Practica Arithmeticae et mensurandis singularis*. Calusci, Milano, (1539). In Hieronymi Cardani Mediolanensis Opera Omnia, vol. IV, Huguetan & Ravaud, Lugduni, (1663), 13-216.
- [4] E. Coumet: Le problème des partis avant Pascal. *Archives Internationelles d'Histoire des Sciences*, **18**, 245-272, (1965).
- [5] P. de Fermat: *Œuvres de Fermat*. Tome II. Gauthier-Villars, Paris, (1894).
- [6] P. Dupont, C.S. Roero: Il trattato *De ratiociniis in ludo aleae* di Christiaan Huygens con le *Annotationes* di Jakob Bernoulli (*Ars Conjectandi*, Parte I) Presentati in traduzione italiana, con commento storico-critico e risoluzioni moderne. *Memorie della Accademia delle Scienze di Torino*, **8** (S. V), 1-258, (1984).
- [7] L. Forestani: *Pratica d'arithmetica e geometria*. II Edizione, Siena, Stamparia del Publico, (1682).
- [8] A. Hald: *History of probability and Statistics and their applications before 1750*. Wiley-Interscience, Hokoben, New-Jersey, U.S.A. (1988).
- [9] C. Huygens: *De Ratiociniis in Ludo Aleae*. In [13], pp. 52-91.
- [10] M.G. Kendall: Studies in the history of probability and statistics. II. The beginnings of a probability calculus. *Biometrika*, **43**, (1956), 1-14.
- [11] M.G. Kendall: Studies in the history of probability and statistics. X. Where shall the history of statistics begin? *Biometrika*, **47**, (1960), 447-449.
- [12] M.-H. Labarthe: Les règles de compagnie, dans les premières arithmétiques imprimées des Espagnes: de la règle marchanda à l'outil mathématique. *Revue d'histoire des mathématiques*, **11**, 257-313.

- [13] N. Meusnier: Le problème des partis bouge... de plus en plus. *Journ@l électronique d'Histoire des Probabilités et de la Statistique*, **3**, 33 pp., (2007).
- [14] L. Pacioli: *Summa de Arithmetica, proportioni et proportionalità*, (1494).
- [15] B. Pascal: *Traité du Triangle arithmétique et de son application*. Desprez, Paris, (1665). In *Oeuvres complètes*, Tome III, Hachette, Paris, (1872), pp. 243-268.
- [16] G.F. Peverone: *Arithmetica e Geometria*. G. di Tornes, Lione, (1581).
- [17] P. Rémond de Montmort: *Essay d'Analyse sur les jeux d'hazard*. II édition Revûe et augmentée de plusieurs Lettres. Quillau, Paris, (1713).

## Capitolo 3

# Dai giochi d'azzardo alle rendite vitalizie: Huygens e De Moivre

### 3.1 Il *De Ratiociniis in Ludo Aleae* di Huygens

Christiaan Huygens era venuto a conoscenza dei problemi trattati nella corrispondenza tra Pascal e Fermat durante il viaggio che lo portò a Parigi, in compagnia del fratello Lodewijk (1631-1699) e del cugino Philips Doublet, tra la metà del luglio e la fine di novembre del 1655. A Parigi Huygens non incontrò Fermat—che viveva a Tolosa—né Pascal e neppure Pierre de Carcavi, che aveva svolto il ruolo di intermediario negli scambi epistolari tra Pascal e Fermat. Tuttavia conobbe un amico di quest'ultimo, Claude Mylon, e Gilles Personne de Roberval, entrambi al corrente della corrispondenza tra Fermat e Pascal: è possibile che essi abbiano informato Huygens del problema della ripartizione della posta e degli altri problemi trattati in tale scambio epistolare. Tornato in Olanda, Huygens studiò i problemi sui giochi d'azzardo e già il 20 aprile del 1656 poté inviare a Frans van Schooten, professore di matematica all'università di Leida, buona parte del materiale che costituirà il *De Ratiociniis in ludo aleae*. Huygens aveva scritto in olandese, anziché in latino, perché molti dei termini tecnici relativi al gioco d'azzardo mancavano. Terminata la stesura in olandese, Huygens aveva scelto i termini necessari alla traduzione latina comunicandoli a van Schooten, sempre nella lettera del 20 aprile 1656. Benché Huygens avesse ventilato la possibilità di scrivere egli stesso la versione latina, da pubblicare prima di quella olandese, van Schooten decise di provvedere da sé alla traduzione che non lasciò del tutto soddisfatto Huygens il quale, pur avendo avuto la possibilità di rivedere il lavoro di van Schooten, così si esprimeva rivolgendosi al matematico vallone René-François de Sluse (1622-1685), in una lettera del 27 luglio 1657:

Ti invierò quanto prima il libro di [van] Schooten da poco pubblicato (...) Ci vedrai annesso un breve mio trattato sui ragionamenti circa i giochi d'azzardo non però agevolmente tradotto in latino a partire dalla lingua Belga nella quale lo redassi.<sup>1</sup>

È possibile che Huygens fosse contrariato perché van Schooten non aveva seguito i suggerimenti inviartigli se non in minima parte. Huygens era ansioso di sapere se le sue soluzioni ai problemi proposti fossero conformi a quelle di Fermat e Pascal e per questo, tramite Carcavi e Mylon, cercò il parere di entrambi che confermarono la correttezza dei risultati da lui ottenuti. Fermat allegò alla risposta una serie di altri problemi che Huygens risolse rapidamente. Conseguita la consapevolezza della bontà dei propri metodi e dei risultati, il *De ratiociniis* fu infine pubblicato nel 1657, tra agosto e settembre, in coda alle *Exercitationes mathematicae* di van Schooten. Si trattò del primo testo a stampa pubblicato sul calcolo delle probabilità: l'edizione olandese comparirà solo nel 1660. Per mezzo secolo, l'opera di Huygens rimarrà l'unico testo di introduzione al calcolo delle probabilità ed avrà un'ottima accoglienza presso gli studiosi. Grazie alla traduzione di van Schooten, il trattato di Huygens non subì la sorte toccata a molti altri suoi manoscritti: quella di essere oltremodo ritardati nella pubblicazione.

Nella lettera a van Schooten con cui si apre il *De ratiociniis*, Huygens giustificò l'inserimento di una appendice dedicata al gioco d'azzardo in questi termini:

Sapendo che, con la pubblicazione dei lodevoli frutti della vostra intelligenza e del vostro zelo, vi proponete anche di far vedere, grazie alla varietà degli argomenti trattati, l'ampiezza del campo sul quale si estende la nostra eccellente Arte *Algebraica*, non dubito che il presente scritto sul calcolo nei giochi d'azzardo potrà servire a raggiungere questo obiettivo. In effetti, quanto più è difficile stabilire con la ragione quanto è incerto e soggetto al caso, tanto più la scienza che raggiunge questo risultato sembrerà degna di ammirazione.<sup>2</sup> ([11], p. 56)

Lo studio dei giochi d'azzardo non rappresenta per Huygens, uno dei maggiori scienziati del '600, un passatempo intellettuale:

anche se alcuni lettori potrebbero pensare che io abbia lavorato su argomenti di poca importanza, essi non condanneranno come del tutto inutile e indegno di ogni lode quanto voi avete accettato di adottare in questo

<sup>1</sup>Schotenij librum recens editum quam primum potero tibi mittam (...) Breve quoque tractatum meum de Ratiocinijs in ludo Aleae, adjunctum videbis, sed non satis commode è lingua Belgica qua fuerat a me conscriptus in latinam conversum.

<sup>2</sup>Sachant qu'en publiant les louables fruits de votre intelligence et de votre zèle, vous vous proposez entre autres de faire voir, par la diversité des sujets traités, la grandeur du champ sur lequel notre excellent Art *Algébrique* s'étend, je ne doute pas que le présent écrit au sujet du Calcul dans les Jeux de hasard ne puisse vous servir à atteindre ce but. En effet, plus il semble difficile de déterminer par la raison ce qui est incertain et soumis au hasard, plus la science qui parvient à ce résultat paraîtra admirable.

modo, come se fosse un'opera vostra, dopo averlo tradotto, non senza fatica, dalla nostra lingua al latino. Voglio però credere che, considerando queste cose con più attenzione, il lettore percepirà presto che non si tratta di un semplice divertimento dello spirito ma che vi si gettano i semi di una riflessione molto interessante e profonda.<sup>3</sup>

Vedremo nelle prossime sezioni i primi passi di Huygens in quella che si può definire l'aspetto più affascinante della storia della probabilità: la sua uscita dalle sale da gioco alla conquista dei più svariati campi del sapere umano.

Huygens iniziò a dichiarare l'ipotesi di fondo che userà per trattare tutti i problemi sui giochi d'azzardo:

in un gioco d'azzardo (in ludo aleae) l'eventualità (sortem) ovvero l'aspettazione (expectatio) che ciascuno ha di ottenere qualcosa deve avere quello stesso valore che, se egli lo possedesse, gli sarebbe possibile procurarsi la stessa eventualità od aspettazione, giocando equamente.<sup>4</sup> ([11], p. 60)

Huygens illustrò questo postulato con questo esempio:

Se qualcuno, a mia insaputa, nasconde 3 monete in una mano e 7 nell'altra, mi desse l'opportunità di prendere le monete dalla mano che preferisco, io dico che ciò ha per me lo stesso valore che se mi fossero date 5 monete. Perché, avendo cinque monete, posso di nuovo giungere al punto in cui possiedo pari aspettazione di avere 3 o 7 monete: e ciò in un gioco equo. ([11], p. 62)

L'aspettazione è dunque una sorta di punto di equilibrio, rispetto al quale si compensano le perdite ed i guadagni. Questo "principio" di Huygens appare una definizione viziosa perché l'aspettazione viene definita per il tramite della nozione di gioco equo che, almeno ai nostri occhi, è tale quando i giocatori hanno aspettative pari alle quote versate all'inizio del gioco. È stato osservato però [22] che Huygens poteva presupporre nota la nozione di gioco equo ai suoi lettori per altra via, trattandosi di un concetto utilizzato in ambito giuridico-economico per definire i contratti aleatori. Per esempio, il giurista e filosofo tedesco Samuel von Pufendorf (1632-1694), a proposito dei contratti aleatori sosteneva che i rischi di guadagnare o perdere debbono essere in giusta proporzione con il denaro investito e che tutti debbono condividere un ugual rischio di guadagnare come di perdere. Pufendorf, che scrisse queste parole nel voluminoso *De jure naturae et gentium libri octo*, pubblicato nel 1672 e quindi dopo

<sup>3</sup>si quelques lecteurs pourraient bien s'imaginer que j'ai travaillé sur des sujets de faible importance, ils ne condamneront néanmoins pas comme complètement inutile et indigne de toute louange ce que vous voulez bien adopter de cette façon comme si c'était votre propre ouvrage, après l'avoir traduit, non sans quelque labeur, de notre langue en Latin. Toutefois je veux croire qu'en considérant ces choses plus attentivement, le lecteur apercevra bientôt qu'il ne s'agit pas ici d'un simple jeu d'esprit, mais qu'on y jette les fondements d'une spéculation fort intéressante et profonde.

<sup>4</sup>in aleae ludo tanti aestimandam esse cujusque sortem seu expectationem ad aliquid obtinendum, quantum si habeat, possit denuo ad similem sortem sive expectationem pervenire, aequa conditione certans.

Huygens, riprese in realtà idee che circolavano da tempo, come nel *De jure belli ac pacis* pubblicato nel 1625 da giurista olandese Huig de Groot (Ugo Grozio, 1583-1645). La definizione di gioco equo che compare in questi testi è solo di natura *qualitativa*, non *quantitativa*.

Anche la Proposizione I del *De Ratiociniis* va letta in quest'ottica, che la libera dal sospetto di essere circolare:

Se io mi aspettassi  $a$  o  $b$ , potendo ottenerli in modo ugualmente facile, bisogna dire che la mia aspettazione è  $\frac{a+b}{2}$ .<sup>5</sup> ([11], p. 62)

Huygens distingue la *costruzione* della Proposizione dalla sua *dimostrazione*. Per la prima, egli chiama  $x$  il valore incognito dell'aspettazione e si pone nella condizione di avere  $x$  in un gioco equo, che costruisce in questo modo. Due giocatori impegnano la stessa somma  $x$  con l'accordo che il vincitore donerà la somma  $a$  a chi perde, una sorta di premio di consolazione, se  $a < x$ , ed intascherà la somma restante  $2x - a$ . Se si richiede che  $2x - a = b$ , si ottiene per  $x$  il valore indicato nella Proposizione I. In questo modo Huygens ha *costruito* la Proposizione I. Per la dimostrazione egli parte dal valore  $\frac{a+b}{2}$  come somma messa in palio da ciascuno dei due giocatori, l'accordo tra i quali è che, chi vince lasci una somma pari ad  $a$  all'altro. In questo modo, se si suppone che i giocatori abbiano la stessa possibilità di vincere o perdere, si ha la stessa possibilità di guadagnare  $a$  come  $b = (a + b) - a$ . Osserviamo come in questa Proposizione gli esiti del gioco abbiano la stessa possibilità di verificarsi. Similmente, nella proposizione II Huygens considera il caso in cui il gioco dia pari possibilità di conseguire  $a$ ,  $b$ , o  $c$ : l'aspettazione è allora  $\frac{a+b+c}{3}$ .

La Proposizione III è interessante per il modo con cui Huygens tratta il caso in cui gli esiti del gioco non sono equivalenti. Essa è il perno di tutti i ragionamenti di Huygens sui giochi d'azzardo, come aveva scritto il 6 luglio 1656 a Carcavi, dicendo di servirsene

in tutti questi problemi di ripartizione di un gioco. ([13], p. 64).

La proposizione è così enunciata:

Sia  $p$  il numero di casi nei quali mi spetta  $a$  e  $q$  il numero di casi in cui mi spetta  $b$ ; nell'ipotesi che tutti questi casi abbiano la stessa possibilità di accadere: la mia aspettazione varrà  $\frac{pa+qb}{p+q}$ .<sup>6</sup> ([11], p. 64)

Huygens chiama  $x$  l'aspettazione e costruisce un gioco equo che possa conferirgli gli stessi esiti di quello proposto. Per questo ipotizza che vi siano  $p+q$  giocatori, ciascuno dei quali gioca la somma  $x$ , cosicché la somma totale disponibile sia  $(p+q)x$ . Un giocatore, diciamo  $G_1$ , stipula con ognuno tra  $q$  di questi giocatori questo *contratto*: chi tra loro vincerà darà  $b$  a  $G_1$  che, vincendo a sua volta, darà  $b$  a ciascuno di loro. Con i restanti  $p-1$  giocatori,  $G_1$  stipula un contratto simile:

<sup>5</sup>Si  $a$  vel  $b$  expectem, quorum utrumvis aequae facile mihi obtingere possit, expectatio mea dicenda est  $\frac{a+b}{2}$ .

<sup>6</sup>Si numerus casuum, quibus mihi eveniet  $a$ , sit  $p$ ; numerus autem casuum, quibus mihi eveniet  $q$ , sumendo omnes casus aequae in proclivi esse: expectatio mea valebit  $\frac{pa+qb}{p+q}$ .

chi tra loro vincesse, riconoscerà a  $G_1$  una somma pari ad  $a$  e  $G_1$ , vincendo a sua volta, cederà la somma  $a$  ad ognuno di questi giocatori. Se tutti gli esiti sono equivalenti,  $G_1$  ha  $q$  aspettative pari a  $b$  (una per ogni successo dei primi  $q$  giocatori),  $p - 1$  pari ad  $a$  (una per ogni successo degli altri  $p - 1$  giocatori) ed una alla cifra residua

$$(p + q)x - bq - a(p - 1)$$

in caso di una sua vittoria. Se si impone che quest'ultimo valore sia pari ad  $a$ , lo spettro degli esiti del nuovo gioco *equo*, in senso qualitativo, sarà di  $p$  aspettative per  $a$  e  $q$  per  $b$ , come nel gioco iniziale. Dunque

$$(p + q)x - bq - a(p - 1) = a$$

e quindi

$$x = \frac{ap + bq}{p + q}. \quad (3.1)$$

Osserviamo che Jakob Bernoulli, commentando nell'*Ars conjectandi* la dimostrazione di questa Proposizione paragonerà la regola di Huygens all'*allegazione* (*alligatio*), una regola che permetteva di ricavare il prezzo della miscela di più prodotti, conoscendo i prezzi unitari e le quantità dei prodotti componenti la miscela.

Queste proposizioni sono la necessaria premessa alla soluzione del problema della ripartizione della posta, che Huygens affrontò a partire dalla Proposizione IV.

Si supponga che qualcuno giochi con me, a questa condizione: chi per primo vince tre volte, si aggiudicherà la cifra in palio, e supponiamo che io abbia già vinto due partite, l'altro una. Desidero sapere se, non volendo proseguire oltre nel gioco ma [volendo] dividere la cifra in palio, per la quale giochiamo, in maniera equa, quanto di questa cifra mi spetta.<sup>7</sup> ([13], p. 66)

Seguiamo il ragionamento di Huygens. Sia  $a$  la cifra in palio ed al giocatore  $G_1$  manchi un punto per conseguire la vittoria mentre ne manchino due all'altro,  $G_2$ . Huygens calcola quale porzione della cifra in palio spetterebbe a ciascuno, continuando il gioco. Siccome  $G_1$  ha vinto due partite, in caso di una sua ulteriore vittoria alla quarta partita, l'intera posta— $a$ — spetterebbe a lui. Se al contrario la quarta partita venisse vinta da  $G_2$ , vi sarebbe una situazione di parità e l'aspettazione di ciascuno è per la somma  $\frac{a}{2}$ . Ora, supponendo che  $G_1$  possa indifferentemente vincere o perdere la quarta partita, egli avrebbe la stessa aspettazione per  $a$  o per  $\frac{a}{2}$  e dunque, per la Proposizione I, la sua aspettazione è  $\frac{3}{4}a$ : questa è la parte della posta che spetta ad  $G_1$  in caso di

<sup>7</sup>Sumpto itaque me cum aliquo certare, hoc pacto: ut qui prius ter vicerit, quod depositum est, lucretur, & me jam bis vicisse, alterum vero semel. Scire cupio, di lusum prosequi non velimus, sed pecuniam, de qua certamus, prout aequum est, partiri, quantum ejus mihi obtingeret.

interruzione del gioco nelle condizioni dichiarate nell'enunciato. Ovviamente a  $G_2$  spetta la quota  $\frac{1}{4}a$ .

Come si vede, il metodo di Huygens è sovrapponibile a quello seguito da Pascal nella corrispondenza con Fermat e, come aveva fatto Pascal in quella sede, egli illustrò nelle proposizioni V-VII altri casi particolari, dall'esame dei quali era possibile risalire al caso generale con un procedimento ricorsivo. Nella Proposizione VII egli esamina il caso in cui ad  $G_1$  manchino 2 punti e a  $G_2$  4, ottenendo che la posta che spetta ad  $G_1$  è  $\frac{26}{32}a = \frac{13}{16}a$ . Huygens osserva che per  $G_1$  questa seconda condizione è migliore della prima, a dispetto del fatto che, nei due casi, il *rapporto* tra il numero di punti mancanti ad  $G_1$  e a  $G_2$  sia  $\frac{1}{2}$  in entrambi i casi.

Con la Proposizione VIII, Huygens generalizzò il problema della ripartizione della posta al caso in cui siano coinvolti tre giocatori. Anche qui egli considerò alcuni casi particolari, ottenendo le soluzioni per ricorrenza. Così, nella Proposizione VIII egli suppose che mancasse  $n = 1$  punto ad  $G_1$  per vincere l'intera posta,  $m = 1$  punto a  $G_2$  e  $p = 2$  punti al terzo giocatore  $G_3$ . Se vincessero  $G_1$  la nuova partita, egli si aggiudicherebbe l'intera posta che invece andrebbe a  $G_2$  se egli vincessero—ed in questo caso  $G_1$  non guadagnerebbe nulla. Infine, se  $G_3$  vincessero, tutti i giocatori sarebbero nelle stesse condizioni e a ciascuno di essi aspetterebbe  $\frac{a}{3}$ . Supponendo i tre risultati equivalenti, otteniamo che l'aspettazione di  $G_1$  vale

$$\frac{a + 0 + \frac{a}{3}}{3} = \frac{4}{9}a :$$

un'uguale aspettazione compete a  $G_2$ , mentre la porzione di posta spettante a  $G_3$  è  $\frac{a}{9}$ . Di questa Proposizione, come della seguente, Huygens diede una dimostrazione articolata in un foglio separato, probabilmente destinato a Carcavi ([13], pp. 92-95). Il ragionamento appena illustrato è lo stesso dell'anonimo autore del secondo manoscritto illustrato nel capitolo precedente. La Proposizione IX, che chiude la prima parte del *De Ratiociniis*, contiene l'enunciato della regola da seguire per effettuare il calcolo della ripartizione della posta nel caso in cui ad alcuni giocatori manchino un certo numero di punti per vincere l'intera posta, ad altri ne manchi un numero maggiore. La regola<sup>8</sup> non fu svolta nel dettaglio ma Huygens allegò una tabella in cui i risultati sono riportati per alcuni valori del numero di partite mancanti a ciascun giocatore.

La seconda parte del *De ratiociniis* contiene problemi sul il gioco dei dadi (*de tesseris*). Dopo aver calcolato il numero di casi possibili nel lancio di uno, due, tre o quattro dadi, egli ricavò il numero di casi favorevoli a ciascuno dei punteggi da 2 a 12 ottenibili lanciando due dadi. A partire dalla Proposizione X, Huygens esaminò i problemi di De Meré, su quanto *convenga* scommettere sull'uscita del numero 6 in un certo numero di lanci di un dado o, meglio, quante volte conviene lanciare un dado se si vuole aver vantaggio nell'ottenere un 6. La convenienza a scommettere si ha quando la probabilità di successo supera  $\frac{1}{2}$ . Se la partita si esaurisse in un solo lancio, l'aspettazione del giocatore  $A$  che

<sup>8</sup>Huygens parla di un numero arbitrario di giocatori ma il suo esempio si riferisce ancora al caso in cui sono coinvolti tre giocatori.

lo effettua sarebbe  $\frac{a}{6}$ ,  $a$  essendo la posta totale, mentre l'altro giocatore  $B$ , che punta sulla mancata uscita del 6, ha una aspettazione di  $\frac{5}{6}a$  per cui occorre che  $A$  punti 1 mentre  $B$  punti 5, affinché il gioco sia equo. Supponiamo ora che la partita consti di due lanci. Il giocatore che lancia il dado può ottenere 6 al primo lancio, aggiudicandosi direttamente l'intera posta  $a$ ; se ciò non accade, egli ha ancora il secondo lancio a disposizione e, per quanto dimostrato appena prima, l'aspettazione è  $\frac{a}{6}$ . Siccome vi è un solo caso a favore dell'uscita del numero 6 al primo lancio e 5 contrari, l'aspettazione complessiva di  $A$  è ora

$$\frac{a + 5\frac{a}{6}}{6} = \frac{11}{36}a$$

mentre quella di  $B$  è  $\frac{25}{36}a$ : occorre dunque che le cifre messe in palio siano nel rapporto di 25 ad 11 per garantire l'equità. Se la partita consta di tre lanci, lo schema si ripropone perché o si vince  $a$  subito (un caso favorevole) oppure si perde al primo lancio nei cinque casi restanti e si hanno ancora due possibilità. L'aspettazione è

$$\frac{a + 5 \times \frac{11}{36}a}{6} = \frac{91}{216}a.$$

Infine, se la partita si decide in quattro lanci, al primo lancio si può vincere  $a$  oppure si hanno ancora tre possibili lanci per ottenere un sei e l'aspettazione è

$$\frac{a + \frac{91}{216}a}{6} = \frac{671}{1296}a \simeq \frac{a}{2}$$

per cui  $n = 4$  risolve il problema proposto. Notiamo come lo schema di soluzione sia ancora una volta iterativo. L'aspettazione  $A_1(n)$  che si ha scommettendo sull'uscita di un 6 in  $n$  lanci verifica l'equazione

$$A_1(n) = \frac{1}{6} [a + 5A_1(n-1)], \quad (3.2)$$

come si vede ricostruendo lo schema generale appena visto per  $n = 1$  ed  $n = 2$ . Come già osservato in precedenza, si potrebbe risolvere il problema rapidamente calcolando la probabilità  $(\frac{5}{6})^n$  dell'evento complementare, cioè che il 6 non si presenti mai in  $n$  lanci, per cui la probabilità cercata è  $1 - (\frac{5}{6})^n$ . Il metodo di Huygens è più prolisso e riproduce le tappe in cui si snoda l'andamento del gioco. Inoltre, il calcolo di Huygens è un calcolo delle *aspettazioni* e non un calcolo delle probabilità. La Proposizione XI pone lo stesso problema quando l'evento che si vuole ottenere è un doppio 6. Con un unico lancio l'aspettazione è  $\frac{a}{36}$ , con due lanci si hanno questi scenari: o il giocatore  $A$  ottiene il doppio 6 al primo lancio, con un'aspettazione di  $a$ , oppure non lo ottiene, ed allora ha ancora un lancio a disposizione per raggiungere l'obiettivo e la sua aspettazione è, per il passo precedente,  $\frac{a}{36}$ . Poiché questa eventualità si presenta in 35 casi equivalenti, l'aspettazione di  $A$  è

$$\frac{a + \frac{35}{36}a}{36} = \frac{71}{1296}a :$$

occorrono 25 lanci perché l'aspettazione di  $A$  sia superiore a quella del banco o giocatore  $B$ . Più delicata è la Proposizione XII:

Trovare il numero di dadi con i quali è conveniente giocare per ottenere due 6 al primo lancio.<sup>9</sup>

Huygens riconobbe l'equivalenza di questo problema con quello di trovare il numero minimo di lanci di un solo dado perché le possibilità che compaia due volte il numero 6 siano maggiori delle possibilità che ciò non accada. Anche qui, Huygens ragionò in termini di aspettazione  $a$ . Chi vuole ottenere due 6 in due lanci di un dado ha aspettazione  $\frac{a}{36}$ . L'aspettazione di ottenere due 6 in tre lanci si calcola in questo modo: se l'esito del primo lancio *non* è 6, chi lancia ha due lanci residui per raggiungere l'obiettivo e, per quanto appena visto, la sua aspettazione è  $\frac{a}{36}$ . Se al primo lancio si ottiene 6, l'aspettazione da calcolare è quella di ottenere un altro 6 nei due lanci restanti e dunque, per la Proposizione X, l'aspettazione è  $\frac{11}{36}a$ . Poiché vi è un solo caso favorevole all'uscita del 6 al primo lancio e cinque contrari, dalla Proposizione III si ha che l'aspettazione complessiva è

$$\frac{5 \frac{1}{36} + 1 \frac{11}{36}}{6} a = \frac{2}{27} a :$$

procedendo, si mostra che il gioco è vantaggioso per chi lancia, se si gioca con almeno 10 dadi. Questa proposizione, come le precedenti, si può riformulare nel linguaggio delle equazioni alle differenze finite. Se diciamo  $A_2(n)$  l'aspettazione che si ha scommettendo di ottenere 2 volte il numero 6 lanciando  $n$  dadi, deve anzitutto essere  $A_2(1) = 0$  e lo schema di Huygens si traduce nell'equazione

$$A_2(n) = \frac{1}{6} [5A_2(n-1) + A_1(n-1)]. \quad (3.3)$$

Anche il problema studiato nella Proposizione XIV è risolto in modo piuttosto articolato:

Se un altro giocatore ed io lanciamo a turno due dadi con l'accordo che io vincerò se avrò ottenuto 7 punti e lui vincerà ottenendo 6 punti; concederò chiaramente a lui il primo lancio: trovare il rapporto tra le mie possibilità e le sue.<sup>10</sup>

Il punto tecnico su cui Huygens rivolse l'attenzione è che le aspettative di un giocatore cambiano a seconda di chi sia di turno a lanciare. Sia  $G_1$  il giocatore che vince se ottiene 7 come punteggio totale e  $G_2$  il giocatore che vince ottenendo 6 e diciamo  $x$  ed  $y$  le aspettative di  $G_1$  quando tocca a  $G_2$  o a  $G_1$  lanciare il dado, rispettivamente. Ora, poiché si ottiene 6 in 5 casi e non lo si ottiene in 31 casi, il valore di  $x$  verifica l'equazione

$$x = \frac{5 \cdot 0 + 31y}{36} = \frac{31}{36}y \quad (3.4)$$

<sup>9</sup>Invenire, quot tesseris suscipere quis possuit, ut primâ vice duos senarios jaciât.

<sup>10</sup>Si ego et alius duabus tesseris alternatim jiciamus, hac conditione, ut ego vincam simul atque septenarium jiciam, ille vero quam primum senarium jaciât; ita videlicet, ut ipsi primum iactum concedam: Invenire rationem meae ad ipsius sortem.

dal momento che, se  $G_2$  non ottiene 6, il gioco torna a  $G_1$  e la sua aspettazione è  $y$ . Quando è il turno di  $G_1$ , poiché il 7 può verificarsi in 6 casi, il valore dell'aspettazione verifica

$$y = \frac{6a + 30x}{36} = \frac{a + 5x}{6}$$

dal momento che, se  $G_1$  non ottiene 7, il gioco passa a  $G_2$  e l'aspettazione di  $G_1$  ritorna ad essere  $x$ . Sostituendo questo valore di  $y$  nell'equazione precedente si ottiene

$$x = \frac{31}{61}a, \quad y = \frac{36}{61}a.$$

Dunque, quando è il turno di  $G_2$ , la sua aspettazione è  $a - x = \frac{30}{61}a$  ed il rapporto richiesto è di 31 : 30.

Il *De Ratiociniis* si conclude con cinque Proposizioni la cui soluzione fu proposta al lettore. Il primo problema è una variante più complessa della Proposizione XIV ed è dovuto a Fermat:

I.  $A$  e  $B$  giocano con la condizione seguente:  $A$  vincerà se ottiene 6 punti;  $B$  se ne ottiene 7.  $A$  farà un solo lancio all'inizio; poi  $B$  ne farà 2, quindi  $A$  due lanci di nuovo, e così via finché uno dei due non sarà riuscito a vincere. Si richiede il rapporto tra la possibilità di successo di  $A$  e di  $B$ .  
Risposta: come 10355 a 12276.

Il secondo quesito è in realtà ambiguo e fu interpretato in tre modi differenti da chi ne affrontò la soluzione:

II. Tre giocatori:  $A$ ,  $B$  e  $C$  prendono 12 gettoni di cui 4 bianchi ed 8 neri. Vincerà chi per primo, bendato, estrarrà un gettone bianco. Il primo a giocare sarà  $A$ , seguito da  $B$  e  $C$ .

L'ambiguità risiede nel fatto che si può intendere il problema con o senza reimbussolamento. Huygens diede una soluzione articolata di questo come del quarto problema, scritta presumibilmente nel 1665. Il terzo problema riguarda il gioco delle carte:

III.  $A$  scommette con  $B$  di riuscire ad estrarre, da un mazzo di 40 carte, dieci per colore, 4 carte di colori tutti diversi. Si trova in questo caso che le possibilità di  $A$  stanno a quelle di  $B$  come 1000 sta ad 8139.

Segue una variante del secondo problema:

IV. Si prendono, come sopra, 12 gettoni, 4 bianchi ed 8 neri.  $A$  scommette con  $B$  che, estraendo bendato 7 gettoni, ne troverà 3 bianchi. Si domanda il rapporto tra le possibilità di  $A$  e quelle di  $B$ .

Infine, si trova un problema che Pascal aveva posto a Fermat e del quale Huygens venne a conoscenza grazie ad una lettera di Carcavi del 28 settembre 1656. Huygens gli inviò la soluzione due settimane più tardi, il 12 ottobre. Si tratta del celebre problema della "rovina" del giocatore, formulato in questi termini.

V.  $A$  e  $B$ , avendo preso ciascuno 12 gettoni, giocano con tre dadi usando queste regole: se il punteggio totale è 11,  $A$  dà un gettone a  $B$  mentre  $B$  darà un gettone ad  $A$  quando il punteggio totale è 14. Il vincitore sarà colui che per primo otterrà tutti i gettoni. Si trova in questo caso che le possibilità di  $A$  stanno a quelle di  $B$  come 244140625 sta a 282429536481.

### 3.2 Probabilità: un nome ambivalente

Se si scorre la corrispondenza tra Fermat e Pascal, il *Traité* di Pascal o, ancora, il *De Ratiociniis* di Huygens, *non* si trova mai la parola probabilità. È noto che le parole possono cambiare significato nel corso del tempo, a volte in modo piuttosto drastico, come nel caso della parola *bravo* che aveva originalmente una connotazione negativa, derivando dal latino *pravus*, cioè cattivo; essa mantenne un significato negativo per lunghissimo tempo, al punto che i *bravi* descritti nei *Promessi Sposi* del Manzoni sono tali nell'accezione originaria del vocabolo, più che in quella moderna dove la bravura ha una connotazione sostanzialmente positiva, anche se spesso legata più all'abilità di compiere operazioni che non ad una valutazione di indole etica del soggetto. La probabilità appartiene a queste parole cangianti, con l'aggravante di racchiudere in sé due concetti non totalmente sovrapponibili: quello di probabilità *aleatoria* e di probabilità *epistemica* [13]. Il cambiamento di significato subito dalla parola probabilità è evidente se si legge un passo, scritto nel 1763 e divenuto celebre, dello storico e uomo politico inglese Edward Gibbon (1737-1794), in cui leggiamo, a proposito delle discrepanze tra i resoconti degli storici antichi sul passaggio di Annibale sulle Alpi:

ne concludiamo, allora, anche se con un certo scetticismo che, benché la narrazione di Livio sia più probabile, quella di Polibio è più veritiera.

Questo flagrante contrasto con la nostra idea che una probabilità crescente faccia avvicinare alla certezza nel realizzarsi di un evento, riflette una stratificazione di significati diversi, espressi dalla sola parola probabilità. In questa frase si riecheggia l'idea, ancora molto viva in tutto il XVII secolo, che una affermazione probabile fosse una affermazione suffragata dalla autorità di più testimoni, in questo caso di fonti storiche. Si tratta di un significato in uso presso i teologi scolastici, ritornato di gran moda appunto nel '600 grazie alla polemica tra i gesuiti ed esponenti di altri ordini religiosi circa il cosiddetto *probabilismo*. In breve, si trattava di fornire delle direttive sulla liceità o meno di una azione quando in proposito le autorità dei padri della chiesa erano discordanti. Secondo i probabilisti, se un certo comportamento morale aveva ricevuto l'approvazione di un numero esiguo ma non nullo di autorità, seguendolo non si sarebbe stati in errore. Il probabilismo fu attaccato da molti, in particolare da Pascal e Leibniz, due tra i testimoni della nascita del calcolo della probabilità come disciplina scientifica.

### 3.3 Intermezzo. Le tavole di Graunt.

Abbiamo visto come Huygens avesse sottolineato l'utilità del nuovo tipo di calcolo, al di là dei problemi legati ai giochi d'azzardo. Per esporre le applicazioni di Huygens del calcolo delle probabilità dobbiamo fare un inciso ed esporre i tratti essenziali dell'opera di John Graunt (1620-1674) *Natural and political observations made upon the Bills of Mortality* [7], pubblicata nel 1662, che può essere vista come il primo lavoro di statistica moderna.

Tra le diverse etimologie proposte per la parola statistica, una delle più accreditate la fa risalire alla parola italiana *stato*. La prima occorrenza si ha in un'opera che Girolamo Ghilini (1589-1668) completò poco prima della morte: il *Ristretto della civile, politica, statistica e militare scienza*. Se la parola statistica è relativamente recente, alcune delle tecniche di *statistica pratica*, coltivate sul campo, sono note da tempi remoti. Se seguiamo, ad esempio, il capitolo iniziale delle lezioni di statistica [6] tenute a Pavia nel 1880 da Antonio Gabaglio<sup>11</sup> (1840-1909) ci possiamo rendere conto che i *censimenti* furono utilizzati presso varie civiltà antiche: dai cinesi agli indiani, dal popolo ebraico ai romani. Le analisi dei beni di uno stato erano però descrittive, di natura *statica* e non contenevano alcuna predizione sull'andamento futuro di una nazione, limitandosi a scattare un'istantanea su cui nessun modello di natura matematica veniva eretto [14]. Queste caratteristiche della *statistica* proseguirono in tutto il Medioevo ed è solo con il Rinascimento italiano che si osservarono i primi timidi segnali di cambiamento, come il fatto che a Venezia, a quanto consta nel XV secolo, per la prima volta la popolazione venne conteggiata nel suo complesso, non solo limitandosi a contare quanti fossero in grado di pagar tasse o combattere un'eventuale guerra.

Tornando a Graunt, della sua vita non conosciamo molto: era figlio di un commerciante di tessuti presso cui lavorò per un certo tempo. Ricevuta una educazione tradizionale, arricchita dallo studio autodidatta di alcune discipline, Graunt fu comunque stimato nei circoli scientifici, al punto che riuscirà a procurare una cattedra in musica presso il Gresham College di Londra per l'amico William Petty (1623-1687). I *Bollettini di Mortalità* (Bills of mortality) erano elenchi pubblicati con cadenza settimanale dalle parrocchie londinesi sin dal 1530 con l'obiettivo di rilevare degli anomali incrementi di mortalità che potessero essere il segnale dell'inizio di una epidemia di peste. In questo modo, i dati potevano essere utilizzati sia dalle autorità per prendere misure contro il diffondersi del contagio, sia da quanti, sufficientemente agiati, potevano decidere di recarsi fuori città dove l'aria era più salubre e per questo meno soggetta al contagio, dal momento che si riteneva, erroneamente, che questo fosse dovuto ad un'infezione dell'aria. A partire del 1538 le parrocchie anglicane avevano l'obbligo di tenere traccia scritta di battesimi, matrimoni e funerali e ciò permette di avere almeno un'idea approssimativa della struttura della popolazione londinese, benché distorta dal fatto che i cittadini di confessione non anglicana

---

<sup>11</sup>Fu incaricato di tenere lezioni di Statistica presso l'università di Pavia e fu uno dei primi a ricoprire tale incarico. Le sue lezioni, tenute all'interno della Facoltà di Giurisprudenza, furono mutate anche dalla Facoltà di Scienze.

non risultassero censiti. Graunt fu il primo ad analizzare questa imponente mole di dati:

Essendo nato ed allevato alla città di Londra ed avendo osservato che molti tra quelli che consultano i [bollettini] di mortalità settimanali non se ne servivano che per guardare in fondo, se i funerali fossero aumentati o diminuiti; e, tra gli incidenti, [guardavano] quelli che erano raramente capitati nella settimana trascorsa e che apparivano straordinari: così esse potevano essere scelte come argomento di conversazione alla prima occasione; ciononostante, nel periodo della peste, come si diffonde o arretra il male, [i bollettini sono utili] affinché il ricco possa decidere sulla necessità di spostarsi, mentre i commercianti possono congetturare quali beni preferirebbero avere sui propri scaffali.<sup>12</sup> ([7], p. 14)

Graunt va oltre la curiosità spicciola dei più, interessati ai dati solo per scopi di—legittimo—tornaconto personale ma ne riconosce l'utilità per chi deve governare una città o uno stato e deve conoscere la composizione della popolazione ed il suo stato di salute per intraprendere eventuali azioni migliorative. In generale, scoprendo verità nascoste nei bollettini, Graunt auspicava di poter sfatare anche qualche luogo comune. I dati contenuti nei bollettini sono molti ed i primi ad essere analizzati da Graunt riguardano le cause di morte certificate che talora lasciano dei dubbi sull'effettiva capacità dei *searchers*, gli incaricati di verbalizzare le cause dei decessi, di distinguere nettamente tra un tipo di malattia ed un altro con sintomi simili. In generale Graunt osserva come i dati contenuti nei bollettini debbano essere interpretati, soprattutto quando si voglia capire la maggiore o minore virulenza di un'epidemia di peste rispetto ad un'altra. Anzitutto, i rapporti dei *searchers* erano imprecisi quando ascrivevano parte dei decessi all'età (*Aged*) dei malati, senza specificare quale fosse l'età dopo la quale si poteva essere considerati anziani. Similmente non viene specificata l'età degli infanti (*infants*) deceduti. Graunt è attento ad osservare come alcune cause di decesso possano essere raggruppate sotto un'unica voce, senza inficiare l'attendibilità dei bollettini:

Ancora, se qualcuno muore all'improvviso, non è molto importante che sia riportato nei bollettini come morte improvvisa o colpo apoplettico o per uno shock.<sup>13</sup>

La prima osservazione di Graunt sui 229250 decessi riportati nei bollettini è che il 36% dei neonati non raggiunge i sei anni di vita; inoltre, la presenza

<sup>12</sup>Having been born and bred in the City of London, and having observed that most of them who constantly took in the weekly Mortality made little other use of them, than to look at the foot, the burials increased or decreased; and, among the casualties, what happened rare and extraordinary in the week current: so as they take the same as a text to talk upon in the next company; and withal, in the Plague-time, how the sickness increased or decreased, that so the rich might judge of the necessity of their removal, and tradesmen might conjecture what doings they were like to have in their respective leadings.

<sup>13</sup>Again, if one dies suddenly, the matter is not great, whether it be reported in the Bills, *Suddenly*, *Apoplexy*, or *Planet-strucken*.

in una località di molte malattie croniche rende poco probabile trovarvi molte persone longeve. I dati su Londra indicavano a questo proposito che il 7% della popolazione moriva di vecchiaia. Pur con tutte le riserve su questo valore, Graunt scriveva:

da qui segue come, se in un altro posto, più di sette ogni cento superano i 70 anni, quel posto va ritenuto più salutare della nostra città di Londra.<sup>14</sup>

Inoltre, mentre i decessi relativi ad alcune malattie contagiose non mantengono un rapporto costante con il numero complessivo di decessi, ve ne sono un gran numero che godono di questa proprietà.

Un dato che colpisce Graunt (Cap. III) è che, a dispetto del gran numero di mendicanti che brulicavano per le vie di Londra, solo una minima percentuale dei decessi è dovuta ad inedia. Pertanto, secondo Graunt, lo stato dovrebbe farsene carico, visto che godono tutto sommato di una buona salute, garantendo loro una vita dignitosa ed insegnando loro un lavoro adatto alle loro capacità, per sottrarli alle condizioni di degrado (*debauchery*) in cui spesso si trovano. Affiora anche in questa circostanza l'uso politico della statistica. Con il capitolo IV Graunt affrontò i dati riguardanti le epidemie di peste a Londra, definendo *anno di peste* (*plague year*) ogni anno in cui si registrino almeno 200 decessi per peste.<sup>15</sup> Gli anni di epidemia per cui esistevano dati erano il 1592, 1593,<sup>16</sup> 1603, 1625 e 1636. I primi dati ad essere commentati erano quelli sul numero di decessi per peste ed il numero totale di decessi nello stesso periodo:

- da marzo a dicembre 1592 morirono 25886 persone, di cui 11503 per la peste;
- nel 1593 morirono 17844 persone di cui 10662 per la peste;
- nel 1603 morirono 37294 persone di cui 30561 per la peste;
- nel 1625 morirono, nello stesso periodo, 51758 persone di cui 35417 per la peste;
- da aprile a dicembre 1636 morirono 23359 persone di cui 10400 per la peste.

Graunt osserva subito che il dato *assoluto* dei morti per peste non basta per stabilire quale sia stata l'epidemia peggiore. Per rispondere a questa domanda, Graunt confronta il *rapporto* tra il numero di morti dovuti alla peste col numero totale di decessi, concludendo che l'epidemia peggiore fu quella del 1603 dove questo rapporto fu circa di 30 su 37, non lontano da  $\frac{4}{5}$ .

Un'analisi successiva, per determinare quale fu l'anno di massima mortalità all'incirca (*at large*), portò Graunt a confrontare i rapporti tra battesimi e decessi avvenuti negli anni di pestilenza: risulta allora che nel 1636 il rapporto fu di 2 a 5, nel 1592 fu di 1 a 6, nel 1603 e nel 1625 fu di circa 1 su 8. Commenta Graunt ([7], p. 32):

<sup>14</sup>it follows from hence, that if in any other country more than seven of the 100 live beyond 70 such country is to be esteemed more healthful than this of our City.

<sup>15</sup>Questa definizione si trova in effetti nel Cap. VI.

<sup>16</sup>Il 1592 ed il 1593 concorrono a formare un'unica pestilenza.

Abbiamo detto che il 1603 fu l'anno di massima pestilenza. Ed ora diciamo che questo medesimo anno non fu un anno di mortalità più grande del 1625.<sup>17</sup>

Perché questo ripensamento? L'argomento di Graunt è il seguente: nel 1625 sembrerebbero essere morti 35417 individui per la peste e 18848 per altre cause.<sup>18</sup> Ora, negli anni immediatamente precedenti e successivi il numero annuo di sepolture, dovute quindi a cause *diverse* dalla peste oscillava tra 7000 e 8000. Il picco a 18848 incontrato nel 1625 si spiega col fatto che circa 11000  $\simeq$  18000–7000 decessi ascritti a cause ordinarie erano in realtà dovuti alla peste: non si tratta di un errore grossolano ma di una possibilità concreta, dal momento che c'era una certa tendenza ad occultare ai *searchers* le reali cause di morte per evitare la quarantena con le sue pesanti conseguenze di segregazione dalla vita civile. In base a questa osservazione, Graunt conclude che i morti per peste sono da ritenersi 35000+11000, portando il totale a 46000 che mostra come non solo nel 1603, ma anche nel 1625 il rapporto tra decessi per peste e decessi totali si aggiri attorno ai  $\frac{4}{5}$ . Da questa analisi Graunt è portato a sospettare (*we may probably suspect*) che le cifre dei decessi per peste vadano aumentate di  $\frac{1}{4}$  rispetto ai dati "ufficiali". Esaminate le "code" degli anni di grande pestilenza, Graunt è propenso a ritenere come la pestilenza del 1603 protrasse i suoi effetti per otto anni e quella del 1636 per dodici anni. Nel capitolo V, Graunt mostrò come vi fosse una relazione tra l'aumento della virulenza del contagio e la diminuzione dei battesimi che costituivano una fonte ufficiale, benché incompleta, del numero di nascite. Le cause di questa relazione erano da ricercare nell'aumento della mortalità delle donne gravide ovvero nella fuga delle puerpere che potevano permettersi un soggiorno fuori Londra o, più verosimilmente, negli aborti o nei bambini nati morti. Terminato lo studio dei bollettini per quanto riguarda la peste, Graunt considerò nel Capitolo VI gli anni "malsani" (*sickly*), definiti come quegli anni che non sono di pestilenza ma nei quali il numero di sepolture supera sia quello dell'anno precedente che quello dell'anno successivo. In quegli anni, notò Graunt, tranne una sola eccezione, si registrava anche una diminuzione del numero di battesimi. Nel capitolo VII Graunt osservò che, a dispetto del fatto che in Londra il numero di battesimi fosse inferiore a quello delle sepolture, la popolazione cresceva, eccezion fatta per gli anni di pestilenza. Le cause dell'apparente anomalia erano il massiccio tasso di immigrazione verso la capitale ed il fatto che diverse categorie di lavoratori mantenevano la famiglia fuori dalla città. Nel Capitolo VIII Graunt notò un dato su cui molti altri rifletteranno in seguito, cioè la differenza tra il numero di maschi e di femmine, con una prevalenza dei primi sulle seconde. In effetti ciò che viene notato è la *stabilità del rapporto* tra numero di maschi e numero di femmine: vi sono in media 14 maschi ogni 13 femmine. Qui Graunt, a dispetto di altri studiosi, non

<sup>17</sup>We said that the year 1603 was the greatest Plague year. And now we say, that the same was not a greater year of mortality than the year 1625.

<sup>18</sup>La discrepanza con i dati precedenti potrebbe essere dovuta al fatto che i valori riportati prima si riferiscono al periodo marzo-dicembre.

cercò le cause del fenomeno ma ne deduce la ragionevolezza della monogamia, vista come garanzia di rispetto dei diritti femminili:

è una benedizione per l'umanità che, grazie a questo eccesso di maschi rispetto alle femmine, ci sia un ostacolo naturale alla poligamia dal momento che in quello stato le donne non possono vivere in quella parità ed uguaglianza nelle sostanze con i loro mariti, come è possibile da noi.<sup>19</sup> ([7], p. 41)

Dopo aver esaminato in dettaglio la crescita di Londra, nei Capitoli IX e X, Graunt illustrò nel Capitolo XI le tavole di mortalità, il suo lascito più importante alla comunità scientifica. A parte i 36 neonati su 100 che muoiono entro i sei anni di vita, Graunt raggruppò i decessi negli anni successivi in decenni. La sua tavola di mortalità è la seguente

anni alla morte	numero morti nel periodo su 100 unità
[0, 6]	36
(6, 16]	24
(16, 26]	15
(26, 36]	9
(36, 46]	6
(46, 56]	4
(56, 66]	3
(66, 76]	2
(76, 86]	1

e la complementare tavola di sopravvivenza è

età	sopravvissuti su 100 unità
6	64
16	40
26	25
36	16
46	10
56	6
66	3
76	1
86	0

che egli utilizzò per avere un'idea della stratificazione in età della popolazione londinese, prima di dedicarsi, nel capitolo XII, allo studio dei dati sulla popolazione relativi a piccoli centri rurali dove, per l'esiguità dei campioni, gli errori sono più rilevanti, eventualità di cui Graunt però non sembrò avvedersi.

<sup>19</sup>It is a blessing to mankind, that by this overplus of males there is this natural bar to polygamy: for in such a state women could not live in that parity and equality of expense with their husbands, as now and here they do.

### 3.4 La corrispondenza tra i fratelli Huygens

Anche se Christiaan Huygens aveva ricevuto una copia del lavoro di Graunt sin dal 1662, non se ne occupò sino a quando, nel 1669, il fratello Lodewijk cominciò ad interessarsi alle tavole di mortalità per determinare il valore delle rendite vitalizie. Egli scrisse al fratello di un proprio metodo per stimare l'aspettazione di vita di una persona, applicandolo a sé e a Christiaan che si mostrò scettico sulla verosimiglianza dei risultati di Lodewijk, per la mancanza di dati ulteriori che permettessero di determinare il numero annuo di decessi su un campione di 100 individui:

Avete fatto molto, avendo saputo svolgere il calcolo delle età, che dite di aver terminato. Affinché questo calcolo sia esatto bisognerebbe possedere una tavola che indichi anno per anno quante persone muoiono su 100 che vengono scelte e voi dovete aver surrogato questo dato in qualche modo. Attraverso quel dato potrò stabilire quanto deve vivere una persona di 6, 16 o 26 anni, ecc. che altrimenti non saprei ricavare ed ancor meno saprei stabilire l'età media di quelle persone, come avete determinato per voi e per me. Credo dunque che la vostra soluzione sia solo approssimata.<sup>20</sup> [[12], p. 484]

Christiaan aggiunse una osservazione che è il primo passo verso una trattazione probabilistica dei dati di Graunt.

Quel che posso concludere di certo dai dati delle tavole è che chi scommettesse che un bambino appena nato (o concepito, come dite voi, anche se mi sembra che l'Inglese non parli di concepiti anche perché non può tenerne il registro) vivrà fino a 16 anni, prenderebbe una cattiva decisione se scommettesse 4 contro 3. Similmente farebbe chi, scommettendo che una persona di 16 anni vivrà fino a 36, giocasse 4 contro 3.<sup>21</sup> ([12], p. 485)

Lodewijk ammise che il proprio metodo non era preciso ma lo espose nel dettaglio al fratello in una lettera del 30 ottobre 1669:

Riconosco che il mio calcolo delle età non è del tutto corretto ma c'è poco da aggiungere su questo, dal momento che la tavola inglese, su cui ci basiamo, non è esatta al proposito. (...) Ecco dunque il metodo che

---

<sup>20</sup>C'est beaucoup fait a vous, d'avoir peu faire le calcul des aages, dont vous dites eitre venu a bout. Mais a fin que ce calcul fuit exact il faudroit avoir une table qui marquai d'année en année combien il meurt des personnes de 100 qu'on suppose, et il faut que vous l'avez supplée par quelque moyen. Comme j'en scay pour cela, ou autrement vous ne scauriez determiner au vray, combien doit vivre une personne de 6, 16 ou 26 ans &c. et encore moins de quelque aage moyen entre ceux la, comme vous l'avez entrepris de vous et de moy. Je crois donc que vous n'en decidez qu'a peu pres.

<sup>21</sup>Ce que je puis conclure de certain par les donnez de la table c'est que qui gageroit qu'un enfant nouveau né (ou conceu comme vous dites, mais il me semble que l'Anglois ne parloit pas des conceus car comment en peut on tenir reginre) vivra a 16 ans, prendroit le mauvais party et hazarderoit 4 contre 3. De mesme qui gageroit qu'une personne de 16 ans vivra jusqu'a 36, il hazarde tout de mesme 4 contre 3.

ho utilizzato. Calcolo dapprima gli anni che queste 100 persone vivranno complessivamente, ottenendo in tutto 1822 anni.<sup>22</sup> ([12], p. 515)

Questa cifra è il risultato di una somma pesata: si suppone che i 36 bambini che muoiono entro il compimento dei 6 anni, muoiano a  $3 = \frac{0+6}{2}$  anni per cui il loro contributo è di  $3 \times 36 = 108$  anni. I 24 individui che muoiono tra 6 e 16 anni si suppone muoiano tutti ad  $11 = \frac{6+16}{2}$  anni per cui contribuiscono con  $24 \times 11 = 264$  anni, e così via sino ad esaurimento del campione. La somma degli anni così ottenuti è 1822 e viene divisa per il numero di individui del campione, 100, ottenendo una stima di poco più di 18 anni e 2 mesi come aspettativa di vita. Lodewijk espose il procedimento per calcolare l'aspettativa di vita per individui che hanno sei anni di età:

Tolgo anzitutto i 108 anni (l'età dei 36 neonati che muoiono sotto i sei anni) dal totale di 1822 anni; restano 1714 anni che debbono essere suddivisi tra le 64 persone ancora in vita, ciò che fa per ognuno, cioè per tutti i bambini di 6 anni, 26 anni e 10 mesi circa. Pertanto restano loro ancora da vivere, alla suddetta età di 6 anni, 20 anni e 10 mesi.<sup>23</sup> ([12], p. 516)

Ripetendo questo processo, Lodewijk ottenne una tabella in cui esprimeva l'aspettativa di vita residua per gli individui che hanno l'età riportata nelle tavole di Graunt:

età	anni da vivere	e mesi
16	20	3
26	19	4
36	17	6
46	15	0
56	12	8
66	8	4
76	5	0
86	0	0

Per chi non ha un'età presente nelle tavole, Lodewijk, che dichiara di non aver compreso l'argomento probabilistico del fratello, procedeva con una proporzione:

Quando voglio determinare l'età di una persona che ha tra i 36 e i 46 anni, ad esempio, come voi e me, regolo gli anni futuri in proporzione a quanto

<sup>22</sup>J'advoïe que mon calcul des aages n'est pas tout à fait juste mais il ij a si peu à dire que cela n'est aucunement considerable, et d'autant moins que la table Angloise, sur laquelle nous nous fondons, n'est pas dans cette dernière justesse aussi bien (...) Voijlà donc la méthode dont je me suis servij. Je compte premièrement les années que toutes ces 100 personnes ensemble doivent avoir vescu, qui font en tout 1822 années.

<sup>23</sup>J'oste premièrement les 108 ans (qui est l'aage des 36 enfans qui meurent au dessous des 6. ans) de tout ce nombre de 1822. ans; reste 1714 ans, lesquels doivent estre partagez entre les 64 personnes qui restent, ce qui fait pour chacun, c'est a dire pour chaque enfant de 6. ans, 26. ans et environ 10 mois de sorte qu'il leur reste encor à vivre au susdit aage de 6. ans, 20 ans et 10 mois.

è il loro eccesso o difetto rispetto al predetto numero di 36 e così via.<sup>24</sup>  
 ([12], p. 516)

Huygens rispose il 21 novembre contestando le conclusioni del fratello con un controesempio in cui considerava un campione di 100 individui, 90 dei quali vivono al più 6 anni mentre gli altri 10 vivono oltre 155 anni, in modo che la somma di tutte le età sia ancora 1822. In questo caso però, scommettere che un individuo del campione superi i 6 anni di vita è molto meno vantaggioso rispetto al caso precedente, dal momento che solo 10 individui riescono a vivere tanto a lungo: non è dunque ragionevole assumere la stima di 18 anni e 2 mesi come *aspettativa* di vita. Inoltre, tornando ai dati della tavola di mortalità di Graunt, Christiaan fece osservare al fratello come, chi avesse scommesso che un neonato avrebbe potuto superare i 16 anni di età, dovrebbe scommettere 2 contro 3 dal momento che solo 40 su 100 arrivano a quella età, per cui vi sono 40 casi favorevoli alla scommessa contro i 60 contrari. Anche per questo non è possibile che l'età di ciascun neonato arrivi al valore di 18 anni e 2 mesi indicato da Lodewijk ma deve arrestarsi a circa 11 anni. Senza indicarne la dimostrazione, Christiaan afferma di aver ottenuto l'aspettativa di vita per una persona di assegnata età grazie all'uso di una curva che riportò nella lettera del 28 novembre, entrando in ulteriori dettagli. In effetti, Christiaan ottenne il valore di 18 anni e 2 mesi per

la possibilità del neonato<sup>25</sup> ([12], p. 528)

cioè come l'aspettazione (*esperance*) di vita del neonato. Tuttavia, egli osservò come questo valore non rappresentasse l'età che verosimilmente il neonato raggiungerà: è possibile che egli muoia molto prima. Per comprendere il suo argomento, vediamo l'obiezione alla conclusione del fratello che, per un bambino di 6 anni, la durata attesa di vita sarà di altri 20 anni. Se dai dati di Graunt è vero che la speranza di vita ha questo valore, tuttavia

non ne potete concludere che, scommettendo che vivrà ancora 20 anni, la partita sarà equa.<sup>26</sup> ([12], p. 537)

Per avere un gioco equo infatti bisogna scommettere 25 contro 39 dal momento che 25 sono gli individui della coorte che raggiungono i 26 anni mentre 39 sono quelli che muoiono nell'intervallo compreso tra 6 e 26 anni. Una scommessa equa con eguale probabilità di successo e di insuccesso ha luogo, per interpolazione, quando si considerano individui di 16 anni e si scommette su quanti arriveranno a 31 anni. Christiaan distinse nettamente l'aspettativa di vita da

l'età alla quale c'è uguale verosimiglianza (*appareance*) di pervenire o non pervenire.<sup>27</sup> ([12], pp. 537-538)

<sup>24</sup>Lors que je veux déterminer l'aage d'une personne qui est entre 36 et 46 par exemple, comme vous et moi je règle leur années futures à proportion de cette qu'ils ont excédé plus ou moins le dit nombre de 36. et ainsi du reste.

<sup>25</sup>la chance de l'enfant conçu

<sup>26</sup>vous ne pouviez pas en conclure qu'en gageant qu'il vivroit encore 20 ans, la partie seroit égale.

<sup>27</sup>l'aage auquel il y a egale appareance qu'il parviendra ou ne parviendra pas.

In termini moderni ([12], p. 108), se  $\ell_n$  rappresenta il numero di individui della coorte di riferimento che sono arrivati all'età  $n$ , il numero  $t$  di anni più probabile di vita per uno di questi individui è la soluzione dell'equazione

$$\ell_{n+t} = \frac{1}{2}\ell_n. \quad (3.5)$$

Il linguaggio del gioco d'azzardo informa anche un altro problema sulla probabilità di sopravvivenza di più persone, che Christiaan pose all'attenzione del fratello: dati due individui di 16 anni, si vuole determinare la durata attesa  $T$  dell'intervallo di tempo dopo il quale entrambi saranno morti. Siccome i due individui hanno 16 anni, essi fanno parte dei 40 individui del campione di 100 che hanno raggiunto quell'età. Questi 40 individui avranno vite di lunghezze differenti: 15 moriranno entro 10 anni, 9 tra 10 e 20 anni, 6 tra 20 e 30 anni, 4 tra 30 e 40 anni, 3 tra 40 e 50 anni, 2 tra 50 e 60 anni ed 1 tra 60 e 70 anni. Assumendo che gli individui che muoiono con una età compresa in ciascuna di queste decadi, muoiano in effetti all'istante medio dell'intervallo, possiamo dire che, dei 40 individui che hanno raggiunto i 16 anni, 15 moriranno dopo 5 anni, 9 dopo 15 anni, 6 dopo 25 anni, 4 dopo 35 anni, 3 dopo 45 anni, 2 dopo 55 anni ed 1 dopo 65 anni. Huygens identifica queste quaranta durate di vita con quaranta biglietti di un'ipotetica lotteria che ciascuno degli individui estrae da un'urna, riponendolo dopo l'estrazione. Consideriamo uno dei due individui e supponiamo di sapere che ha estratto uno dei 15 biglietti che gli danno 5 anni di aspettativa di vita. Huygens calcolò, *dato* il risultato dell'estrazione del primo individuo, l'aspettativa di vita del secondo individuo. Questi potrebbe ancora estrarre uno di questi 15 biglietti e, supponendo uniforme la distribuzione dei decessi all'interno di ogni decennio, si può dire che egli ha 7.5 possibilità di sopravvivere fino a 5 anni ed altrettante di sopravvivere dai 6 ai 10 anni ovvero, per uniformità, di vivere ancora  $8 = \frac{6+10}{2}$  anni. Per il problema considerato, le prime 7.5 possibilità contano ciascuna per 5 anni di vita dal momento che, se anche visse di meno, il primo individuo vivrebbe comunque 5 anni e dunque il valore di  $T$  non sarebbe modificato. Infine, gli restano 25 possibilità di morire in un decennio successivo raggiungendo un'età attesa di 29.4 anni, valore quest'ultimo ottenuto calcolando la media ponderata delle età corrispondenti ai 25 "biglietti" restanti:

$$\frac{9 \times 15 + 6 \times 25 + 4 \times 35 + 3 \times 45 + 2 \times 55 + 1 \times 65}{25} = 29.4 \text{ anni}$$

Dunque, *sapendo* che il primo individuo ha estratto un biglietto che garantisce 5 anni di sopravvivenza, il periodo atteso  $T_1$  dopo il quale nessuno dei due individui sarà in vita è

$$T_1 = \frac{7.5 \times 5 + 7.5 \times 8 + 25 \times 29.4}{40} = 20.3 \text{ anni.}$$

Se il primo individuo "estrae" uno dei 9 biglietti che gli garantiscono una sopravvivenza di 15 anni, il secondo ha a questo punto 15 possibilità che gli garantiscono una sopravvivenza inferiore ai 15 anni e che, ai fini del computo di

$T$ , contribuiscono con 15 anni; 9 possibilità delle quali 4.5 gli garantiscono una sopravvivenza di 15 anni e le altre 4.5 una sopravvivenza di  $18 = \frac{16+20}{2}$  anni. Morendo in un decennio successivo, gli restano 16 possibilità che gli garantiscono una sopravvivenza di

$$\frac{6 \times 25 + 4 \times 35 + 3 \times 45 + 2 \times 55 + 1 \times 65}{16} = 37.5 \text{ anni.}$$

Sapendo che il primo individuo ha estratto un biglietto da 15 anni di sopravvivenza, il valore atteso del periodo  $T_2$  dopo il quale nessuno dei due individui sarà in vita è

$$T_2 = \frac{19.5 \times 15 + 4.5 \times 18 + 16 \times 37.5}{40} = 24.3 \text{ anni.}$$

Il procedimento può essere ora ripetuto calcolando i valori  $T_i$  ( $i = 3, 4, \dots, 7$ ) di  $T$  sapendo che il primo individuo ha estratto un biglietto che gli indica la sopravvivenza nel decennio  $i$ -esimo. Si ottiene

$$T_3 = 30.2 \text{ anni} \quad T_4 = 37.6 \text{ anni.}$$

$$T_5 = 46.1 \text{ anni} \quad T_6 = 55.3 \text{ anni.}$$

e, infine,

$$T_7 \simeq 65 \text{ anni.}$$

Per ricavare il valore di  $T$ , Huygens calcolò la media pesata dei valori  $T_i$  appena ottenuti:

$$T = \frac{15 \times 20.3 + 9 \times 24.3 + 6 \times 30.2 + 4 \times 37.6 + 3 \times 46.1 + 2 \times 55.3 + 1 \times 65}{40} = 29.22 \text{ anni.}$$

Il calcolo di Christiaan Huygens mostra in definitiva come ottenere un valore atteso a partire dalla conoscenza di valori attesi *condizionati*. Portare il punto di vista di uno scommettitore (*gager*) in questo ambito gli permise di distinguere tra valore di vita atteso e valore più probabile e costituì il suo contributo più innovativo.

### 3.5 Ulteriori sviluppi: le tavole di Breslavia

Come abbiamo visto, Graunt aveva notato che l'eccesso dei funerali sui battesimi a Londra, a dispetto della crescita della popolazione, fosse dovuto ad importanti flussi migratori che alteravano i dati. Inoltre mancavano informazioni precise sull'età dei deceduti né era nota la struttura anagrafica precisa della popolazione, cioè il numero di persone viventi di una data età. Tavole prive di questi difetti furono stilate per la città di Breslavia, in Slesia, dal pastore protestante Caspar Neumann (1648-1715). In esse erano registrati i dati raccolti dal 1687 al 1691 e relativi alle cinque parrocchie protestanti di quella città. Neumann le inviò

a Leibniz che, verosimilmente [10], informò la Royal Society di Londra, dove la mancanza di questo tipo di tavole era sentita. Il segretario della Società, Henri Justel (1620-1693), ottenne da Neumann copia delle tavole che gli furono inviate in due tornate. Alla morte di Justel, la lettera inviata da Neumann fu letta durante una seduta della Royal Society ed il materiale fu consegnato ad Edmund Halley (1656-1742) per uno studio approfondito. L'esito delle indagini di Halley fu pubblicato nel 1693 [9] in un lavoro dove egli illustrò diversi usi delle tavole di Neumann da cui ottenne la seguente tabella che presenta la distribuzione delle età nella popolazione di 34000 persone di Breslavia.

età	N	età	N	età	N	età	N	età	N	età	N
1	1000	8	680	15	628	22	586	29	539	36	481
2	855	9	670	16	622	23	579	30	531	37	472
3	798	10	661	17	616	24	573	31	523	38	463
4	760	11	653	18	610	25	567	32	515	39	454
5	732	12	646	19	604	26	560	33	507	40	445
6	710	13	640	20	598	27	553	34	499	41	436
7	692	14	634	21	592	28	546	35	490	42	427
età	N	età	N	età	N	età	N	età	N	età	N
43	417	50	346	57	272	64	202	71	131	78	58
44	407	51	335	58	262	65	192	72	120	79	49
45	397	52	324	59	252	66	182	73	109	80	41
46	387	53	313	60	242	67	172	74	98	81	34
47	377	54	302	61	232	68	162	75	88	82	28
48	367	55	292	62	222	69	152	76	78	83	23
49	357	56	282	63	212	70	142	77	68	84	20

La tavola poté essere utilizzata per stimare il numero di persone abili alla guerra: i maschi da 18 a 56 anni. Si considera la popolazione di età in questa fascia e si osserva che *almeno* la metà di questi sono maschi, per cui arruolabili. Più interessante è l'uso della tavola per stimare la *vitalità* della popolazione ad una certa età. Siano  $\ell_n$  ed  $\ell_{n+1}$  il numero di persone della coorte con età  $n$  ed  $n+1$  anni, rispettivamente. Allora la frazione

$$\frac{\ell_{n+1}}{\ell_n - \ell_{n+1}}$$

dà il rapporto tra le possibilità di sopravvivere un anno e di morire entro un anno: così, un individuo di 30 anni può scommettere 523 contro 8 che raggiungerà i 31 anni di età. Similmente, se si considera  $\ell_{n+t}$ , il rapporto

$$\frac{\ell_{n+t}}{\ell_n - \ell_{n+t}}$$

misura il rapporto tra le possibilità di sopravvivere  $t$  anni e di morire entro  $t$  anni per un individuo che oggi ha un'età di  $n$  anni. Per esempio, le possibilità che un individuo di  $n = 50$  anni, viva altri  $t = 10$  anni rispetto a quelle che

egli muoia entro 10 anni sono 104 (=346-242) su 242. Per potere trarre simili conclusioni Halley doveva avere la certezza che la struttura della popolazione non subisse fluttuazioni annue eccessive. Come Huygens, Halley considerava l'età più probabile raggiungibile da un individuo di  $n$  anni come la soluzione  $t$  dell'equazione (3.5) ed utilizzava il rapporto  $\frac{\ell_{n+t}}{\ell_n}$  per stabilire il prezzo di una polizza di assicurazione sulla vita. Quanto alle rendite vitalizie, Halley dichiarò quale dovesse essere il principio su cui fondarne il pagamento:

è evidente che l'*acquirente* dovrebbe pagare solo tanta parte del valore della *rendita* quante sono le possibilità che sia ancora in vita; e ciò andrebbe calcolato annualmente e, sommati tutti quei valori annui, si otterrebbe il valore della *rendita* per la *vita* della persona proposta.<sup>28</sup> ([9], p. 602)

Halley fornì per l'importo di una rendita nel caso di un tasso di interesse  $i$  annuo l'espressione

$$\frac{1}{\ell_n} \sum_{t=1}^{\omega-n-1} (1+i)^{-t} \ell_{n+t} \quad (3.6)$$

in cui  $\omega$  è l'età tale che,  $\forall n \geq \omega$ ,  $\ell_n \equiv 0$ . La quantità  $(1+i)^{-t}$  rappresenta l'importo da versare *oggi* per ottenere tra  $t$  anni una sterlina, in regime di capitalizzazione composta, al tasso annuo costante dell' $i\%$ . Per capire l'equazione (3.6), osserviamo che, se  $n$  è l'età del contraente al momento della stipula del contratto e  $C_0$  il capitale versato, con un tasso di interesse annuo dell' $i\%$ , dopo un anno il capitale diverrà

$$C_1 = C_0 + iC_0 = C_0(1+i);$$

esso viene reinvestito producendo dopo un altro anno, quindi dopo due anni dalla stipula del contratto, il capitale

$$C_2 = C_1 + iC_1 = C_1(1+i) = C_0(1+i)^2.$$

Nelle rendite vitalizie si *fissa* la quantità di denaro che si vuole garantire al contraente dopo ciascun anno. Se fissiamo che  $C_1 = 1$  allora occorre versare inizialmente la quota

$$\frac{1}{1+i} < 1$$

che viene della *sconto*. Se si vuole garantire *anche* dopo due anni la fruizione di un capitale unitario lo sconto, ovvero il premio da versare all'inizio del contratto, è

$$\frac{1}{(1+i)^2} < 1$$

e così via per gli anni successivi.

<sup>28</sup>It is plain that the *Purchaser* ought to pay for only such a part of the value of the *Annuity*, as he has chances that he is living; and this ought to be computed yearly, and the sum of all those yearly values being added together, will amount to the value of the *Annuity* for the *Life* of the person proposed.

Se il contraente fosse certo di essere ancora in vita allo scadere del contratto, si assicurerebbe una rendita vitalizia versando il premio

$$\frac{1}{1+i} + \frac{1}{(1+i)^2} + \dots$$

e questo prezzo della rendita è detto *rendita annua certa*. Nella formula di Halley, il premio da corrispondere per avere un capitale unitario dopo  $t$  anni viene ulteriormente scontato moltiplicandone l'importo per la frazione  $\frac{\ell_{n+t}}{\ell_n}$  che dà la probabilità empirica, cioè desunta dalle tabelle, che un individuo di età  $n$  anni ha di sopravvivere altri  $t$  anni.

A fornire maggiori dettagli fu Abraham de Moivre nel 1725 [3] che illustrò il calcolo di una rendita vitalizia che fruttava un interesse annuo del 5%. Se vogliamo concedere al beneficiario 100 sterline dopo un anno occorre *scontare* tale cifra al tasso nominale del 5%, chiedendo il versamento di un premio di

$$\frac{100}{(1+0.05)^1} \simeq 95.23 \text{ sterline.}$$

Se vi fosse la *certezza* che il beneficiario sarà vivo al momento del riscatto, questo sarebbe il prezzo da pagare. Tuttavia, come aveva già scritto Halley, il beneficiario è tenuto a pagare solo una parte proporzionale alle possibilità di essere ancora in vita. Nell'esempio di De Moivre, il contraente ha 30 anni e la probabilità empirica di restare ancora in vita un anno è, secondo la tavola di Halley,  $\frac{523}{531}$  per cui il valore della rendita su un anno è

$$95.23 \times \frac{523}{531} \simeq 93.80.$$

Similmente, il valore scontato di 100 sterline per goderne dopo due anni dalla stipula del contratto sarebbe

$$\frac{100}{(1+0.05)^2} \simeq 90.70 \text{ sterline}$$

e siccome la probabilità empirica che un individuo di 30 anni sia ancora vivo due anni dopo è  $\frac{513}{531}$ , il valore da versare è

$$90.7 \times \frac{513}{531} = 87.90 \text{ sterline.}$$

Per ottenere il valore complessivo della rendita occorre sommare i vari prezzi scontati, modulati con le rispettive probabilità di sopravvivenza, ottenendo la formula di Halley (3.6). De Moivre osservò la regolarità del numero di adulti deceduti annualmente presente nella tavola di Halley. Questo dato venne riletto da De Moivre in termini di probabilità di sopravvivenza di un individuo di una data età e lo portò a formulare una *ipotesi di lavoro*:

Supponendo che le probabilità di restare in vita decrescano in progressione aritmetica, trovare il valore corrispondente di una rendita vitalizia.<sup>29</sup> ([3], p. 11)

<sup>29</sup>Supposing the probabilities of life to decrease in Arithmetic progression, to find the value of an Annuity upon such a life.

Proponendo un *modello* sulla legge di mortalità, De Moivre forniva una base razionale di indagine ed una conseguente semplificazione dei calcoli. Egli introdusse il montante unitario  $r = 1 + i$  (interesse annuo di una sterlina sommato al capitale di una sterlina) ed il numero  $m$  di anni che intercorrono tra l'età del beneficiario e l'età massima raggiungibile da un individuo che appartiene alla coorte di riferimento che, per le tavole di Halley, è 86 anni. Se, al momento della stipula del contratto, vi sono  $a$  individui nella popolazione che hanno l'età del contraente, la ragione della progressione aritmetica sarà  $-\frac{a}{m}$  e le probabilità del contraente di sopravvivere 1, 2, 3, ...,  $m$  anni sono, per l'ipotesi fatta,

$$\frac{a - \frac{a}{m}}{a} = \frac{m-1}{m}, \quad \frac{a - \frac{2a}{m}}{a} = \frac{m-2}{m}, \quad \frac{m-3}{m}, \quad \dots, \quad \frac{m-m}{m}.$$

Se il contraente fosse certo di essere in vita dopo 1, 2, ...  $m$  anni, il prezzo scontato di una rendita vitalizia unitaria sarebbe, rispettivamente

$$\frac{1}{r}, \quad \frac{1}{r^2}, \quad \frac{1}{r^3}, \quad \frac{1}{r^4}, \quad \dots.$$

Questi prezzi vanno ridotti moltiplicandoli per la probabilità che l'individuo sia in vita dopo 1, 2, 3, ... anni. I singoli termini diventano

$$\frac{m-1}{mr}, \quad \frac{m-2}{mr^2}, \quad \frac{m-3}{mr^3}, \quad \dots, \quad \frac{m-m}{mr^m}$$

e il valore della rendita vitalizia è

$$\mathcal{Q} := \sum_{k=1}^{m-1} \frac{m-k}{mr^k} = \sum_{k=1}^{m-1} \frac{1}{r^k} - \sum_{k=1}^{m-1} \frac{k}{mr^k}. \quad (3.7)$$

La prima somma, che indichiamo con  $\mathcal{P}$ , rappresenta la *rendita certa* (*annuity certain*), cioè quanto si dovrebbe versare per ottenere un vitalizio unitario nella certezza di sopravvivere  $m-1$  anni. Eseguendo il calcolo meno semplice di  $\mathcal{Q}$ , De Moivre ottenne

$$\mathcal{Q} = \frac{1 - \frac{r}{m}\mathcal{P}}{r-1}. \quad (3.8)$$

Se la rendita si prolunga per un intervallo di  $n$  anni, all'inizio del quale la popolazione è di  $b$  individui che si riducono a  $a < b$  al termine dell'intervallo, sempre supponendo un decremento della popolazione secondo una progressione aritmetica, di ragione  $\frac{a-b}{n} < 0$ , negli  $n$  anni in cui è attivo il contratto, la popolazione si riduce secondo la legge

$$b, \frac{nb-b+a}{n}, \frac{nb-2b+2a}{n}, \frac{nb-3b+3a}{n}, \dots, a$$

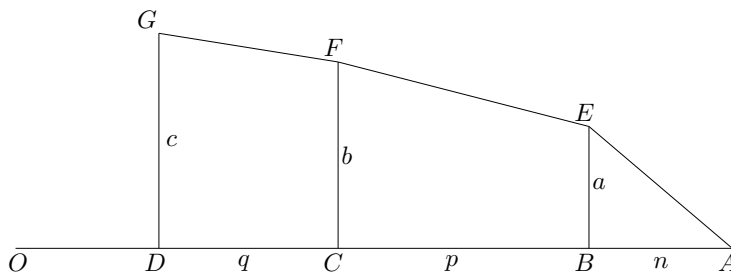
e dunque le probabilità di sopravvivenza del contraente 1, 2, ...,  $n$  anni dopo la stipula del contratto sono

$$\frac{nb-b+a}{nb}, \frac{nb-2b+2a}{nb}, \frac{nb-3b+3a}{nb}, \dots, \frac{a}{b}.$$

Utilizzando il procedimento che ha portato all'equazione (3.8), De Moivre concluse che il valore della rendita per il periodo proposto è

$$Q + \frac{a}{b}(P - Q). \quad (3.9)$$

Per ottenere un migliore accordo con le tavole, De Moivre mostrò come adattare il proprio metodo di calcolo al caso in cui l'ipotesi di un tasso costante di riduzione della probabilità di sopravvivenza fosse verificata solo a tratti. Per questo egli affrontò il problema di trovare il valore di una rendita, attiva in un certo lasso temporale, all'inizio del quale la popolazione è di  $c$  unità. De Moivre considerò un grafico in cui  $OA$  è la durata massima della vita nella coorte mentre all'interno degli intervalli  $AB$ ,  $BC$ ,  $CD$ , ... che hanno durata rispettivamente pari ad  $n$ ,  $p$ ,  $q$ , ... anni, la mortalità annua è costante.



Infine,  $BE = a$ ,  $CF = b$ ,  $DG = c$  rappresentano il numero di individui che sono vivi agli istanti  $B$ ,  $C$  e  $D$ , mentre in  $A$  nessun individuo della popolazione è vivo. Per calcolare il valore della rendita vitalizia, De Moivre procedette dall'ultimo intervallo  $AB$  sino al primo,  $DC$ . Il valore della polizza nell'intervallo  $AB$  è dato dall'equazione (3.8), ed è detto  $A$  da De Moivre. Questo valore va ora *scontato* sul periodo  $BC$  pari a  $p$  anni, per cui la rendita certa nell'intervallo di vita  $BC$  è  $\frac{A}{r^p}$  e quindi la rendita vitalizia per il tratto  $AB$ , stimata a partire da  $C$ , è  $\frac{a}{b} \frac{A}{r^p}$ , avendo moltiplicato il valore della rendita certa per la probabilità di sopravvivenza  $\frac{a}{b}$  che un individuo, di età misurata dal segmento  $OC$ , ha di sopravvivere i  $p$  anni che lo portano all'età rappresentata dal punto  $B$ . Per ottenere una rendita che copra tutto l'intervallo  $AC$ , occorre aggiungere al valore precedente quello della rendita (3.9), ora scritto come  $Q + \frac{a}{b}(P - Q)$ , che copre l'intervallo  $BC$ . La procedura viene iterata fino ad esaurire gli intervalli entro cui si è suddiviso  $OA$ . In altri problemi De Moivre passò al continuo per calcolare i valori delle rendite vitalizie in particolari condizioni.

Sottolineiamo la natura *empirica* del concetto di probabilità che entra nelle considerazioni sulle rendite vitalizie. Fu senza dubbio l'uso da parte di Huygens dello schema della lotteria per calcolare del tempo di vita atteso per la morte di due individui a gettare un ponte tra il nascente calcolo delle probabilità (o delle aspettative) e la matematica attuariale (delle assicurazioni). È opportuno però ricordare che la base razionale fornita da De Moivre al calcolo delle rendite vitalizie non entrò immediatamente nell'uso degli addetti ai lavori. Occorse del tempo perché si abbandonassero schemi empirici, consolidati dall'uso,

per rivolgersi all'impostazione matematica, di cui abbiamo ripercorso le prime tappe.

# Bibliografia

- [1] J. Bernoulli: *Ars Conjectandi, opus posthumum*, Basel, (1713).
- [2] L. Daston: *Classical probability in the Enlightenment*. Princeton University Press, Princeton, New Jersey, U.S.A., (1988).
- [3] A. De Moivre: *Annuities upon lives*. W. P., London, (1725).
- [4] A. De Moivre: *Miscellanea analytica de seriebus et quadraturis*. Tonson & Watts, London, (1730).
- [5] P. Dupont, C.S. Roero: Il trattato *De ratiociniis in ludo aleae* di Christiaan Huygens con le *Annotationes* di Jakob Bernoulli (*Ars Conjectandi*, Parte I) Presentati in traduzione italiana, con commento storico-critico e risoluzioni moderne. *Memorie della Accademia delle Scienze di Torino*, **8** (S. V), 1-258, (1984).
- [6] A. Gabaglio: *Storia e teoria generale della statistica*. Hoepli, Milano, (1880). Ristampato da Kessinger Legacy Reprints.
- [7] J. Graunt: *Natural and political Observations made upon the Bills of Mortality*, (1662). In *Journal of the Institute of Actuaries*, **90**, (1964), 1-61. Con una premessa di B. Benjamin.
- [8] A. Hald: *History of probability and Statistics and their applications before 1750*. Wiley-Interscience, Hoboken, New-Jersey, U.S.A. (1988).
- [9] E. Halley: An Estimate of the Degrees of the Mortality of Mankind, Drawn from Curious Tables of the Births and Funerals at the City of Breslaw; With an Attempt to Ascertain the Price of Annuities upon Lives. *Philosophical Transactions of the Royal Society of London*, **17**, 596-610, (1693).
- [10] C.C. Heyde, E. Seneta: *Statisticians of the centuries*. Springer, New York, ((2001).
- [11] C. Huygens: *De Ratiociniis in Ludo Aleae*. In [13], pp. 52-91.
- [12] C. Huygens: *Oeuvres complètes de Christiaan Huygens* Tome VI. Correspondance 1666-1669. Nijhoff, La Haye, (1895).

- [13] C. Huygens: *Oeuvres complètes de Christiaan Huygens* Tome XIV. Calcul des probabilités. Travaux de mathématiques pures. 1655-1666. Nijhoff, La Haye, (1920).
- [14] M.G. Kendall: Studies in the history of probability and statistics. X. Where shall the history of statistics begin? *Biometrika*, **47**, (1960), 447-449.

## Capitolo 4

# Nuovi giochi e nuovi metodi di soluzione

In questo capitolo esaminiamo alcune soluzioni a dei problemi di calcolo delle probabilità ottenute nel XVIII secolo ed all'inizio del XIX secolo per mostrare l'evoluzione subita dagli strumenti analitici impiegati. Ci dedicheremo anzitutto al *gioco delle concordanze* (*jeu de rencontre*) o gioco del tredici studiando sia la soluzione di Pierre Rémond de Montmort (1678-1719) che quelle proposte da De Moivre (1667-1754) ed Eulero (1707-1783), con le applicazioni proposte da due scienziati: l'alsaziano Johann-Heinrich Lambert (1728-1777) e l'inglese Thomas Young (1773-1829). Vedremo poi come De Moivre risolse il problema della durata del gioco, una variante del problema della rovina del giocatore proposto da Pascal ed Huygens e considereremo un gioco di strategia, *le her*, trattato da Montmort, che sarà riscoperto all'inizio del XX secolo nell'ambito della teoria dei giochi. Infine, daremo conto della soluzione del problema della ripartizione della posta tra tre o più giocatori proposta da De Moivre.

### 4.1 Il gioco delle concordanze

Come abbiamo visto, il matematico francese Pierre Rémond de Montmort pubblicò due edizioni dell'*Essay d'Analyse sur les Jeux de Hazard*, nel 1708 e nel 1713, quest'ultima successiva alla pubblicazione dell'*Ars Conjectandi* e del *De mensura sortis* di De Moivre. Appartenente ad una nobile famiglia, Montmort [4] fuggì di casa per non essere costretto ad assecondare i voleri del padre che voleva farne un magistrato. Viaggiò in Inghilterra, Olanda e Germania dove, a casa di un cugino, lesse *La recherche de la vérité* di Nicholas de Malebranche. A seguito di quella lettura subì una sorta di conversione che lo spinse a riconciliarsi col padre, tornando a casa nel 1699. La morte del padre, avvenuta quando Montmort aveva 22 anni, lo obbligò ad occuparsi delle cospicue ricchezze ereditate ma non gli impedì di proseguire gli studi in matematica sotto la guida dello stesso Malebranche. Sposatosi nel 1704, si ritirò nella tenuta di campagna di

Montmort, in Borgogna, dove iniziò a lavorare al calcolo delle probabilità. La motivazione dietro la scelta di questo tema di studio può essere rintracciata tra le righe della Prefazione alla prima edizione dell'*Essay* ([19], p. iii):

Da molto tempo i matematici si vantano di poter scoprire coi loro metodi tutte le verità presenti nelle scienze naturali che sono alla portata dell'intelletto umano; è fuor di dubbio che, grazie alla meravigliosa alleanza che la matematica ha stretto con la fisica dagli ultimi cinquant'anni, le persone sono obbligate a riconoscere che, quanto detto a favore della matematica non è privo di fondamento. Quale gloria sarà per questa scienza se potrà anche servire a regolare i giudizi e guidare gli uomini nella vita di ogni giorno!<sup>1</sup>

Non si tratta di retorica ma della consapevolezza che questo desiderio è ormai a portata di mano grazie al lavoro di Jakob Bernoulli, di cui Montmort era a conoscenza quantomeno dai resoconti scritti da Bernard le Bovier de Fontenelle e Joseph Saurin ed apparsi sulla *Historie de l'Académie des Sciences* e sul *Journal des Sçavants*, rispettivamente. Anche se Montmort non osò cimentarsi nelle applicazioni della probabilità alle questioni economico-giuridiche, lo studio dei giochi d'azzardo offriva un terreno ideale per la missione della scienza di cacciare i fantasmi della superstizione grazie ad un corretto uso della ragione:

È soprattutto nei giochi d'azzardo che si evidenzia la debolezza dello spirito umano e la sua inclinazione alla superstizione. Non vi è nulla di più comune che vedere dei giocatori attribuire la propria sfortuna alle persone che li avvicinano e ad altre circostanze che non sono meno indifferenti allo svolgersi del gioco. (...) Altri (...) si ostinano a prendere le carte che hanno perso, ritenendo che, avendo perso più volte, sia meno verosimile che possano perdere ancora, come se il passato potesse decidere qualcosa circa il futuro.<sup>2</sup> ([19], p. vi)

Per Montmort non vi è dubbio che le leggi di natura siano deterministiche e che, in senso stretto, non vi sia posto per il caso, inteso come forza cieca che regola l'evoluzione dell'universo. Il caso entra in gioco per la limitatezza delle facoltà umane che non riescono a cogliere i nessi causali presenti nel progetto della creazione:

---

<sup>1</sup>Il y a long-temps que les geometres se vantent de pouvoir par leurs méthodes decouvrir dans les Sciences naturelles, toutes les vérités qui sont à la portée de l'esprit humain; et il est certain que par le merveilleux alliage qu'ils ont fait depuis cinquante ans de la Geometrie avec la Physique, ils ont forcé les hommes à reconnoître que ce qu'ils disent à l'avantage de la Geometrie n'est pas sans fondement. Quelle gloire seroit-ce pour cette Science si elle pouvoit encore servir à regler les jugemens et conduire des hommes dans la pratique des choses de la vie!

<sup>2</sup>C'est particulièrement dans les Jeux de hazard que paroît la foiblesse de l'esprit humain et la pente qu'il a à la superstition. Rien n'est si ordinaire que de voir des joueurs attribuer leur malheur aux personnes qui le approchent, et à d'autres circonstances qui ne sont pas moins indifferentes aux événements du jeu. (...) D'autres (...) s'attachent à prendre les cartes perdantes, dans l'opinion qu'ayant plusieurs fois perdu, il est moins vrai-semblable qu'elles perdront encore, comme si le passé pouvoit décider quelque chose pour l'avenir.

A essere precisi, nulla dipende dal caso; quando si studia la natura, ci si convince ben presto che il suo Autore agisce in modo generale ed uniforme, che reca il tratto di una saggezza e una prescienza infinita. Pertanto, per legare a questa parola *caso* un'idea che sia conforme alla vera filosofia, bisogna pensare che, benché tutte le cose siano rette da leggi certe, il cui ordine ci è molto spesso sconosciuto, dipendono dal caso quelle la cui causa naturale ci è nascosta. Con questa definizione possiamo dire che la vita dell'uomo è un gioco in cui regna il caso.<sup>3</sup> ([19], p. xiv)

Dopo questa breve premessa generale sull'impostazione dell'*Essay* che può vedersi come un'aggiornata rassegna del trattamento probabilistico dei principali giochi d'azzardo, passiamo ad analizzare il primo studio sul gioco delle concordanze, che Montmort chiama *gioco del tredici*, un gioco di carte che troverà posto fisso, con le sue varianti, nella letteratura sul calcolo delle probabilità. Le regole sono chiaramente esposte da Montmort: si abbia un numero di giocatori qualsivoglia, uno dei quali (*Pierre*) sia di turno a distribuire le carte. Il gioco viene effettuato con un mazzo di 52 carte rimescolate a caso; *Pierre* le scopre in successione pronunciando il numero 1 quando scopre la prima carta, due quando scopre la seconda e così via fino a 13: il ruolo dei numeri 11, 12, 13 viene svolto, nell'ordine, da Fante, Regina e Re. Se *nessuna* delle 13 carte estratte reca il numero pari alla sua posizione, *Pierre* deve pagare a ciascuno degli altri giocatori quanto questi aveva scommesso e passerà la mano a chi si trova alla sua destra. Al contrario, se *Pierre* estrae una carta il cui valore coincide con la posizione nell'estrazione (ad esempio, se la sesta carta estratta è il numero sei), allora *Pierre* vince tutto il denaro scommesso dagli altri giocatori e tiene la mano del gioco, ricominciando il processo di estrazione. Se, ad un certo punto, le carte terminano, per esempio, con la settima di un gruppo di 13 carte, il mazzo viene rimescolato e si riparte pronunciando il numero otto e così via. Questa è la descrizione del gioco ma, al momento di farne una trattazione matematica, Montmort affrontò un problema più semplice in cui i giocatori sono solo due *Pierre* e *Paul* ed il numero di carte varia da due a tredici, tutte dello stesso seme. L'obiettivo, qui come in generale nell'*Essay*, è vedere quanto il gioco si allontani dalla condizione di essere equo. Montmort procede per gradi, partendo dal caso più semplice in cui le carte sono solo due, numerate con 1 e 2. In questo caso vi sono solo due sequenze possibili: (1, 2) e (2, 1) e solo la prima fa vincere *Pierre* che ha una probabilità di successo pari ad  $\frac{1}{2}$ . Se ci sono tre carte (1, 2, 3) si hanno  $6 = 3!$  possibilità. Per calcolare il numero di quelle favorevoli a *Pierre*, Montmort usa questo argomento di *inclusione-esclusione*: sono favorevoli a *Pierre* le sequenze che hanno 1 in prima posizione: esse sono  $2! = (3 - 1)! = 2$ ; quelle che *non* hanno 1 in prima posizione ma hanno 2 in

<sup>3</sup>A parler exactement, rien ne dépend du hazard; quand on étudie la nature, ou est bientôt convaincu que son Auteur agit d'une maniere generale et uniforme, qui porte le caractere d'une sagesse et d'une prescience infinie. Ainsi pour attacher à ce mot *hazard* une idée qui soit conforme à la vraye Philosophie, on doit penser que toutes choses étant réglées selon des loix certaines, dont le plus souvent l'ordre ne nous est pas connu, celles-là dépendent du hazard dont la cause naturelle nous est cachée. Après cette définition on peut dire que la vie de l'homme est un jeu où regne l'hazard.

seconda posizione ve ne è solo una; la sola permutazione che non ha  $n\acute{e}$  1 in prima posizione  $n\acute{e}$  2 in seconda posizione ma ha 3 in terza posizione. In definitiva, quattro delle sei permutazioni possibili di  $(1, 2, 3)$  sono favorevoli a *Pierre* che quindi ha una probabilità di successo uguale a  $\frac{2}{3}$ . Si delinea dunque il primo passo nel computo dei casi favorevoli a *Pierre* perché, se le carte sono  $n$ , le permutazioni che iniziano con 1 sono  $(n-1)!$ . Ora, vi sono anche  $(n-1)!$  permutazioni di  $n$  che hanno come secondo elemento 2 *ma*, per non contare più volte una stessa permutazione, occorre sottrarre quelle che hanno anche 1 in prima posizione. Poiché le permutazioni di  $n$  elementi del tipo  $(1, 2, \dots)$  sono  $(n-2)!$ , i casi favorevoli al successo con la seconda carta sono  $(n-1)! - (n-2)!$ . Dalle  $(n-1)!$  permutazioni che hanno 3 come terzo elemento occorre sottrarre quelle del tipo  $(1, *, 3, \dots)$  e quelle del tipo  $(*, 2, 3, \dots)$  che, in tutto, sono  $2(n-2)!$  ed *aggiungere* quelle del tipo  $(1, 2, 3, \dots)$  che sono state sottratte *due* volte. Queste ultime sono  $(n-3)!$  e dunque abbiamo

$$(n-1)! - 2(n-2)! + (n-3)!$$

permutazioni di  $n$  elementi per le quali la *prima* coincidenza tra rango di estrazione e numero della carta avviene con il numero 3. Come ultimo esempio, consideriamo il caso in cui la prima coincidenza si abbia con la carta 4. Si parte sempre dalle  $(n-1)!$  permutazioni che hanno 4 in posizione 4. A queste occorre togliere quelle del tipo  $(1, *, *, 4, \dots)$ ,  $(*, 2, *, 4, \dots)$  e  $(*, *, 3, 4, \dots)$  che in tutto sono  $3(n-2)! = (4-1)(n-2)!$ . Si ripristinano tutte le permutazioni  $(1, 2, *, 4, \dots)$ ,  $(1, *, 3, 4, \dots)$  e  $(*, 2, 3, 4, \dots)$ : ciascun insieme è formato da  $(n-3)!$  elementi ed in tutto ve ne sono  $3(n-3)! = \binom{4-1}{2} (n-3)!$ . Infine occorre *togliere* le permutazioni del tipo  $(1, 2, 3, 4, \dots)$  che sono in numero di  $(n-4)!$ : in tutto le permutazioni favorevoli *distinte* dalle precedenti sono

$$(n-1)! - 3(n-2)! + 3(n-3)! - (n-4)!$$

In generale il numero  $\alpha_{m,n}$  di permutazioni per le quali la *prima* coincidenza tra rango di estrazione e numero della carta è il numero  $m$ , in un mazzo di  $n$  carte, è

$$\begin{aligned} \alpha_{m,n} := & (n-1)! - \binom{m-1}{1} (n-2)! + \binom{m-1}{2} (n-3)! - \dots + \\ & + (-1)^{m-1} \binom{m-1}{m-1} (n-m)!. \end{aligned} \tag{4.1}$$

L'equazione (4.1) non figura in nessuna delle edizioni dell'*Essay*. Nella prima edizione, Montmort si limitò a dare il risultato, senza fornire traccia della dimostrazione. Il ragionamento discusso sopra compare, per sommi capi, nella seconda edizione ed era stato comunicato da Nicholas Bernoulli a Montmort nel 1710 in una delle lettere che costituiscono l'appendice di [19], dove troviamo anche l'espressione dell'aspettazione (*sors*), cioè la probabilità di vittoria, di

*Pierre* quando si considera la prima concordanza al posto  $m$ :

$$S_{m,n} = \frac{\alpha_{m,n}}{n!} = \frac{1}{n} - \binom{m-1}{1} \frac{1}{n(n-1)} + \binom{m-1}{2} \frac{1}{n(n-1)(n-2)} - \cdots +$$

$$+ (-1)^{m-1} \binom{m-1}{m-1} \frac{1}{n(n-1)\cdots(n-m+1)}.$$

Infine, la probabilità che *Pierre* vinca, *indipendentemente* dalla posizione in cui si manifesta la prima concordanza, purché entro le prime  $m$  carte, è

$$S_n(m) := \sum_{\mu=1}^m S_{\mu,n} = \frac{1}{n} \sum_{\mu=1}^m \binom{\mu-1}{0} - \frac{1}{n(n-1)} \sum_{\mu=2}^m \binom{\mu-1}{1} +$$

$$+ \frac{1}{n(n-1)(n-2)} \sum_{\mu=3}^m \binom{\mu-1}{2} + \cdots + (-1)^{m-1} \frac{1}{n(n-1)\cdots(n-m+1)} \sum_{\mu=m}^m \binom{\mu-1}{m-1}.$$

Per il calcolo di  $\sum_{\mu=k+1}^m \binom{\mu-1}{k}$  possiamo tornare alla disposizione del triangolo aritmetico—che per primo Montmort chiamò triangolo di Pascal—già vista nel Cap. I ed osservare che gli addendi che figurano in questa somma sono disposti sulla riga<sup>4</sup>  $k+1$ . Inoltre, notiamo che gli addendi partono sempre dal termine 1 con cui la riga inizia. Grazie alla regola di formazione dell'elemento  $a_{n,k}$  del triangolo aritmetico, la somma dei primi  $m-k-1$  termini della riga  $k+1$ -esima coincide con l'elemento sulla riga  $k+1$ -esima e la colonna  $m$ , per cui

$$\sum_{\mu=k+1}^m \binom{\mu-1}{k} = \binom{m}{k+1}$$

e dunque

$$S_n(m) := \sum_{\mu=1}^m S_{\mu,n} = \frac{m}{n} - \frac{m(m-1)}{2!} \frac{1}{n(n-1)} + \frac{m(m-1)(m-2)}{3!} \frac{1}{n(n-1)(n-2)} -$$

$$- \frac{m(m-1)(m-2)(m-3)}{4!} \frac{1}{n(n-1)(n-2)(n-3)} + \cdots + (-1)^{m-1} \frac{(m-1)!}{(m-1)!} \frac{1}{n(n-1)\cdots(n-m+1)},$$

In particolare, quando si prende  $n = m$  la probabilità di vittoria di *Pierre* è

$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n-1} \frac{1}{n!}, \quad (4.2)$$

risultato di cui Montmort ([19], p. 134) fornì un'interpretazione geometrica legata alla curva logaritmica. Quando  $n \rightarrow \infty$ , la probabilità di vittoria di *Pierre* tende al valore limite  $1 - \frac{1}{e}$ , dove  $e$  è la base dei logaritmi naturali. Questo legame con i logaritmi sarà ripreso anche da Eulero che trattò il gioco delle concordanze in una nota pubblicata nel 1753 [10], e da De Moivre, che lo presentò come Problema XXXVI del trattato *The Doctrine of Chances*. Prima di esaminare la soluzione di De Moivre, è opportuno riportare l'*incipit* di quest'opera, perché vi compare la definizione cosiddetta "classica" della probabilità di un evento, già pubblicata nel *De mensura sortis*, apparso nel 1711:

<sup>4</sup>Ricordiamo che le righe sono numerate dall'alto a partire dalla riga numero 1 mentre le colonne sono numerate da sinistra, partendo dal numero 0.

La probabilità di un evento è maggiore o minore a seconda del numero di casi in cui può realizzarsi, confrontato con il numero di casi in cui può o realizzarsi o non realizzarsi.

Pertanto, se formiamo una frazione il cui numeratore è il numero di casi in cui un evento può realizzarsi ed il denominatore è il numero totale di casi in cui esso può realizzarsi o meno, quella frazione sarà una indicazione adeguata della probabilità del realizzarsi [di un evento].<sup>5</sup> ([8], pp. 1-2)

Questa definizione, rispetto a quella del *De Mensura sortis*, non fa riferimento all'ipotesi che i casi favorevoli e quelli contrari alla realizzazione di un evento debbano essere *aeque faciles*, cioè "equiprobabili" ([6], p. 215). L'aspettazione è definita subordinatamente al concetto di probabilità di un evento, al contrario di quanto aveva fatto Huygens. La vecchia impostazione viene comunque ripresa come conseguenza della nuova:

Da quanto detto segue necessariamente che, se sono assegnati tanto il valore di un'aspettazione quanto il valore della cosa attesa, allora dividendo il primo valore per il secondo, il quoziente esprimerà la probabilità di ottenere la somma attesa.<sup>6</sup> ([8], p. 4)

In termini moderni, se  $\bar{x}$  è il valore atteso di una variabile aleatoria che può assumere i valori  $x$  con probabilità  $p$  e  $0$  con probabilità  $1 - p$ , allora da

$$\bar{x} = px + (1 - p)0$$

segue che  $p = \frac{\bar{x}}{x}$ , come asserito da De Moivre. Vedremo come Thomas Bayes prenderà questa come *definizione* della probabilità di un evento. De Moivre mette anche in chiaro il concetto di indipendenza di due eventi, così come quello di probabilità condizionata, enunciando correttamente la regola per calcolarla:

Due eventi sono indipendenti quando non c'è tra loro alcuna connessione ed il realizzarsi di uno né agevola né ostacola il realizzarsi dell'altro.

Due eventi sono dipendenti, quando sono così collegati che la probabilità che si realizzi uno di loro è modificata dal realizzarsi dell'altro.<sup>7</sup> ([8], p. 6)

---

<sup>5</sup>The probability of an event is greater or less, according to the number of chances by which it may happen, compared with the whole number of chances by which it may either happen or fail.

Wherefore, if we constitute a fraction whereof the numerator be the number of chances whereby an event may happen, and the denominator the number of all the chances whereby it may either happen or fail, that fraction will be a proper designation of the probability of happening.

<sup>6</sup>From what precedes, it necessarily follows that if the value of an expectation be given, as also the value of the thing expected, then dividing the first value by the second, the quotient will express the probability of obtaining the sum expected.

<sup>7</sup>Two events are independent, when they have no connexion one with the other, and that the happening of one neither forwards nor obstructs the happening of the other.

Two events are dependent, when they are so connected together as that the probability of either's happening is altered by the happening of the other.

De Moivre considera, come esempio, l'estrazione dell'asso da un mazzo di 13 carte, tutte dello stesso seme. La probabilità di estrarre questa carta due volte successive, rimettendola nel mazzo dopo la prima estrazione è  $\frac{1}{13} \times \frac{1}{13} = \frac{1}{169}$  e le due estrazioni corrispondono ad eventi indipendenti. La dipendenza si verifica quando, come secondo evento, si consideri l'estrazione di un 2, *senza* che l'asso estratto sia reinserito nel mazzo. La probabilità dell'evento composto è  $\frac{1}{13} \times \frac{1}{12} = \frac{1}{156}$  e De Moivre commenta:

Da ciò si può dedurre che la probabilità che si realizzino due eventi tra loro dipendenti è il prodotto della probabilità che si realizzi uno di essi per la probabilità che l'altro avrà di realizzarsi quando si consideri il primo come realizzato.<sup>8</sup> ([8], p. 7)

Torniamo al gioco delle concordanze, che De Moivre risolse ricorrendo ad un'originale notazione. Siano  $n$  le carte. La probabilità che una carta  $a$  sia estratta alla  $a$ -esimo turno è

$$p(a) = \frac{(n-1)!}{n!} = \frac{1}{n} =: r;$$

De Moivre indica la probabilità che un'altra carta ( $b$ ) sia in  $b$ -esima posizione con  $+b = p(b)$ , mentre  $-b = 1 - p(b)$  è la probabilità che  $b$  non sia estratta al  $b$ -esimo turno: è chiaro che i segni  $\pm$  *non* sono usati nell'accezione aritmetica. Infatti egli prosegue dicendo

Esprimiamo la probabilità che  $a$  sia al primo posto e  $b$  al secondo come  $a + b$ : similmente, esprimiamo con  $a - b$  la probabilità che,  $a$  occupando il primo posto,  $b$  sia esclusa dal secondo posto<sup>9</sup> ([8], p. 111)

e dunque la sua notazione è da intendere in senso logico-insiemistico. Per calcolare le probabilità relative al gioco delle concordanze De Moivre scompose un evento nell'unione di due eventi disgiunti. Dette allora

$$s := p(a, b) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \quad t := p(a, b, c) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)},$$

$$v := p(a, b, c, d) = \frac{(n-4)!}{n!} = \frac{1}{n(n-1)(n-2)(n-3)}$$

le probabilità che  $a$  e  $b$  (in realtà una qualunque coppia di carte) siano al posto giusto,  $a, b, c$  (o una qualsiasi terna) lo siano, oppure ancora che  $a, b, c, d$  (o una quaterna qualsiasi) lo siano e posto

$$p(\bar{a}, b)$$

<sup>8</sup>From whence it may be inferred, that the probability of happening of two events dependent, is the product of the probability of the happening of one of them, by the probability which the other will have of happening, when the first is considered as having happened.

<sup>9</sup>Let the Probability that  $a$  being in the first place,  $b$  shall be in the second, expressed by  $a + b$ : Likewise let the probability that  $a$  being in the first place,  $b$  shall be excluded the second, be expressed by  $a - b$ .

per indicare la probabilità che  $b$  sia estratta al posto corretto (cioè al  $b$ -esimo) posto *ma* che  $a$  non lo sia, si avrà

$$p(b) = p(a, b) + p(\bar{a}, b) \quad (4.3)$$

e quindi

$$p(\bar{a}, b) = r - s \quad (4.4)$$

è la probabilità richiesta nella soluzione del gioco delle concordanze. Ora, anche  $p(b, c) = s$  perché la seconda e la terza carta sono nelle stesse condizioni della prima e della seconda mentre, per definizione,

$$p(a, b, c) = t$$

e

$$s = p(b, c) = p(a, b, c) + p(\bar{a}, b, c).$$

Si avrà allora

$$p(\bar{a}, b, c) = s - t.$$

D'altro canto,  $p(\bar{a}, c) = p(\bar{a}, b) = r - s$  e quindi, siccome

$$p(\bar{a}, c) = p(\bar{a}, b, c) + p(\bar{a}, \bar{b}, c)$$

si ottiene

$$p(\bar{a}, \bar{b}, c) = (r - s) - (s - t) = r - 2s + t.$$

Anche se questo metodo non ha la stessa rapidità di quello di Montmort e Nicholas Bernoulli, l'aspetto saliente è l'uso del principio delle probabilità totali, espresso in questi termini:

Il numero di possibilità che la lettera  $a$  ha di essere al primo posto contiene il numero di possibilità in cui, essendo  $a$  al primo posto,  $b$  può essere al secondo posto oppure no: questo è un assioma di buon senso, con lo stesso grado di evidenza dell'assioma che l'intero è uguale a tutte le sue parti.<sup>10</sup>  
([8], p. 110)

Grazie a questo principio egli riuscì non solo a risolvere il problema delle concordanze ma a determinare le probabilità di eventi composti, almeno quando questi eventi sono simmetrici.<sup>11</sup>

Montmort e Nicholas Bernoulli avevano succintamente esposto una formula ricorsiva per determinare la probabilità di vincere al gioco delle concordanze con  $n$  carte, in termini delle probabilità di successo quando le carte erano  $n - 1$  ed  $n - 2$ . Eulero [10] giunse ad un'altra formula ricorsiva esplicitando però il procedimento. Supponiamo di giocare con  $n$  carte; le  $M := n!$  permutazioni

<sup>10</sup>The number of chances for the letter  $a$  to be in the first place, contains the number of chances by which  $a$  being in the first place,  $b$  may be in the second, or out of it: this is an axiom of common sense of the same degree of evidence, as that the whole is equal to all parts.

<sup>11</sup>Ciò significa che le probabilità, per esempio, di estrarre una coppia di carte al turno corrispondente al loro punteggio, non dipende dai punteggi delle carte.

possibili vengono ripartite in  $n$  classi formate ciascuna da  $\frac{M}{n} = (n-1)!$  permutazioni ciascuna delle quali ha il primo, il secondo, ..., l' $n$ -esimo elemento fisso. Occorre sapere come diminuire  $\frac{M}{n}$  perché il risultato dia le permutazioni che hanno un certo elemento fisso e *nessuno* degli elementi precedenti fisso. Eulero considera l'esempio in cui vi sono quattro carte: 1, 2, 3, 4. Le sei permutazioni che hanno il numero 3 in terza posizione sono

1	1	2	2	4	4
2	4	4	1	1	2
3	3	3	3	3	3
4	2	1	4	2	1

*Eliminando* la terza riga possiamo cercare il numero di permutazioni che abbiano il numero 1 o il 2 estratti al primo o secondo turno analizzando la tabella ridotta

1	1	2	2	4	4
2	4	4	1	1	2
—	—	—	—	—	—
4	2	1	4	2	1

Ciò è sufficiente ad Eulero per esporre il procedimento in generale. Nel caso di  $n$  carte, siano  $\alpha, \beta, \gamma, \dots$  il numero di casi in cui *Pierre* vince al primo, al secondo, al terzo turno e così via. Quando le carte sono  $n+1$  e si hanno  $M' = (n+1)! = (n+1)M$  permutazioni possibili, indichiamo il numero di permutazioni che fanno vincere *Pierre* al primo, al secondo, al terzo turno, ecc. con  $\alpha', \beta', \gamma'$ . Ora

$$\alpha' = \frac{M'}{n+1} = M$$

e per trovare  $\beta'$  si elimina la carta numero 2 rimanendo con  $n$  carte e si sottraggono da  $\alpha' = M$  tutte le permutazioni tra  $n$  carte che fanno vincere al *primo* turno, vale a dire  $\alpha$ . In altre parole

$$\beta' = M - \alpha = \alpha' - \alpha$$

Procedendo allo stesso modo si ottiene

$$\gamma' = M - \alpha - \beta = \beta' - \beta, \quad \delta' = M - \alpha - \beta - \gamma = \gamma' - \gamma$$

e via di seguito. Il passaggio alla probabilità di vittoria di *Pierre* è ora diretto perché, se  $\varrho$  è il numero di permutazioni vincenti per *Pierre* ad un certo passo, quando si gioca con  $n$  carte;  $\varrho'$  il numero di permutazioni vincenti, allo *stesso* passo, quando le carte sono  $n+1$ ;  $\sigma'$  le permutazioni vincenti al passo *successivo* quando le carte sono sempre  $n+1$ , allora deve essere

$$\sigma' = \varrho' - \varrho. \tag{4.5}$$

La probabilità  $P$  che *Pierre* vinca giocando con  $n$  carte ad un certo passo, sarà dunque  $P = \frac{\varrho}{M} = \frac{\varrho}{n!}$ , quella di vincere allo stesso passo ma giocando con una

carta in più sarà  $Q = \frac{\varrho'}{(n+1)!} = \frac{\varrho'}{(n+1)M}$  e quella  $R$  di vincere al passo successivo con  $n + 1$  carte è  $R = \frac{\sigma'}{(n+1)!} = \frac{\sigma'}{(n+1)M}$  ovvero, grazie all'equazione (4.5),

$$R = \frac{\varrho' - \varrho}{(n+1)M} = Q - \frac{P}{n+1}.$$

Una prima riflessione di natura generale basata sul gioco delle concordanze fu proposta da Johann-Heinrich Lambert (1728-1777), in un lavoro [15] del 1771 dove voleva dimostrare infondata la pretesa dei curatori di almanacchi di poter prevedere il tempo atmosferico per tutto un anno solare. Lambert si era sforzato invano di far capire quanto non fosse ben riposta la stima accordata alle previsioni contenute in questi almanacchi, che erano molto diffusi. L'opinione di Lambert si scontrava col fatto che le previsioni meteorologiche degli almanacchi non erano, in definitiva, troppo lontane dal vero. Con il lavoro [15] egli si rivolse al calcolo delle probabilità per mostrare come questa circostanza fosse dovuta al modo un po' ambiguo con cui certe condizioni atmosferiche erano descritte ed al numero ristretto di termini usati per indicare i cambiamenti del tempo. Lambert voleva mettere in chiaro, grazie al calcolo delle probabilità, che se i termini usati per descrivere diverse condizioni meteorologiche non erano troppi, le possibilità di *indovinare* i cambiamenti del tempo atmosferico crescevano. Per questo egli considerò proprio il gioco delle concordanze che conosceva dal lavoro di Eulero [10]:

Supponiamo ora che il compilatore di un almanacco sfumi e vari le predizioni in modo da assegnare 30 diverse qualificazioni del tempo atmosferico. In questo caso il lettore sarà presto convinto, perché l'almanacco non si accorderà che molto raramente con il tempo atmosferico che si presenta.<sup>12</sup> ([15], p. 414)

Nella parte matematica del lavoro, Lambert si chiese quante fossero, tra le  $n!$  permutazioni di  $n$  oggetti, quelle che lasciano un certo numero di oggetti disposti secondo il proprio rango. In generale, sia  $n_k$  il numero di permutazioni di  $n$  oggetti che *spostano*  $k$  oggetti *assegnati*. In primo luogo  $n_0 = 1$  mentre, poiché non esistono permutazioni in cui un solo oggetto è fuori posto, si ha  $n_1 = 0$ . Quanto alle permutazioni che spostano 2 oggetti, Lambert osserva che questa coppia di oggetti si può scegliere in  $\binom{n}{2}$  modi diversi e che, scelta una coppia, non è possibile effettuare ulteriori permutazioni perché ogni loro permutazione ordinerebbe *tutti* gli  $n$  oggetti e dunque  $n_2 = 1$ . In generale, le permutazioni in cui  $k$  oggetti *qualsiasi* non occupano la posizione corrispondente al proprio rango sono  $n_k \binom{n}{k}$ . In definitiva, si ha una partizione delle  $n!$  permutazioni

<sup>12</sup>Supposons maintenant que le faiseur d'almanacs nuance et varie ses prédictions en sorte qu'il assigne 30 modifications différentes de tems. Donnons encore au tems qu'il fait effectivement 30 modifications différentes. Dans ce cas le public seroit bientôt détrompé, parce que l'almanac ne s'accorderoit que très rarement avec le tems qu'il fait.

del tipo

$$n! = n_0 + \sum_{k=2}^n n_k \binom{n}{k}. \quad (4.6)$$

Per trovare i coefficienti  $n_k$ , che *non* dipendono da  $n$ , Lambert riscrisse (4.6), successivamente, per  $n = 0, 1, 2, \dots$ . Per  $n = 0$  si ha, come già visto,

$$1 = n_0$$

mentre per  $n = 2$  l'equazione (4.6) si riduce a

$$2 = n_0 + n_2$$

da cui  $n_2 = 1$ , come già osservato. Quando  $n = 3$  si ha

$$6 = n_0 + 3n_2 + n_3$$

e quindi  $n_3 = 2$ . Successivamente si ottiene

$$24 = n_0 + 6n_2 + 4n_3 + n_4 \quad \text{da cui} \quad n_4 = 9$$

e

$$120 = n_0 + 10n_2 + 10n_3 + 5n_4 + n_5 \quad \text{da cui} \quad n_5 = 44.$$

Lambert ha abbastanza elementi per fornire la regola ricorsiva che permette di calcolare tutti i coefficienti  $n_k$ :

$$n_k = kn_{k-1} + (-1)^k.$$

Tornando al problema che aveva dato origine alle considerazioni sul gioco delle concordanze, l'esito dell'analisi è che, se i redattori degli almanacchi compilassero le loro previsioni adoperando  $n = 30$  diverse sfumature per indicare il tempo atmosferico, la probabilità di successo sarebbe molto bassa: riducendo il numero di termini e basandosi sulle osservazioni di quanto successo in passato, è invece possibile dare alle proprie previsioni una parvenza di attendibilità che però è *spiegabile* grazie alle regole della probabilità. Lambert ha dunque fatto subire una prima importante metamorfosi al gioco delle concordanze, trasformandolo da esercizio ad alleato nella battaglia contro il dilagare di pratiche superstiziose.

Il problema delle concordanze riemerse in un'altra variante nella *Théorie Analytique des probabilités* di Laplace ([16], p. 217):

Immaginiamo che un'urna contenga  $r$  palle numerate con 1,  $r$  numerate con 2,  $r$  numerate con il 3 e così via fino al numero  $n$ . Queste palle vengono estratte in successione dopo essere state ben mescolate; si richiede la probabilità che almeno una delle palle venga estratta in posizione corrispondente al suo numero o che ne saranno estratte almeno due o almeno tre, ecc. in queste condizioni.<sup>13</sup>

<sup>13</sup>Concevons dans une urne,  $r$  boules marquées du n°1,  $r$  boules marquées du n°2,  $r$  boules marquées du n°3, et ainsi de suite jusqu'au n°  $n$ . Ces boules étant bien mêlées dans l'urne, on les tire toutes successivement; on demande la probabilité qu'il sortira au moins une de ces boules, au rang indiqué par son numéro, ou qu'il en sortira au moins deux, ou au moins trois, etc.

L'aspetto nuovo sta nel fatto che la numerazione si estende da 1 ad  $n$  mentre il numero totale di palle nell'urna è  $rn$ , strettamente maggiore di  $n$ , non appena sia  $r > 1$ . La prima osservazione di Laplace è che, se ci possono essere delle coincidenze, queste debbono aver luogo non oltre le prime  $n$  estrazioni (senza reimbussolamento). Per valutare allora la probabilità di avere almeno una coincidenza, occorre calcolare il numero dei casi possibili che è pari alle *disposizioni* di  $rn$  elementi presi a gruppi di  $n$ : occorre considerare le disposizioni e non le combinazioni perché l'ordine di estrazione è importante. Il numero di casi possibili è allora

$$rn(rn - 1) \cdots (rn - n + 1) =: D(rn, n).$$

Il numero di casi favorevoli al fatto che una palla con il numero 1 esca per prima è invece  $rD(rn - 1, n - 1)$ , dove il fattore  $r$  tiene memoria del fatto che vi sono  $r$  palle che recano il numero 1. Poiché questo ragionamento si applica a tutti gli altri  $n - 1$  gruppi di palle, si hanno

$$rnD(rn - 1, n - 1)$$

casi favorevoli, *non* tutti distinti, all'evento in cui si presenti almeno una concordanza. Per eliminare i termini ridondanti, Laplace applica ancora l'argomento di inclusione-esclusione.

Come ultima applicazione del gioco delle concordanze, riportiamo quella pubblicata nel 1819 da Thomas Young (1773-1829), personaggio poliedrico, noto per i suoi studi sulla teoria ondulatoria della luce e sulla capillarità ma anche per essere stato tra i primi a cimentarsi con successo nella decifrazione dei caratteri geroglifici, poco prima del francese Jean-François Champollion. In una lettera inviata a Henry Kater, pubblicata sulle *Philosophical Transactions* della Royal Society di Londra [25], Young mise in guardia da superficiali applicazioni del calcolo della probabilità, proponendo un esempio tratto dalla linguistica:

Ci sono casi in cui un poco di aiuto può essere ottenuto dalla teoria delle probabilità in quanto concerne questioni di letteratura e storia: anche qui però sarebbe estermamente facile corrompere questa applicazione in modo da renderla strumentale allo scopo di avvolgere un ragionamento fallace in un manto di rigore. Così, se stessimo studiando le relazioni reciproche tra due lingue con lo scopo di stabilire fino a che punto esse indichino un'origine comune da una lingua più antica o uno scambio occasionale tra le due nazioni che le parlano, sarebbe importante stabilire—nell'ipotesi che le possibili varietà di monosillabi o parole molto semplici, debbano essere limitate ad un certo numero a causa dell'estensione dell'alfabeto; e che questi nomi siano stati assegnati in modo casuale allo stesso numero di cose—quale sarebbe la possibilità che 1, 2, 3, o più nomi siano stati associati alle stesse cose in due situazioni indipendenti.<sup>14</sup> ([25], p. 80)

<sup>14</sup>These are cases in which some little assistance may be derived from the doctrine of chances with respect to matters of literature and history: but even here it would be extremely easy to

Young ripeté succintamente lo stesso argomento di Lambert, senza citarlo, determinando i coefficienti  $n_k$ , per  $k$  fino a 10. In questo caso, ad esempio, le probabilità che su 10 parole uguali usate in due lingue, nessuna sia stata usata per indicare la stessa cosa è 0.36788. La presenza di un numero più grande di concordanze potrebbe però indurre a conclusioni poco verosimili:

Sembra dunque che non sia possibile dedurre nulla a riguardo delle relazioni tra due lingue a partire dalla coincidenza di un'unica parola tra di loro; e che si può scommettere 3 contro 1 contro l'accordo di due parole: ma se vi sono tre parole identiche, si può scommettere più di 10 contro 1 che le lingue debbono avere un'origine comune o introdotte in qualche altro modo; sei parole darebbero 1700 possibilità contro 1 ed 8 quasi 100000 contro 1: in questi ultimi casi dunque l'evidenza equivale praticamente a certezza. Ora, nel vocabolario che accompagna l'elegante saggio del barone W.[ILHELM] VON HUMBOLDT troviamo, per esempio, nella lingua Basca, cioè l'antica lingua della Spagna, le parole *berria*, nuovo; *ora*, cane; *guchi*, piccolo; *oguia*, pane; *otsoa*, lupo, da cui le parole spagnole *onza*; *zazpi* o, come la scrive LACROZE, *shashpi*, sette. Nell'egiziano antico, nuovo è BERI; cane, UHOR; piccolo, KUDCHI; pane, OIK; lupo, UONSH; e sette, SHASHF; se consideriamo sufficientemente identiche queste parole da potervi applicare i calcoli precedenti, vi saranno più di mille possibilità contro una che, in tempi molto remoti, una colonia egizia si sia stabilita in Spagna, dal momento che nessuna lingua di popoli vicini reca tracce tali da poter essere stata il veicolo di trasmissione di queste parole. (...) ma la relazione, (...) amplificata a diventare una identità, sembra in generale essere quella di una flebile somiglianza: e questo è proprio l'esempio di un caso nel quale tentare di ridurre l'argomento a calcolo significherebbe ingannare noi stessi.<sup>15</sup> ([25], pp. 81-82)

---

pervert this application in such a manner, as to make it subservient to the purpose of clothing fallacious reasoning in the garb of demonstrative evidence. Thus if we were investigating the relations of two languages to each other, with a view of determining how far they indicated a common origin from an older language, or an occasional intercourse between the two nations speaking them, it would be important to inquire, upon the supposition that the possible varieties of monosyllabic or very simple words must be limited by the extent of the alphabet to a certain number; and that these names were to be given promiscuously to the same number of things, what would be the chance that 1, 2, 3 or more of the names would be applied to the same things in two independent instances.

<sup>15</sup>It appears therefore that nothing whatever could be inferred with respect to the relation of two languages from the coincidence of the sense of any single word in both of them; and that the odds would only be 3 to 1 against the agreement of two words: but if three words appeared to be identical, it would be more than 10 to 1 that they must be derived in both cases from some parent language, or introduced in some other manner; six words would give near 1700 chances to 1, and 8 near 100,000: so that in these last cases the evidence would be little short of absolute certainty. In the Biscayan, for example, or the ancient language of Spain, we find in the vocabulary accompanying the elegant essay of Baron W. VON HUMBOLDT, the words *berria*, new; *ora*, a dog; *guchi*, little; *oguia*, bread; *otsoa*, a wolf, whence the Spanish *onza*; and *zazpi*, or, as LACROZE writes it, *shashpi*, seven. Now in the ancient Egyptian, new is BERI; a dog, UHOR; little, KUDCHI; bread, OIK; a wolf, UONSH; and seven, SHASHF; and if we consider these words as sufficiently identical to admit of our calculating upon them, the chances will be more than a thousand to one, that, at some very remote period, an Egyptian

Essendosi appena conclusa l'avventura napoleonica, il contesto culturale inglese in cui operava Young era di opposizione alla Francia, dove si assisteva al tentativo di applicare la probabilità all'ambito giuridico-sociale. Le osservazioni di Young testimoniano quanta attenzione occorra prestare quando il calcolo delle probabilità viene applicato a situazioni concrete che esulano dagli schemi classici dei giochi d'azzardo. A margine dei risultati ottenuti concludiamo con un'osservazione sul valore limite *irrazionale*  $1 - \frac{1}{e}$  per la probabilità di successo di *Pierre* nel gioco delle concordanze quando il numero  $n$  di carte tenda all'infinito. Può sorprendere come, dopo la definizione di probabilità data da De Moivre, da cui si può concludere che la probabilità di un evento è un numero *razionale*, questo risultato non abbia condotto ad una riflessione sulla definizione di probabilità. Occorre però osservare che, quando Montmort o De Moivre si occuparono del problema, la natura del numero e *non* era nota. Saranno Eulero e Lambert a dimostrarne l'irrazionalità nella seconda metà del XVIII secolo. Inoltre,  $1 - \frac{1}{e}$  è una probabilità *limite* ed era accettato che il limite di una successione composta da numeri razionali, potesse essere irrazionale. A questo proposito ricordiamo che il politico e matematico olandese Jan De Witt (1625-1672), che fu anche primo ministro e si occupò dell'emissione di rendite vitalizie come strumento di finanziamento dello stato, già nel 1671 aveva notato, parlando del concetto di aspettazione, legato ai valori di oggetti preziosi, la possibilità di coinvolgere anche numeri irrazionali<sup>16</sup>:

Occorre poi osservare che mi sono qui servito espressamente di un esempio o caso in cui sono coinvolti tre oggetti di valore, senza introdurre alcuna somma ma parlando di una perla, di un rubino o di un diamante, per fare in modo che la dimostrazione si potesse applicare a tutti i tipi di numeri, alle frazioni come ai numeri interi, ai numeri irrazionali come ai razionali, visto che tutti i numeri immaginabili si possono utilizzare per esprimere il valore di queste pietre preziose.<sup>17</sup> ([9], p. 90)

## 4.2 La durata del gioco

Il problema della durata del gioco è legato a quello della rovina del giocatore già incontrato nel *De Ratiociniis* di Huygens. Montmort e Nicholas Bernoulli ebbero su questo problema una corrispondenza pubblicata in [19] ed il problema

---

colony established itself in Spain: for none of the languages of the neighbouring nations retain any traces of having been the medium through which these words have been conveyed. (...) but the relation, (...) magnified into identity, appears in general to be that of a very faint resemblance: and this is precisely an instance of a case, in which it would be deceiving ourselves to attempt to reduce the matter to a calculation.

<sup>16</sup>Ho utilizzato la traduzione in inglese contenuta nella biografia di De Witt [9]. L'originale olandese ha per titolo "Waardye van Lyf-renten naer proportie van Los-Renten", traducibile in inglese come *Value of life annuities in proportion to redeemable annuities* [12].

<sup>17</sup>And it is to be observed, that I have here expressly made use of an example or case of three objects of value, without expression of any sum, as in speaking of a pearl, a ruby, or a diamond, so as to cause the demonstration to be applicable to all sorts of numbers, to fractions as well as integer numbers, to irrational as well as to rational numbers, since all imaginable numbers may be applied to the value of these jews.

fu affrontato a più riprese da De Moivre. Ci occupiamo proprio dei metodi di De Moivre che mostrano la necessità di introdurre nuovi strumenti matematici, che affianchino l'analisi combinatoria dello sviluppo del binomio, per una soluzione generale del problema. È lo stesso De Moivre a ricordare, nel *Doctrine of Chances*, le tappe che lo condussero ad affinare i propri metodi. Egli cominciò ad occuparsi del problema nel 1708, partendo dalla versione che sarà riprodotta come Problema LXV di [8]:

*A e B, le cui abilità stanno tra loro come a sta a b, giocano insieme. Qual è la probabilità che uno di loro, sia A, possa, in un numero assegnato di partite, vincere a B un certo numero di gettoni? Si lasci del tutto ingiudicato se B, prima del termine del gioco, possa trovarsi o meno nella circostanza di vincere lo stesso numero o un qualunque numero di gettoni di A.*<sup>18</sup> ([8], p. 208)

De Moivre pubblicò una prima soluzione del problema nel *De Mensura sortis*, accennando anche ad una soluzione più semplice e riservandosi la possibilità di ritornare sul problema in un secondo momento. Leggendo la seconda edizione del trattato di Montmort e non riuscendo a seguire l'intricato processo di soluzione di Nicholas Bernoulli, De Moivre studiò la soluzione di Montmort e, trovatala in difetto, ne pubblicò una versione corretta nella prima edizione del *Doctrine of Chances* del 1718, attribuendola interamente a Montmort. Non avendo ricevuto alcun ringraziamento, ripubblicò la propria soluzione, rivendicandone l'originalità. Seguiamo la presentazione del problema della durata del gioco nell'edizione del 1756 del *Doctrine of Chances*. L'approccio di De Moivre è, al solito, graduale: illustrazione di casi semplici per familiarizzare il lettore con il metodo di soluzione, che viene poi generalizzato. Il problema LVIII cui ci riferiamo è formulato in questi termini ([8], p. 191):

Due giocatori d'azzardo *A e B*, le cui abilità stanno tra loro come *a* sta a *b*, hanno ciascuno un certo numero di gettoni e giocano con queste regole: ogni volta in cui *A* vince, *B* gli dà un gettone; ogni volta in cui *B* vince, *A* gli darà un gettone; essi non smettono di giocare finché uno dei due non ha preso tutti i gettoni dell'avversario: ora supponiamo due astanti *R* ed *S* che si interessino a come finisce il gioco; il primo di loro scommette che il gioco finirà entro un certo numero di partite che egli assegna, l'altro scommette sull'evento opposto. Trovare la probabilità che *S* ha di vincere la scommessa.<sup>19</sup>

---

<sup>18</sup>If *A* and *B*, whose proportion of skill is supposed as *a* to *b*, play together: What is the probability that one of them, suppose *A*, may in a number of Games not exceeding a number given, win of *B* a certain number of Stakes? leaving it wholly indifferent whether *B*, before the expiration of those Games, may or may not have been in the circumstance of winning the same, or any other number of Stakes of *A*.

<sup>19</sup>Two Gamesters *A* and *B* whose proportion of skill is as *a* to *b*, each having a certain number of Pieces, play together on condition that as often as *A* wins a Game, *B* shall give him *one* piece; and that they cease not to play till such time as either one or the other has got all the Pieces of his Adversary: now let us suppose two Spectators *R* and *S* concerning themselves about the ending of the Play, the first of them lying that the Play will be ended

Il ruolo di  $R$  ed  $S$  è del tutto fittizio: si tratta di trovare la probabilità che il gioco termini entro un certo numero di partite assegnato a priori. De Moivre parte dal caso in cui ciascun giocatore ha  $n = 2$  gettoni e si scommette che il gioco terminerà in 2 partite. Sviluppato

$$(a + b)^2 = a^2 + 2ab + b^2,$$

egli interpreta ogni termine come un particolare esito del gioco: così  $a^2$  indica due vittorie di  $A$ ,  $ab$  una vittoria per  $A$  ed una per  $B$ , nell'ordine,  $ba$  una vittoria per  $B$  seguita da una per  $A$  e  $b^2$  indica che  $B$  ha vinto due partite. Allora i termini  $a^2 + b^2$  sono favorevoli ad  $R$  mentre  $2ab = ab + ba$  sono i casi favorevoli ad  $S$  la cui probabilità di aggiudicarsi la scommessa è dunque

$$\frac{2ab}{(a + b)^2}.$$

Quando i giocatori hanno ancora  $n = 2$  gettoni ciascuno e si scommette sul termine del gioco in  $3 = 2 + 1$  partite, De Moivre sviluppò

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

ed osservò che i termini estremi sono ancora sfavorevoli ad  $S$  mentre i due termini centrali richiedono un'analisi ulteriore, prima di essere ascritti come favorevoli ad uno degli astanti. De Moivre risolve il problema scrivendo

$$3a^2b + 3ab^2 = a^2b + aba + ba^2 + ab^2 + bab + b^2a : \quad (4.7)$$

solo i quattro termini centrali— $aba + ba^2 + ab^2 + bab$ —sono favorevoli ad  $S$  perché corrispondono a situazioni di gioco in cui nessuno dei giocatori riesce ad avere in una fase intermedia dello svolgimento della partita un vantaggio sull'avversario di 2 punti, cioè pari al numero di gettoni inizialmente in possesso di ciascuno. La probabilità che  $S$  ha di vincere la scommessa è dunque

$$\frac{2ab(a + b)}{(a + b)^3} = \frac{2ab}{(a + b)^2}$$

e coincide dunque con quella del caso precedente. Il motivo è che

vincere un numero pari di gettoni in un numero dispari di partite è impossibile, a meno che ciò non accada nel numero pari di partite che precede immediatamente il numero dispari.<sup>20</sup> ([8], pp. 192-193)

L'ultimo caso particolare ad essere trattato è quello in cui, sempre partendo con  $n = 2$  gettoni ciascuno, si scommette sulla fine del gioco in  $4 = 2 + 2$  partite. De Moivre sviluppò

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$

in a certain number of Games which he assigns, the other laying to the contrary. To find the Probability that  $S$  has of winning his wager.

<sup>20</sup>the winning of a certain number of even Pieces in and odd number of Games is impossible, unless it was done in the even number of Games immediately preceding the odd number.

in cui i casi potenzialmente favorevoli ad  $S$  sono solo quelli presenti nel termine centrale

$$6a^2b^2$$

dal momento che in tutti gli altri casi il numero di vittorie di uno dei giocatori supera quello dell'altro di una quantità non inferiore ad  $n$ . Tra i termini centrali, quattro— $abab$ ,  $abba$ ,  $baab$ ,  $baba$ —sono favorevoli ad  $S$  la cui probabilità di vittoria è dunque

$$\frac{4a^2b^2}{(a+b)^4}$$

che si mantiene inalterata quando il numero di partite è  $n = 5$ . A questo punto De Moivre enunciò il risultato nel caso in cui ciascun giocatore abbia due gettoni e si scommetta sulla fine del gioco in  $2 + d$  turni. La probabilità di vittoria di  $S$  è

$$\begin{aligned} \frac{(2ab)^{1+\frac{d}{2}}}{(a+b)^{2+d}} & \text{ se } d \text{ è pari} \\ \frac{(2ab)^{\frac{1+d}{2}}}{(a+b)^{1+d}} & \text{ se } d \text{ è dispari.} \end{aligned}$$

Quando i gettoni in possesso di ciascun giocatore sono 3 e si scommette sul termine del gioco in  $3 + d$  turni, la probabilità di  $S$  è

$$\begin{aligned} \frac{(3ab)^{1+\frac{d}{2}}}{(a+b)^{2+d}} & \text{ se } d \text{ è pari} \\ \frac{(3ab)^{\frac{1+d}{2}}}{(a+b)^{1+d}} & \text{ se } d \text{ è dispari.} \end{aligned}$$

È a questo punto che De Moivre illustrò verbalmente l'algoritmo che si cela dietro la soluzione del problema, nel caso generale in cui ciascun giocatore abbia inizialmente  $n$  gettoni e si scommetta sul termine della partita in  $n + d$  turni. De Moivre non fornì la dimostrazione della validità dell'algoritmo ma si limitò ad enunciare i passi da seguire per ottenere la corretta probabilità di vittoria per  $S$ , illustrando l'algoritmo con esempi che egli riteneva sufficientemente chiari per convincere il lettore della correttezza del metodo.

1. Si sviluppi  $(a + b)^n$  e si eliminino i termini estremi  $a^n$  e  $b^n$
2. Si moltiplichino ciò che resta per  $(a + b)^2$
3. Si eliminino i termini estremi e si iteri il processo per  $\frac{d}{2}$  o  $\frac{d-1}{2}$  volte, a seconda che  $d$  sia pari o dispari.
4. Si divida la somma dei termini residui per  $(a + b)^{n+d}$  o per  $(a + b)^{n+d-1}$  a seconda che  $d$  sia pari o dispari: la frazione ottenuta è la probabilità cercata.

Per spiegare il significato di questi passi, ricordiamo che il gioco termina quando la *differenza* tra il numero di partite vinte da ciascun giocatore è uguale ad  $n$ . Il primo passaggio di De Moivre è allora chiaro:  $a^n$  e  $b^n$ , i termini estremi dello sviluppo di  $(a + b)^n$  corrispondono proprio al caso in cui uno dei due giocatori

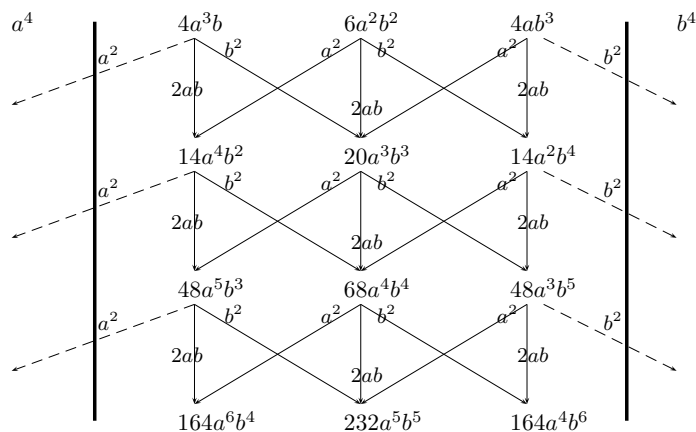


Figura 4.1: Schema per illustrare la generazione della probabilità che il gioco non termini entro  $4 + 2d$  turni, con  $d = 0, 1, \dots$ .

ha vinto tutte le prime  $n$  partite e dunque ha vinto tutti gli  $n$  gettoni in possesso dell'avversario. Se non li si eliminasse, dovremmo affrontare tutte le ambiguità nell'attribuire a favore di  $R$  od  $S$  i termini successivi dello sviluppo.

Perché De Moivre moltiplicò per  $(a + b)^2$ , che contiene gli esiti possibili di due partite, e non, più semplicemente, per  $a + b$ ? La risposta è anche qui racchiusa negli esempi già illustrati. La probabilità di vittoria di  $S$  entro  $n + 2\ell$  turni ed entro  $n + 2\ell + 1$  coincidono. De Moivre, moltiplicando passo passo per  $(a + b)^2$  vuole evidenziare i *cambi* nei valori delle probabilità di vittoria di  $S$ . L'eliminazione dei termini estremi procede anche nei passi successivi perché essi avevano al passo precedente una differenza negli esponenti di  $a$  e  $b$  pari ad  $n - 2$  che diventa  $n$  con la moltiplicazione per  $a^2$  o  $b^2$ . I termini residui sono tutti e soli quelli favorevoli ad  $S$  e quindi l'algoritmo di De Moivre fornisce la soluzione del problema della durata del gioco.

Per illustrare quanto detto consideriamo un caso semplice:  $n = 4$ . L'algoritmo di De Moivre si può ricondurre alla figura 4.2 in cui la prima riga contiene tutti i termini dello sviluppo di  $(a + b)^4$ .

Le linee in grassetto separano i termini esterni, favorevoli ad  $R$ , da quelli interni, favorevoli ad  $S$ . La somma dei termini della prima riga compresi tra queste due linee fornisce la probabilità  $p_4(4)$  che il gioco *non* termini entro le  $4 = 4 + 0$  partite:

$$p_4(4) = \frac{4a^3b + 6a^2b^2 + 4ab^3}{(a + b)^4}.$$

I termini interni vengono moltiplicati per  $(a + b)^2$ : ciascun termine può restare nella stessa colonna, quando viene moltiplicato per  $2ab$  o migrare a sinistra o a destra di una colonna, quando viene moltiplicato per  $a^2$  o  $b^2$ , rispettivamente.

Quando, a seguito di questa moltiplicazione, un termine si sposta fuori dalle linee in grassetto, ciò significa che la differenza degli esponenti di  $a$  e  $b$  ha superato  $n$  e dunque il termine non va più considerato perché uno dei due avversari ha perso tutti i gettoni. Se diciamo  $\alpha_m$ ,  $\beta_m$  e  $\gamma_m$  i coefficienti numerici interni, posti sulla riga  $m$ -esima, da sinistra a destra, abbiamo:

$$\gamma_m = \alpha_m$$

per la simmetria del problema e

$$\alpha_m = 2\alpha_{m-1} + \beta_{m-1} \quad \beta_m = 2\beta_{m-1} + \alpha_{m-1} + \gamma_{m-1} = 2(\beta_{m-1} + \alpha_{m-1}) \quad (4.8)$$

per le regole di formazione di una riga a partire dalla precedente. Usando queste regole, è possibile determinare

$$\begin{aligned} p_4(6) &= \frac{14a^4b^2 + 20a^3b^3 + 14a^2b^4}{(a+b)^6} \\ p_4(8) &= \frac{48a^5b^3 + 68a^4b^4 + 48a^3b^5}{(a+b)^8} \\ p_4(10) &= \frac{164a^6b^4 + 232a^5b^5 + 164a^4b^6}{(a+b)^{10}} \\ p_4(12) &= \frac{560a^7b^5 + 732a^6b^6 + 560a^5b^7}{(a+b)^{12}} \end{aligned}$$

e così via.

Il metodo consente dunque di ottenere le probabilità richieste ma risulterebbe impraticabile se  $d$  fosse sufficientemente grande. De Moivre osservò però che i coefficienti numerici seguono una legge ricorsiva

$$\alpha_{m+1} = 4\alpha_m - 2\alpha_{m-1}, \quad \beta_{m+1} = 4\beta_m - 2\beta_{m-1} \quad (4.9)$$

In questo caso semplice la relazione si può trovare partendo dalle (4.8) e dalle equazioni che si ottengono da essa, cambiando  $m$  in  $m+1$ :

$$\alpha_{m+1} = 2\alpha_m + \beta_m \quad \beta_{m+1} = 2(\beta_m + \alpha_m) \quad (4.10)$$

e cominciando ad eliminare le  $\beta$  nelle relazioni di sinistra di (4.8) e (4.10). Per questo, sostituiamo  $\beta_m$  nell'equazione (4.10), che diventa

$$\alpha_{m+1} = 2\alpha_m + 2\alpha_{m-1} + 2\beta_{m-1};$$

siccome, moltiplicando la (4.8) per 2, si ha  $2\alpha_m = 4\alpha_{m-1} + 2\beta_{m-1}$ , possiamo eliminare  $\beta_{m-1}$ , ottenendo

$$\alpha_{m+1} = 4\alpha_m - 2\alpha_{m-1}. \quad (4.11)$$

Eliminando le  $\alpha$ , si ottiene la (4.9) per i coefficienti  $\beta$ . Per le successive considerazioni di De Moivre, è importante che tutte le colonne della tabella che fornisce le probabilità di successo di  $S$  seguano la *stessa* relazione di ricorrenza.

De Moivre si rese allora conto della necessità di fornire uno strumento più duttile per il calcolo delle probabilità di vittoria di  $S$  e per un insieme di altri problemi che richiedevano la valutazione dei termini e delle somme di particolari serie finite, le serie ricorrenti. Egli sviluppò la teoria di queste serie già nella *Miscellanea Analytica* del 1730 dove troviamo questa definizione:

Chiamo *ricorrente* una serie che sia formata in modo tale che, preso a piacere un certo numero di suoi termini, ogni termine successivo è legato allo stesso numero di termini precedenti secondo una legge di relazione immutabile.<sup>21</sup> ([8], p. 220)

De Moivre chiarì subito con un esempio, considerando la serie

$$A + B + C + D + E + F + \dots = 1 + 2x + 3x^2 + 10x^3 + 34x^4 + 97x^5 + \dots$$

in cui

$$D = 3Cx - 2Bx^2 + 5Ax^3 \quad E = 3Dx - 2Cx^2 + 5Bx^3 \quad F = 3Ex - 2Dx^2 + 5Cx^3 \dots :$$

le quantità  $3x - 2x^2 + 5x^3$  o la terna  $(3x, -2x^2, 5x^3) \equiv (3, -2, 5)$  formano la *scala di relazione* della serie ricorrente. In questo caso, assegnati ad arbitrio i primi tre termini della serie  $1, 2x, 3x^2$ , che vengono riscritti nell'ordine  $(3x^2, 2x, 1) \equiv (3, 2, 1)$ , gli altri si ricavano dalla moltiplicazione con la scala di relazione:

$$D = (3 \times 3 - 2 \times 2 + 5 \times 1)x^3 = 10x^3.$$

Per ottenere il termine successivo si moltiplica termine a termine la scala di relazione  $(3, -2, 5)$  con  $(10x^3, 3x^2, 2x) \equiv (10, 3, 2)$  ottenendo

$$E = (3 \times 10 - 2 \times 3 + 5 \times 2)x^4 = 34x^4$$

e così via. Nella *Miscellanea* [7], De Moivre aveva insegnato a calcolare la somma di una serie ricorrente utilizzando l'approccio consueto: esempi particolari, enunciato della legge generale e sua dimostrazione. Seguiamo il problema I nella *Miscellanea* ([7], p. 72) in cui si chiede di trovare la somma di una serie ricorrente

$$a + bx + cx^2 + dx^3 + ex^4 + \dots = P + Q + R + S + T + \dots := z$$

con scala di relazione  $(f, -g) \equiv (fx, -gx^2)$ . I due termini iniziali della serie vanno prescritti ad arbitrio:  $a = P$  e  $bx = Q$  e considerati nell'ordine  $(Q, P)$  si parte con l'impiego della scala di relazione per trovare

$$\begin{aligned} P &= P \\ Q &= Q \\ R &= fQx - gPx^2 \\ S &= fRx - gQx^2 \\ T &= fSx - gRx^2 \\ &\dots \end{aligned}$$

<sup>21</sup>I call that a *recurring* series which is so constituted, that having taken at pleasure any number of its terms, each following term shall be related to the same number of preceding terms, according to a constant law of relation.

Le due colonne di destra hanno somme rispettivamente  $P + Q + fx(z - P)$  e  $-gx^2z$  per cui si ottiene

$$z = \frac{P + Q - fPx}{1 - fx + gx^2} = \frac{a + (b - af)x}{1 - fx + gx^2}. \quad (4.12)$$

Quando la scala di relazione è formata da tre termini  $(f, -g, h) \equiv (fx, -gx^2, hx^3)$  la somma  $z$  si ottiene con procedimento analogo

$$z = \frac{a + (b - af)x + (c - bf + ga)x^2}{1 - fx + gx^2 - hx^3},$$

risultato che suggerì la regola di formazione generale della somma  $z$  quando la scala di relazione sia composta da un numero arbitrario di termini, schematizzata da De Moivre nel *Doctrine of Chances* con una serie di passaggi:

1. si prendano tanti termini della serie quanti ne sono presenti nella scala di relazione;
2. si sottragga da 1 la scala di relazione: il risultato è detto *scala differenziale*  $\Delta$ ;
3. si moltiplichi la scala differenziale per la somma dei termini della serie individuati al primo passo, trascurando i termini ridondanti e si indichi con  $\Sigma$  il risultato ottenuto:

la somma della serie ricorrente è  $\frac{\Sigma}{\Delta}$ . Osserviamo che l'argomento di De Moivre è formale, perché non vi è alcuna attenzione alla convergenza della serie, quando la si prolunghi all'infinito. D'altra parte De Moivre era interessato ad applicazioni in cui occorreva calcolare un numero *finito* di addendi. In questo caso bisogna essere accorti per considerare opportunamente il troncamento.

Servirsi della definizione di serie ricorrente per ottenerne un termine qualsiasi è poco efficace quando il termine cercato è sufficientemente lontano da quello iniziale. Per questo De Moivre trasformò una serie ricorrente in più progressioni geometriche. Anzitutto, nel Capitolo II della *Miscellanea*, egli mostrò come rileggere una progressione geometrica come serie ricorrente; presa la progressione geometrica di elementi successivi  $A, B, C, D, E, \dots$  e di ragione  $m$ , si ha

$$B = mA, \quad C = mB, \dots$$

Si moltiplichi la prima relazione per un numero *qualsiasi*  $p \neq 0$ , ottenendo  $pB - mpA = 0$  e quindi

$$C = mB = mB + 0 = mB + pB - mpA = (m + p)B - mpA \quad (4.13)$$

che ha la struttura di serie ricorrente, con scala di relazione  $(m + p, -mp)$ . Seguendo De Moivre, consideriamo ora due progressioni geometriche: la prima, di elementi  $A, B, C, \dots$  e ragione  $m$ ; la seconda di elementi  $H, K, L, \dots$  e ragione  $p$ . Grazie alla (4.13), possiamo scrivere

$$C = (m + p)B - mpA \quad L = (m + p)K - mpH \quad (4.14)$$

ed ottenere ancora una serie ricorrente

$$C + L = (m + p)(B + K) - mp(A + H). \quad (4.15)$$

Quello che sembra un esercizio formale acquista un significato più profondo con il corollario II ([7], pp. 29-30) che è una interpretazione algebrica della (4.15). De Moivre osservò infatti che i coefficienti della scala di relazione (4.15) sono funzioni simmetriche delle due quantità  $m$  e  $p$  che si possono ottenere risolvendo l'equazione

$$x^2 - (m + p)x + mp = 0,$$

risultato che si può estendere al caso in cui tre o più progressioni geometriche sono combinate per formare una serie ricorrente. L'interesse di questa osservazione si comprende con il corollario III ([7], p. 30) dove De Moivre osservò che, data una serie ricorrente con scala di relazione  $(a, -b)$  e costruita l'equazione di secondo grado

$$x^2 - ax + b = 0,$$

le cui radici siano  $m$  e  $p$ , allora la serie ricorrente può spezzarsi in due progressioni geometriche che hanno tali radici come ragioni.

Vediamo di applicare il procedimento all'equazione (4.11), la cui equazione caratteristica è

$$x^2 - 4x + 2 = 0$$

di radici  $m = 2 + \sqrt{2}$  e  $p = 2 - \sqrt{2}$ . Il termine di posizione  $\ell$  ha la forma

$$Am^\ell + Bp^\ell \quad (4.16)$$

dove le costanti  $A$  e  $B$  sono da determinare inserendo opportune condizioni iniziali. Osserviamo che lo schema (4.2) parte dalle probabilità che il gioco non termini in 4 od un numero pari ma più grande di partite. Mancano due termini iniziali, quelli che danno le probabilità che il gioco non termini in 0 e 2 partite, che sono entrambe pari ad 1 e che corrispondono a numeratori formati inserendo tutti i termini degli sviluppi di  $1 = (a + b)^0$  e di  $(a + b)^2$ , rispettivamente. Per trovare i coefficienti  $\alpha_m$ , poniamo  $\ell = 0$  ed  $\ell = 1$  in (4.16) ed osserviamo che i primi due coefficienti  $\alpha_0$  ed  $\alpha_1$  sono rispettivamente 0 ed 1 per cui la soluzione del sistema

$$\begin{cases} A + B = 0 \\ Am + Bp = 1 \end{cases}$$

è

$$A = \frac{1}{m - p} \quad B = -\frac{1}{m - p}$$

ed il coefficiente  $\alpha_\ell$  è uguale a

$$\alpha_\ell = \frac{1}{m - p}(m^\ell - p^\ell).$$

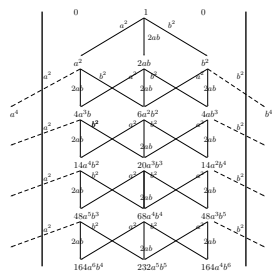


Figura 4.2: Schema per ottenere i coefficienti  $\alpha_m$  che formano la serie ricorrente (4.11).

A titolo di esempio verifichiamo che  $\alpha_2 = 4$  ed  $\alpha_3 = 14$ . Infatti, per  $\ell = 2$  abbiamo

$$\alpha_2 = \frac{1}{m-p}(m^2 - p^2) = m + p = 4$$

e

$$\alpha_3 = \frac{1}{m-p}(m^3 - p^3) = \frac{1}{m-p}(m-p)(m^2 + mp + p^2) = m^2 + mp + p^2 = 14.$$

Il Teorema IV della *Miscellanea analytica* ([7], pp. 33-35) è di particolare interesse, sia per le successive applicazioni, sia perché presenta il cosiddetto determinante di Vandermonde (1735-1796), qualche anno prima della nascita di questo matematico francese. De Moivre vuole scomporre una serie ricorrente

$$a + br + cr^2 + dr^3 + er^4 + \dots \quad (4.17)$$

con scala di relazione  $(f, -g, h, -k)$  in quattro progressioni geometriche. Per questo egli introduce le radici  $m, p, q$  ed  $s$  dell'equazione

$$x^4 - fx^3 + gx^2 - hx + k = 0$$

e costruisce le quattro progressioni

$$\begin{aligned} A + Am + Am^2 + Am^3 + Am^4 + \dots \\ B + Bp + Bp^2 + Bp^3 + Bp^4 + \dots \\ C + Cq + Cq^2 + Cq^3 + Cq^4 + \dots \\ D + Ds + Ds^2 + Ds^3 + Ds^4 + \dots \end{aligned}$$

che riprodurranno la serie ricorrente se le quantità  $A, B, C$ , e  $D$  risolveranno il sistema lineare

$$\begin{cases} A + B + C + D = a \\ Am + Bp + Cq + Ds = br \\ Am^2 + Bp^2 + Cq^2 + Ds^2 = cr^2 \\ Am^3 + Bp^3 + Cq^3 + Ds^3 = dr^3 \end{cases}$$

in cui la matrice dei coefficienti è proprio la matrice di Vandermonde

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ m & p & q & s \\ m^2 & p^2 & q^2 & s^2 \\ m^3 & p^3 & q^3 & s^3 \end{pmatrix}$$

il cui determinante è il prodotto di tutte le differenze delle quantità  $m, p, q, s$ . Pertanto

$$A = \frac{dr^3 - (p + q + s)cr^2 + (pq + ps + qs)rb - pqsa}{(m-p)(m-q)(m-s)} :$$

formule analoghe valgono per  $B, C, D$  e De Moivre ne fornì la descrizione verbale per scale di relazioni formate da un numero arbitrario di termini. La somma

(infinita) dei termini della serie ricorrente (4.17) si può allora esprimere come somma delle quattro progressioni geometriche ora ottenute:

$$\frac{A}{1-m} + \frac{B}{1-p} + \frac{C}{1-q} + \frac{D}{1-s}$$

o, se i termini da sommare sono  $\ell$ :

$$A \frac{1-m^{\ell+1}}{1-m} + B \frac{1-p^{\ell+1}}{1-p} + C \frac{1-q^{\ell+1}}{1-q} + D \frac{1-s^{\ell+1}}{1-s}$$

Come detto, De Moivre si accorse dell'utilità di queste serie per trattare il problema della durata del gioco:

Il metodo di determinare la probabilità che il gioco non finisca entro un numero assegnato di partite non è altro che quello per trovare un termine di una serie ricorrente.<sup>22</sup> ([8], p. 230)

Abbiamo già detto che la scala di relazione dipende dal valore del numero  $n$  di gettoni in possesso di ciascun giocatore ma tutte le serie hanno in comune questa proprietà:

$$p_n(0) = p_n(2) = \dots p_n(n-2) = 1,$$

dal momento che il gioco *non* può terminare in un numero di partite inferiore ad  $n$ .

Questo metodo soffre tuttavia di un inconveniente tecnico dal momento che il grado dell'equazione algebrica che fornisce le ragioni delle progressioni geometriche usate per scomporre la serie ricorrente cresce con il numero di termini che ne formano la scala di relazione. Per ovviare a questa difficoltà, De Moivre espose, con una dimostrazione incompleta, un approccio trigonometrico alla soluzione dell'equazione relativa alla scala di relazione, che permette un notevole risparmio di tempo basandosi sull'identità formale tra le formule ottenute per trovare la probabilità che il gioco non termini in  $n+d$  passi e quella che esprime  $\cos n\alpha$  in funzione di  $\cos \alpha$  [16].

### 4.3 Intermezzo: la ripartizione della posta tra tre o più giocatori in De Moivre

A partire dalla *Miscellanea Analytica* del 1730, De Moivre propose un metodo per evitare la prolissità del metodo delle combinazioni di Fermat e dello schema iterativo di Huygens. De Moivre, come Montmort, aveva esposto una versione algebrica del metodo di Fermat nella *Doctrine of Chances*, basata sullo sviluppo del trinomio e su una gestione separata dei casi dubbi. Supponiamo, per semplicità, di avere tre giocatori  $G_1$ ,  $G_2$  e  $G_3$  le cui abilità siano tra loro come i numeri  $a$ ,  $b$  e  $c$ . Supponiamo infine che il gioco venga terminato quando ai tre

<sup>22</sup>The method of determining the probability of the play's not ending in any number of game given, is no more than the finding of a term in a recurrent series.

giocatori mancano, nell'ordine,  $p$ ,  $q$ , ed  $r$  punti cosicché la sfida avrà termine al più dopo  $p + q + r - 2$  partite, se ciascuna assegna un punto. Per determinare la frazione della posta che spetta, per esempio, a  $G_1$ , De Moivre sviluppò

$$(a + b + c)^{p+q+r-2}.$$

Gli esponenti di  $a$ ,  $b$ ,  $c$  nei singoli termini dello sviluppo rappresentano il numero di vittorie ottenute dai tre giocatori se la partita si svolgesse secondo la modalità descritto dal termine in esame. Si assegnano allora ai singoli giocatori i termini (moltiplicati per i coefficienti dello sviluppo) che conferiscono la vittoria ad uno solo di essi e si considerano a parte i casi dubbi. Per esempio, se  $p = 1$  e  $q = 2$  ed  $r = 3$ , nello sviluppo di

$$(a + b + c)^{p+q+r-2} = (a + b + c)^4$$

sono da ascrivere a  $G_1$  i termini

$$a^4 + 4a^3b + 4a^3c + 6a^2c^2 + 12a^2bc + 12abc^2$$

perché in essi  $G_1$  ha vinto almeno una partita, *senza* che gli altri giocatori risucissero a vincerne, rispettivamente, 2 o 3. Similmente, vanno assegnati a  $G_2$  i termini  $b^4 + 4b^3c + 6b^2c^2$  e a  $G_3$  i termini  $c^4 + 4bc^3$ . Un termine ambiguo come  $6a^2b^2$  viene spezzato in questo modo

$$6a^2b^2 = a^2b^2 + abab + ab^2a + ba^2b + baba + b^2a^2$$

e ciascun monomio va interpretato in base all'*ordine* con cui compaiono le abilità dei singoli giocatori per cui  $a^2b^2$  significa che  $G_1$  vince le prime due partite e  $G_2$  le ultime due: il termine va dunque attribuito a  $G_1$ . Al contrario  $b^2a^2$  va ascritto a  $G_2$  che raggiunge prima degli altri giocatori il punteggio richiesto: in definitiva, i termini da attribuire a  $G_1$  ammontano a  $5a^2b^2$ , mentre a  $G_2$  va attribuito  $a^2b^2$ . Si procede similmente per gli altri termini ambigui e, raggruppati gli addendi che vengono attribuiti ai singoli giocatori, si divide tutto per  $(a+b+c)^4$  ottenendo le frazioni della posta spettanti ai singoli giocatori. È chiaro che l'ultimo passaggio deriva dal fatto che De Moivre non introdusse subito le probabilità di successo ma, separatamente, i casi favorevoli, normalizzando il risultato alla fine del procedimento. De Moivre introdusse la soluzione ricorsiva, come quella di Pascal ed Huygens, per formare una tabella con i valori delle ripartizioni, utile quando non manchino troppi punti alla fine del gioco. Nella *Miscellanea Analytica* De Moivre introdusse un'astuta semplificazione dei casi dubbi, suddivisa in più passi. Per ottenere la frazione della posta che spetta al giocatore di abilità  $a$ :

1. scrivi l'unità;
2. scrivi le lettere  $b$ ,  $c$ ,  $d$ , che esprimono le abilità degli altri giocatori;
3. combina queste lettere a due a due, poi a tre a tre, a quattro a quattro;
4. elimina, tra le combinazioni ottenute al passo precedente, quelle che fanno vincere certamente uno degli altri giocatori perché contengono  $b^q$  o  $c^r$ ,...;

5. moltiplica per  $a^{p-1}$  tutti i termini ottenuti sinora;
6. moltiplica ciascun termine per il numero delle sue permutazioni, cioè il numero di modi diversi in cui si può scrivere;
7. raggruppa i monomi che hanno lo stesso grado;
8. somma i termini all'interno di ogni gruppo e disponi le somme in ordine crescente, in base al grado dei monomi da cui sono composte;
9. detta  $S = a + b + c + d + \dots$  dividi i termini dei singoli gruppi, rispettivamente per  $S^{p-1}, S^p, S^{p+1}, \dots$ ;
10. moltiplica il tutto per  $\frac{a}{S}$ : il risultato ottenuto è la frazione di posta che spetta al giocatore scelto.

La semplificazione è attuata al punto 4 perché con l'eliminazione che si prescrive di fare, scompaiono automaticamente i termini ambigui. La moltiplicazione prescritta al passo 6 tiene conto del numero di modi in cui può presentarsi un certo monomio, cioè il numero di schemi che compongono una partita vinta da  $G_1$  con un certo punteggio.

Seguiamo, con qualche commento, l'applicazione al caso di tre giocatori cui manchino  $p = 2$ ,  $q = 3$  ed  $r = 5$  punti.

- Passi 1-4.

$$1, b + c, b^2 + bc + c^2, b^2c + bc^2 + c^3, b^2c^2 + bc^3 + c^4, b^2c^3 + bc^4, b^2c^4$$

Il grado massimo dei monomi da considerare è  $q + r - 2 = 6$ .

- Passo 5. Dopo la moltiplicazione per  $a^{p-1} = a$  e per l'opportuno numero di combinazioni, si ha

$$a, 2ab + 2ac, 3ab^2 + 6abc + 3ac^2, 12ab^2c + 12abc^2 + 4ac^3, \\ 30ab^2c^2 + 20abc^3 + 5ac^4, 60ab^2c^3 + 30abc^4, 105ab^2c^4$$

dove per il calcolo dei coefficienti di  $ab^k c^h$  De Moivre usò la formula  $\frac{(h+k+1)!}{k!h!}$ .

- Passi 7-9. L'espressione per la frazione di posta da assegnare a  $G_1$  è

$$\frac{a}{S} \left[ \frac{a}{S} + \frac{2a(b+c)}{S^2} + \frac{3a(b^2+2bc+c^2)}{S^3} + \frac{4a(3b^2c+3bc^2+c^3)}{S^4} + \right. \\ \left. \frac{5a(6b^2c^2+4bc^3+c^4)}{S^5} + \frac{30abc^3(2b+c)}{S^6} + \frac{105ab^2c^4}{S^7} \right]$$

Se osserviamo che  $\frac{a}{S}, \frac{b}{S}, \frac{c}{S}, \dots$  sono le probabilità di successo  $p_i$  del giocatore  $G_i$  in ciascuna partita, possiamo dare alla formula trovata da De Moivre nel caso in cui vi siano  $n$  giocatori, dicendo per comodità di notazione  $a_i$  i punti che mancano a  $G_i$  per vincere, la forma seguente

$$\sum_{x_2=0}^{a_2-1} \dots \sum_{x_n=0}^{a_n-1} \frac{(a_1 - 1 + x_2 + \dots + x_n)!}{(a_1 - 1)!x_2!x_3! \dots x_n!} p_1^{a_1-x_2-x_3-\dots-x_n} p_2^{x_2} p_3^{x_3} \dots p_n^{x_n} : \quad (4.18)$$

le probabilità di successo per gli altri giocatori si ottengono in modo analogo.

## 4.4 Un gioco strategico

Tra i problemi di cui Montmort propose la soluzione nell'*Essay*, particolarmente intrigante è il gioco *Le Her* (il galantuomo), un gioco di carte nato in Francia nel XVI secolo e noto anche con i nomi di *le Hère* o *coucou*. Montmort lo descrisse nella versione a quattro giocatori nella prima edizione dell'*Essay* ed in quella a tre giocatori nella seconda edizione ([19], pp. 278-279), basandovi un problema di ripartizione della posta proposto ai lettori per la soluzione. In effetti, il numero di giocatori mutava durante la partita perché questi erano eliminati quando perdevano tutti i gettoni in loro possesso all'inizio del gioco. Nella corrispondenza con Nicholas Bernoulli, Montmort si concentrò sullo stadio finale del gioco, quando i giocatori sono due, osservando che il problema presentava delle difficoltà singolari. Ecco dunque le regole del gioco nel caso di due giocatori, che diremo *A* e *B*. Si gioca con un mazzo di 52 carte suddivise in quattro semi, numerate da 1 a 10 e con le figure *J*, *Q*, *K* che equivalgono ad 11, 12 e 13, rispettivamente. Supponiamo che *B* sia di turno a distribuire le carte. Dal mazzo egli estrae una carta per *A* ed una per sé. Se *A* non è soddisfatto della carta ricevuta, può obbligare *B* allo scambio e *B* dovrà accettare, *a meno di* non avere un *K*. Dal canto suo, se *B* non ha un *K* può estrarre un'altra carta dal mazzo e tenerla se il punteggio non è inferiore a quello della carta in suo possesso. Se però *B* pescherà un *K* egli è costretto a tenere la carta che voleva scambiare. Operata la scelta, i giocatori scoprono la carta in loro possesso: chi realizza il punteggio più alto vince e, in caso di parità è *B* a vincere. Il gioco non è dunque di solo azzardo ma il suo esito dipende anche dalla *strategia* seguita dai giocatori. Bernoulli, in una lettera dell'11 novembre 1711, scrisse a Montmort comunicando tra l'altro di aver trovato il vantaggio di un giocatore rispetto all'altro, supponendo che entrambi scelgano la strategia ritenuta più vantaggiosa: *A* deve tenere la carta ricevuta se questa è superiore a 7 mentre deve cambiarla in caso contrario; *B* dovrà tenere la carta ricevuta se superiore ad 8, cambiarla in caso contrario. Con queste scelte, Bernoulli afferma che il rapporto tra le probabilità di successo di *B* e di *A* è di 2828 contro 2796, risultato che si può anche ottenere quando *A* decida di tenere il 7, purché allora *B* decida di tenere l'8. Laconicamente, Bernoulli aggiunse:

È tuttavia più vantaggioso per Paolo (il giocatore *A*) di non tenersi il sette, piuttosto che tenerlo, ciò che costituisce un enigma che vi lascio da sviluppare.<sup>23</sup> ([19], p. 334)

Oltre a Montmort e Bernoulli, la discussione intorno a questo gioco coinvolse due conoscenti di Montmort: l'abate di Orbais, che risiedeva poco lontano

<sup>23</sup>Il est néanmoins plus avantageux pour Paul de ne pas s'en tenir à un sept que de s'y tenir, ce qui est une énigme que je vous laisse à développer.

dalla tenuta della famiglia Montmort, e un nobile inglese, Charles Waldegrave. Secondo l'abate di Orbais non era possibile risolvere il problema, dal momento che  $A$  non può decidere fino a quale carta richiedere lo scambio, senza sapere quale sarà la strategia di  $B$  e viceversa, generando un circolo vizioso. L'accordo tra i risultati di Bernoulli e Montmort obbligò però gli interlocutori a riconsiderare seriamente il problema. Per seguire l'argomento di Waldegrave e dell'abate d'Orbais, vediamo un esempio dei calcoli necessari a trovare le aspettative dei due giocatori, calcoli che furono sviluppati completamente dal matematico svizzero Jean Trembley (1749-1811), in un lavoro del 1804 [17]. Supponendo che un giocatore riceva 1 in caso di vittoria e 0 altrimenti, vediamo come impostare il calcolo dell'aspettazione di  $A$  quando, avendo egli un 7, decida di obbligare  $B$  allo scambio. Tra le altre 51 carte, le 24 con punteggi inferiori al 7, le altre 3 carte col numero 7 ed i 4  $K$  faranno perdere  $A$  mentre le 20 carte restanti, dagli 8 a  $Q$  gli daranno un'aspettazione  $x$ . In definitiva, l'aspettazione  $y$  di  $A$ , quando egli ha un 7 e lo cambia, è

$$y = \frac{31 \times 0 + 20 \times x}{51} = \frac{20x}{51}.$$

Per calcolare il valore di  $x$  si passa al giocatore  $B$  che deve per forza pescare una carta dal mazzo, se vuole vincere. Se  $B$  aveva un 8, ora passato ad  $A$ , egli ha 50 carte tra cui scegliere e di queste sono a suo sfavore i 4  $K$ , i tre 7 rimasti e le 24 carte con punteggio inferiore a 7, mentre restano a suo vantaggio i 9, 10,  $J$  e  $Q$ , oltre ai tre 8. In definitiva l'aspettazione di  $B$ , quando aveva ricevuto un 8 che ha dovuto cedere ad  $A$ , che a sua volta aveva ricevuto un 7, è

$$\frac{19 \times 1 + 31 \times 0}{50} = \frac{19}{50}.$$

Con un ragionamento simile si trova che le aspettative di  $B$ , se avesse ricevuto un 9, un 10, un  $J$  od una  $Q$  che poi ha dovuto cedere ad  $A$ , sono rispettivamente:  $\frac{15}{50}$ ,  $\frac{11}{50}$ ,  $\frac{7}{50}$  e  $\frac{3}{50}$ . In definitiva, l'aspettazione di  $B$  è

$$\frac{19 + 15 + 11 + 7 + 3}{5 \times 50} = \frac{11}{50}$$

da cui segue il valore di  $x = \frac{39}{50}$  e quindi l'aspettazione di  $A$ , quando decide di non tenere il 7 ricevuto all'inizio, è

$$y = \frac{780}{2550}.$$

Questo valore non dipende dalla strategia adottata dal giocatore  $B$  che, in questo caso, è obbligata ma, se  $A$  tenesse il 7 ricevuto, lo scenario cambierebbe. Infatti in questo caso, se  $B$ , ricevuto un 8, lo tenesse, ci sono tre possibilità:

- 1)  $B$  ha ricevuto una delle 24 carte con punteggio superiore a 7. Allora  $A$  perde certamente e la sua aspettazione è nulla.
- 2) Anche  $B$  ha ricevuto uno dei tre 7 rimasti. Diremo  $z$  l'aspettazione di  $A$ .
- 3)  $B$  ha ricevuto una delle 24 carte con punteggio inferiore al 7. Diremo  $x$  l'aspettazione di  $A$ .

L'aspettazione  $y$  di  $A$  è

$$y = \frac{24 \times 0 + 3 \times z + 24 \times x}{51} = \frac{3z + 24x}{51}.$$

Ora, se  $B$  ha uno dei tre 7 restanti,  $A$  potrà vincere solo se  $B$  estrae dal mazzo una delle 24 carte con punteggio inferiore al 7, mentre perderà nei 26 casi restanti, per cui  $z = \frac{24}{50}$ . D'altro canto, se  $B$  ha una carta inferiore a 7,  $A$  vincerà quando  $B$  dovesse pescare uno dei 4  $K$ , ovvero una delle altre 23 carte con punteggio inferiore a 7, per cui  $x = \frac{27}{50}$  e, in definitiva,

$$y = \frac{720}{2550}.$$

Infine, se la strategia di  $B$  è di scambiare anche un 8,  $A$ , avendo tenuto il 7 ricevuto, perderebbe sicuramente se  $B$  avesse ricevuto, tra le 51 altre carte, una delle 20 con punteggio superiore ad 8, mentre l'aspettazione di  $A$  varrebbe  $z$ , se  $B$  avesse un 8 od uno degli altri tre sette e varrebbe  $x$  se  $B$  avesse una carta con punteggio inferiore a 7. Per calcolare  $x$  e  $z$ , si osserva che per  $B$  nulla è cambiato rispetto a prima per cui

$$y = \frac{20 \times 0 + 7 \times z + 24 \times x}{51} = \frac{7 \times 24 + 24 \times 27}{2550} = \frac{816}{2550}.$$

Poiché, a seconda della strategia di  $B$ , le aspettative di  $A$  quando decide di tenere il 7 si modificano, Waldegrave e l'abate d'Orbais ritenevano che ci dovesse essere una motivazione di un certo peso per convincere  $A$  a cambiare un 7. Ora, un discorso analogo vale per il giocatore  $B$ , cui serve una motivazione di peso generalmente diverso per cambiare il proprio 8. Il peso che convince  $A$  a cambiare un 7, convince anche  $B$  a cambiare un 8 ma ciò che porta  $B$  a prendere questa decisione, porta  $A$  a tenersi un 7. Dunque i giocatori dovrebbero tenere due comportamenti contraddittori e da questo stallo, l'abate d'Orbais e Waldegrave concludevano l'impossibilità di risolvere il problema. Bernoulli confermò la correttezza dei calcoli dei suoi interlocutori, pur ritenendo un sofisma il loro ragionamento dal momento che, una volta che  $A$  ha deciso di cambiare il 7, ignorando ciò che farà  $B$  che, conoscendo questa decisione, cambierebbe l'8, non è più lecito ritornare al giocatore  $A$  per giungere alla contraddizione, dal momento che così facendo si sta ipotizzando nello stesso momento che  $A$  prima ignori e poi conosca la strategia di  $B$ . Per comprendere le posizioni, seguiamo i calcoli di Trembley [17]. Anzitutto, il calcolo visto sopra va ripetuto per tutte le altre carte che  $A$  potrebbe avere ricevuto e, fissata questa carta, occorre calcolare quattro aspettative per  $A$ , una per ciascuna di queste strategie di gioco:

$S_{11}$  :  $A$  tiene carte dal 7 in su;  $B$  tiene carte da 8 in su.

$S_{12}$  :  $A$  tiene carte dal 7 in su;  $B$  tiene carte da 9 in su.

$S_{21}$  :  $A$  tiene carte dall'8 in su;  $B$  tiene carte da 8 in su.

$S_{22}$  :  $A$  tiene carte dall'8 in su;  $B$  tiene carte da 9 in su.

Calcolate le 13 aspettative di  $A$ , ciascuna condizionata dal valore della carta ricevuta, si ottiene questa tabella in cui le quattro colonne più a destra indicano le aspettative di  $A$  per le quattro strategie descritte, moltiplicate per  $25 \times 51$ :

carta di $A$	$S_{11}$	$S_{12}$	$S_{21}$	$S_{22}$
$K$	1200	1200	1200	1200
$Q$	1058	1052	1058	1052
$J$	902	888	902	888
10	746	724	746	724
9	590	560	590	560
8	434	476	434	476
7	360	408	390	390
6	444	444	444	444
5	490	490	490	490
4	528	528	528	528
3	558	558	558	558
2	580	580	580	580
1	594	594	594	594

Ad ogni strategia  $S_{ij}$  si associa una aspettazione  $m_{ij}$  per  $A$  ottenuta sommando tutte le aspettative che stanno nella colonna corrispondente ad  $S_{ij}$  e dividendo per 13 la somma, dal momento che tutte le carte ottenute da  $A$  hanno la stessa probabilità di presentarsi:

$$m_{11} = \frac{2828}{5525} \quad m_{12} = \frac{2834}{5525}$$

$$m_{21} = \frac{2838}{5525} \quad m_{22} = \frac{2828}{5525}$$

le aspettative  $m'_{ij} = 1 - m_{ij}$  di  $B$  si ottengono immediatamente

$$m'_{11} = \frac{2697}{5525} \quad m'_{12} = \frac{2691}{5525}$$

$$m'_{21} = \frac{2687}{5525} \quad m'_{22} = \frac{2697}{5525}.$$

Se  $A$  decide di tenere anche il 7 ma ignora la strategia di  $B$ , la sua aspettazione, se le strategie di  $B$  sono pesate allo stesso modo, sarà

$$m_{A1} = \frac{m_{11} + m_{12}}{2} = \frac{2831}{5525}$$

mentre, se decide di scambiare anche il 7, ignorando ciò che fa  $B$ , questa aspettazione diventa

$$m_{A2} = \frac{m_{21} + m_{22}}{2} = \frac{2811}{5525}.$$

Similmente se  $B$  tiene anche l'8, la sua aspettazione, ignorando ciò che fa  $A$ , le cui strategie sono pesate allo stesso modo, è

$$m_{B1} = \frac{m'_{11} + m'_{21}}{2} = \frac{2692}{5525},$$

che diventa

$$m_{B2} = \frac{m'_{12} + m'_{22}}{2} = \frac{2694}{5525}$$

se sceglie di cambiare anche l'8 ma ignora la scelta di  $A$ . Ecco come Trembley riassume il dilemma:

Si vede dunque che se Pietro [il giocatore  $B$ ] tiene l'otto, a Paolo [il giocatore  $A$ ] conviene cambiare il sette<sup>24</sup> e che, se Pietro cambia l'otto, conviene che Paolo tenga il sette<sup>25</sup>. Il sig. Montmort ed i suoi amici conclusero, contro il parere di Nicholas Bernoulli, che questo problema è insolubile perché, dicevano, se Paolo sapesse che Pietro tiene l'otto, scambierebbe il sette ma, se Pietro viene a sapere della scelta di Paolo di scambiare il sette, scambierebbe l'otto, il che costituisce un circolo vizioso.

Torniamo ora alla corrispondenza tra Montmort, Bernoulli ed i loro interlocutori. Né Waldegrave né l'abate di Orbais si dichiararono convinti dagli argomenti di Bernoulli che, a sua volta, pur ammettendo che ogni scelta di  $A$  potrebbe essere usata da  $B$  a detrimento di  $A$  stesso, insisteva nel non ritenere indifferenti le strategie di  $A$ , se tenere o cambiare un 7. Se  $A$  non sapesse quale decisione prendere, potrebbe affidarsi al caso inserendo in un'urna un ugual numero di gettoni bianchi e neri e decidendo che, se l'estrazione darà un gettone bianco, egli manterrà il 7 mentre lo cambierà, estraendo un gettone nero. In questo modo, secondo i calcoli di Bernoulli, non riportati in esteso, la sua aspettazione calerebbe leggermente portandosi da 780/2550 a 774/2550. Per questo è meglio che  $A$  si mantenga nel proposito di cambiare un 7 e che  $B$  cambi un 8.

Bernoulli insistette nel difendere il proprio argomento dalle obiezioni ricevute per poi chiarire il comportamento ottimale che debbono seguire i giocatori:  $A$ , volendo tenere il 7, deve chiedersi quali sono le sue probabilità di vittoria nel caso che  $B$  scelga o meno di tenersi un 8. Similmente  $A$  dovrebbe porsi la stessa domanda se volesse cambiare il 7 e scegliere la strategia che lo espone al rischio minore.

Non riuscendo a far breccia in Waldegrave e nell'abate d'Orbais, Bernoulli chiese a Montmort di dare un parere, che si trova in una lettera del 15 novembre 1713. Montmort è d'accordo con Bernoulli nel contestare la prima soluzione di Waldegrave e dell'abate d'Orbais per la quale era indifferente per  $A$  tenere o meno un 7 e per  $B$  tenere o scambiare un 8. Montmort sostenne d'altra parte ([19], p. 403):

sono però persuaso che la soluzione del problema sia impossibile, cioè che non si possa prescrivere la condotta di Paolo quando ha un sette né quella di Pietro, quando ha un 8.<sup>26</sup>

<sup>24</sup>Poiché  $m_{11} < m_{21}$ .

<sup>25</sup>Poiché  $m_{22} < m_{12}$ .

<sup>26</sup>Je suis au contraire persuadé que la solution du Problème est impossible, c'est à dire qu'on ne peut prescrire à Paul la conduite qu'il doit tenir quand il a un sept, et à Pierre quand il a un 8.

Per cercare una via di uscita, Montmort propose di calcolare la probabilità di vittoria di  $A$  supponendo che questi metta in un sacchetto  $a$  gettoni bianchi e  $b$  neri, decidendo di cambiare il 7 se estrae un gettone bianco e tenere il 7 se ne estrae uno nero.  $B$ , sapendo del ricorso di  $A$  a questa tecnica, metterà  $c$  gettoni bianchi e  $d$  neri in un altro sacchetto ed estrarrà a sorte, cambiando l'8 se estrae un gettone bianco e tenendolo se estrae un gettone nero. Montmort calcolò l'aspettazione  $p$  di  $A$  in funzione del numero di gettoni presenti nei sacchetti, ottenendo

$$p = \frac{m_{11}ac + m_{21}bc + m_{12}ad + m_{22}bd}{(a+b)(c+d)} \quad (4.19)$$

ovvero, numericamente

$$p = \frac{2828ac + 2834bc + 2838ad + 2828bd}{5525(a+b)(c+d)} : \quad (4.20)$$

è lo stesso procedimento utilizzato per il calcolo di  $m_{1A}$  ed  $m_{2A}$ , solo che le aspettative  $m_{ij}$  sono pesate dalle probabilità di estrarre un gettone di un certo colore da parte di ciascun giocatore.

Montmort studiò il comportamento di  $p$  nei casi limite in cui  $a \gg b$  e  $c \gg d$  ovvero quando  $a \ll b$  e  $c \ll d$  ma non riuscì a stimare il valore ottimale di  $p$ . Comunque, tutto ciò per lui confermava la presenza di un circolo vizioso ed infatti egli sostenne che la conclusione opposta da Waldegrave e dall'abate d'Orbais al ragionamento di Bernoulli non fosse ridicola perché, se ciascuno dei due giocatori conoscesse la decisione dell'altro e si comportasse razionalmente, le conclusioni non potrebbero essere diverse da quelle di Waldegrave, appoggiate dall'Abate d'Orbais. In conclusione, per Montmort ([19], p. 405)

da tutte le ragioni che sembrano obbligare Paolo [il giocatore  $A$ ] a scambiare anche il 7, ricavo solo che, nella pratica, egli deve attenersi alla regola di cambiare il 7 più spesso di quanto non lo tenga ma, quanto più spesso e, in particolare, ciò che deve fare (*hic et nunc*), che è il punto principale della questione, il calcolo non lo insegna e ritengo impossibile una decisione.<sup>27</sup>

Poco prima di chiudere la lettera, Montmort ricevette da Waldegrave alcune considerazioni sul valore di  $p$  contenuto nella (4.20) che si affrettò a trasmettere a Bernoulli. Waldegrave notò che, se  $a = 3$  e  $b = 5$ , il valore di  $p$  è *costante*, cioè indipendente dai valori di  $c$  e  $d$  e dunque dalla decisione adottata da  $B$  quando trova un 8. Non solo, il valore di  $p$  in questo caso è superiore a quello trovato da Bernoulli per cui Waldegrave poté concludere di aver mostrato la falsità delle conclusioni di Bernoulli. In effetti, dal punto di vista della odierna teoria dei giochi, Waldegrave ha trovato la strategia *mista* ottimale [13] e gli apparenti

<sup>27</sup>de toutes les raisons qui semblent devoir engager Paul à prendre la maxime de changer au 7, j'en conclus seulement qu'il sera bien dans la pratique de faire une lois de changer plus souvent au sept que de s'y tenir, mais combien plus souvent il doit changer que s'y tenir, et en particulier ce qu'il doit faire (*hic et nunc*) car c'est la principalement la question: le calcul n'apprend rien la dessus, et j'en tiens la décision impossible.

paradossi che avevano tormentato gli interlocutori sono dovuti al fatto che la matrice di *payoff*  $a_{ij}$ , che indica le aspettative dei giocatori a seconda della strategia  $i$  scelta da  $A$  tra le  $m$  possibili e quella  $j$  scelta da  $B$  tra le  $n$  possibili, non ha un punto di sella, cioè la disuguaglianza

$$\max_{i \leq m} \min_{j \leq n} a_{ij} \leq \min_{i \leq m} \max_{j \leq n} a_{ij}$$

vale in senso stretto. L'aspettazione (4.20) si può riscrivere in termini delle probabilità  $x = \frac{a}{a+b}$  ed  $y = \frac{c}{c+d}$  che ciascun giocatore ha di estrarre un gettone bianco, come ([13], p.500)

$$S(x, y) = \frac{2828xy + 2834(1-x)y + 2838x(1-y) + 2828(1-x)(1-y)}{5525} = \frac{2(5x+3y-8xy) + 2828}{5525} \quad (4.21)$$

o ancora, indicando con  $\alpha$  e  $\beta$  due opportune costanti

$$S(x, y) = \alpha s(x, y) + \beta \quad \text{con} \quad s(x, y) := 5x + 3y - 8xy.$$

Scrivendo

$$s(x, y) = (3 - 8x)y + 5x \quad (4.22)$$

si vede che, scelto  $x = x_0 = \frac{3}{8}$ , l'aspettazione di  $A$  è sempre pari a  $\frac{15}{8}\alpha + \beta$ . Ora, se  $A$  scegliesse un valore diverso per  $x$ ,  $B$  potrebbe sempre trovare un valore di  $y$  che riduce l'aspettazione di  $A$ . Infatti, posto  $x = x_0 + \Delta x$ , si ha

$$s(x_0 + \Delta x, y) = (5 - 8y)\Delta x + 5x_0$$

per cui, se  $\Delta x \neq 0$ ,  $B$  potrà sempre scegliere un valore  $y^*$  di  $y$  che *riduca* l'aspettazione di  $A$ , dimostrando quanto asserito. D'altro lato, scrivendo

$$s(x, y) = (5 - 8y)x + 3y, \quad (4.23)$$

si vede che, se  $B$  sceglie una strategia mista con  $y = y_0 = \frac{5}{8}$ , il valore di  $s(x, y_0) = \frac{15}{8}$  è indipendente da ciò che sceglie  $A$ . Se  $B$  utilizzasse una strategia diversa,  $A$  potrebbe sempre scegliere un valore  $x^*$  di  $x$  tale da aumentare la propria aspettazione. Il fatto che  $s(x_0, y) = s(x, y_0)$  è una manifestazione del teorema del "minimax", teorema fondamentale nella teoria dei giochi. È certamente curioso osservare come questo gioco strategico sia stato studiato e correttamente risolto da Waldegrave più di due secoli prima dell'avvento della teoria dei giochi iniziata da Émile Borel (1871-1956) nel 1921 e sviluppata considerevolmente a partire dal 1928 da Janos (John) von Neumann (1903-1957) [17] che contribuì in modo decisivo alla sua maturazione come disciplina matematica di alto livello, grazie soprattutto al poderoso volume *Theory of Games and Economic Behavior* scritto in collaborazione con l'economista austriaco Oskar Morgenstern (1902-1977) e pubblicato nel 1944 [18].

# Bibliografia

- [1] F.N. David: *Games, Gods and Gambling. A history of probability and statistical ideas*. General Publishing Company, Toronto, (1962). Ristampato da Dover, New York, (1998).
- [2] M. Dresher: *Games of strategy. Theory and applications*. Prentice-Hall, Englewood Cliffs, New Jersey, U. S. A., (1961).
- [3] A. De Moivre: De mensura sortis seu; de probabilitate eventuum in ludiis a caso fortuito pendentibus. *Transactions of the Royal Society of London*, **27**, 213-264, (1710-1712).
- [4] A. De Moivre: *Miscellanea analytica de seriebus et quadraturis*. Tonson & Watts, London, (1730).
- [5] A. De Moivre: *The Doctrine of Chances*. Millar, London, (1756).
- [6] J. De Witt: *Treatise on Life annuities*. In: R.G. Barnwell: *A sketch of the life and times of John De Witt, Grand Pensionary of Holland*. Pudney & Russel, New York, (1856).
- [7] L. Euler: Calcul de la probabilité dans le jeu de rencontre. *Mémoires de l'Académie des Sciences de Berlin*, **7**, (1753), 255-270.
- [8] A. Hald: *History of probability and Statistics and their applications before 1750*. Wiley-Interscience, Hokoben, New-Jersey, U.S.A. (1988).
- [9] J. Henny: Niklaus und Johann Bernoullis Forschungen auf den Gebiet der Wahrscheinlichkeitsrechnung in ihrem Briefwechsel mit Pierre Rémond de Montmort. In *Die Werke von Jakob Bernoulli*. Band 3, pp. 457-507, (1975).
- [10] M.G. Kendall: Studies in the history of probability and statistics. XVIII. Thomas Young on coincidences. *Biometrika*, **55**, (1968), 249-250.
- [11] J.-H. Lambert: Examen d'une espece de superstition ramenée au calcul des probabilités. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Berlin*, (1771), 411-420.
- [12] P.-S. de Laplace: *Théorie analytique des probabilités*. II Édition. Courcier, Paris, (1814).

- [13] J. von Neumann: Zur Theorie der Gesellschaftsspiele *Mathematische Annalen*, **100**, 295-320, (1928).
- [14] J. von Neumann, O. Morgenstern: *The Theory of Games and Economic Behavior*. Princeton University Press, Princeton, U.S.A., (1944).
- [15] P. Rémond de Montmort: *Essay d'Analyse sur les Jeux de Hazard*. II Édition. Quillau, Paris, (1713).
- [16] I. Schneider: Der Mathematiker Abraham De Moivre (1667-1754). *Archive for History of exact Sciences*, **5**, (1968), 177-317.
- [17] J. Trembley: Sur le calcul d'un jeu d'hazard. *Mémoires de l'Académie Royale des Sciences de Berlin, Classe de mathématique*, 88-102 (1804).
- [18] T. Young: Remarks on the probabilities of error in physical observations, and on the density of the Earth, considered, especially with regard to the reduction of experiments on the pendulum. *Philosophical Transactions of the Royal Society of London*, **109**, (1819), 70-95.

## Capitolo 5

# L' *Ars Conjectandi* di Jakob Bernoulli

### 5.1 Analisi dell'opera

L' *Ars Conjectandi* fu pubblicata postuma nel 1713, otto anni dopo la morte di Jakob Bernoulli (1654-1705). Il lungo lasso di tempo intercorso tra la morte dell'autore e la pubblicazione dell'opera è in parte dovuto alle incertezze sulla politica editoriale: se stampare il manoscritto nella versione originale incompleta oppure se trovare qualcuno che si facesse carico del non agevole compito di completare il trattato. Come ci informa Nicolaus Bernoulli<sup>1</sup> (1687-1759) nella prefazione di [4], la proposta di completare il manoscritto fu fatta in prima battuta al fratello di Jakob, Johann, che non aveva avuto buoni rapporti con il fratello e che, distratto dai molti impegni, declinò l'invito. La richiesta fu quindi rivolta a Nicolaus al suo ritorno in Svizzera dopo un lungo viaggio per l'Europa: non sentendosi all'altezza, anche Nicolaus rifiutò l'incarico. Notando però che il manoscritto era stato completato in massima parte, egli ne propose la stampa nello stato lasciato dall'autore e l' *Ars Conjectandi* poté vedere finalmente la luce. Si tratta di un punto di riferimento per le successive generazioni di studiosi che vi hanno trovato ispirazione. La ricchezza dell' *Ars Conjectandi* si comprende osservando come fautori delle varie impostazioni, a volte contrastanti, sulla teoria della probabilità vi abbiano fatto riferimento, vedendo in Jakob Bernoulli un progenitore del proprio approccio.

L'opera consta di quattro parti, di cui l'ultima incompiuta, e di un'appendice dedicata alle serie. La prima parte dell' *Ars Conjectandi* riproduce il *De Ratiociniis in Ludo Aleae* di Huygens cui Bernoulli aggiunse delle annotazioni. Segue una seconda parte dedicata al calcolo combinatorio ed una terza parte

---

<sup>1</sup>Il capostipite della dinastia Bernoulli fu Nicolaus Bernoulli (1623-1708) che fu il padre dei matematici Jakob e Johann (1667-1748) e di Nicolaus Bernoulli (1662-1716) che fu il padre del "nostro" Nicolaus. Da Johann Bernoulli nacque Daniel (1700-1782), che incontreremo nel capitolo dedicato al paradosso di San Pietroburgo.

nella quale sono esaminati dal punto di vista probabilistico i giochi d'azzardo più diffusi all'epoca. Infine, la quarta parte, quella su cui ci concentreremo maggiormente, è di capitale importanza perché Bernoulli intende allargare in modo considerevole il raggio di azione del calcolo delle probabilità. Il titolo stesso della sezione è programmatico:

Artis Conjectandi  
Pars Quarta  
tradens,  
Usum et applicationem praecedentis doctrinae  
in civilibus, moralibus et oeconomicis

Come vedremo, un progetto tanto ambizioso richiede necessariamente che si esca dai limiti angusti delle applicazioni ai giochi d'azzardo.

## 5.2 Le *Annotationes* di Jakob Bernoulli

La prima parte dell'*Ars conjectandi* di Jakob Bernoulli contiene la riproduzione del *De Ratiociniis* di Huygens ed una serie di *Annotationes* al testo, particolarmente interessanti sia come esempio di lettura critica di un testo matematico da parte di un altro matematico, sia come testimonianza dell'evoluzione della teoria della probabilità. In questa sezione presentiamo una selezione delle note di Bernoulli, rinviando per una analisi esaustiva a [5]. Nella prima *Annotatio* Bernoulli osserva come il termine *expectatio* non abbia il senso attribuitogli nel linguaggio comune dove indica una sorta di estremo superiore alle nostre speranze. Al contrario, il suo valore indica

qualcosa di intermedio tra l'ottimo che speriamo ed il pessimo che temiamo.<sup>2</sup> ([4], p. 5)

Anche Bernoulli vede nell'aspettazione una quantificazione del *giusto mezzo* aristotelico. Quanto al problema della divisione della posta, Bernoulli ribadisce senza ambiguità quanto già Cardano aveva osservato nella *Practica Arithmeticae*:

Infatti nel calcolo delle eventualità occorre tener conto solo delle partite da disputare, non di quelle già disputate.<sup>3</sup> ([4], p. 12)

Una annotazione molto interessante è quella sulla Proposizione X, laddove Huygens aveva calcolato la speranza matematica del giocatore che vuole ottenere un 6 in 4 lanci come  $\frac{671}{1296}a \simeq \frac{a}{2}$ :

se qualcuno, aspirando ad ottenere un 6 in quattro lanci, avesse circa la stessa aspettazione di vincere o perdere, cioè, se vincesse o perdesse

<sup>2</sup>per valorem ejus semper significetur intermedii quidpiam inter optimum quod speramus et pessimum quod metuimus.

<sup>3</sup>Adeoque in computandis sortibus solummodo futurorum lusuum, nulla praeteritorum habenda est ratio.

con ugual facilità, succederebbe che, giocando per un po', tante volte vincerebbe quante perderebbe, se possiede una fortuna equilibrata; a tal punto che, tante volte uscirà un sei su quattro lanci, quante volte esso non uscirà; per cui si otterrà un 6 ogni 8 lanci e dunque, ad esempio, 75 volte il 6 su 600 lanci.

Consideriamo ora altri 6 giocatori, che giochino con queste condizioni: il primo vince se esce l'1, il secondo se esce il 2, il terzo se esce il 3, ecc.; stabilita questa condizione certamente si affronteranno con una pari aspettazione; ma se giocano anche con pari fortuna, accadrà allora necessariamente che su 600 lanci 100 realizzeranno il 6: dunque, poiché si gioca con uguali possibilità ed ugual fortuna, in 600 lanci il 6 comparirà cento volte e meno di cento volte, ciò che è assurdo.<sup>4</sup> ([4], pp. 26-27)

Bernoulli esplora le conseguenze del risultato di Huygens esaminando cosa succederebbe se si ripetesse tante volte una sequenza di 4 lanci non solo in condizione di equa aspettazione (*aequa sors*) ma anche di fortuna equa (*aequa fortuna*). Questa seconda condizione appare molto simile all'*aequalitas* di Cardano perché presuppone una distribuzione *rigida* dei risultati che si possono presentare ripetendo un certo "esperimento" (il lancio di un dado). Bernoulli mette in guardia dalle interpretazioni superficiali del concetto di aspettazione equa e per questo compie un passo che si rivelerà decisivo per una interpretazione dei concetti probabilistici sulla base della ripetizione di un esperimento. L'assurdo nasce dal confronto tra due giochi per i quali l'aspettazione di vittoria a seguito dell'uscita del sei è diversa: circa  $\frac{a}{2}$  nel primo caso ed  $\frac{a}{6}$  nel secondo caso. La ripetizione di 600 lanci porta nel primo caso a 150 quaterne, nel secondo caso ogni partita si decide subito e quindi se ne giocano 600. Se allora la *fortuna* è equa, il 6 si deve presentare 75 volte nel primo caso, 100 nel secondo, il che appare assurdo a Bernoulli. Per uscire da questo stallo, egli entra nel dettaglio del primo gioco:

Perché questo errore venga corretto, affermo senza dubbio che, quando si giochi con una fortuna equa, allora il sei deve comparire cento volte su 600 lanci; nego però che, se qualcuno giocherà per ottenere un sei su quattro lanci, avrà per questo bisogno di quattro lanci; può succedere infatti che il sei si presenti al primo o al secondo o al terzo lancio, nel qual caso i lanci restanti vengono attribuiti alla quaterna successiva; cosicché per vincere una volta e perdere un'altra possono bastare meno di otto lanci. In qual modo ciò si adatti al vero, lo dimostro in questa maniera: suppongo che in tutte le quaterne di lanci dalle quali esco vincitore, il 6 sia sempre

<sup>4</sup>Si quis quatuor jactibus senarium jacere contendes, aequam circiter ad vincendum ac perdendum expectationem habeat, hoc est, aequae facile vincat ac perdat, fiet ut aliquandiu ludens toties vincat quoties perdit, si et pari fortuna utatur; adeoque ut e quaternis jactibus toties unus senarius existat, quoties e quaternis aliis nullus; quare in octonis jactibus unus reperietur senarius, ac proinde in sexcentis verb. gr. jactibus senarij 75. Sunt jam sex alij, qui ea conditione ludant, ut primus vincat, si unum punctum jaciatur, secundus si duo, tertius si tria etc. quo utique pacto aequa sorte certabunt; sed ludant etiam pari fortuna, sic fiet necessario, ut in sexcentis jactibus centum eveniant senarii: idcirco cum aequa sorte et pari fortuna luditur, in sexcentis jactibus senarij prodibunt centum et pauciores quam centum, quod absurdum.

uscito al primo lancio; poiché allora non si impiegano che cento lanci per vincere cento volte, i restanti 500 lanci, divisi per 4, indicano che sono destinato a perdere 125 volte. Se in verità il 6 si presenta all'ultimo lancio di ciascuna delle quaterne, impiegherò 400 lanci per vincere cento volte ed i restanti duecento indicheranno che sono destinato a perdere cinquanta volte. A motivo di ciò, poiché in qualche caso perderò più spesso di quanto vinca, in altri vincerò più spesso di quanto perda, concludo che è possibile giocare a questa condizione con un'aspettazione equa. Al contrario, se qualcuno volesse l'uscita del 6 una volta ogni tre lanci, costui in alcuni casi vincerebbe tante volte quanto perderebbe, se il sei apparisse come terzo in un gruppo di tre lanci; in altri casi perderebbe molto più di quanto guadagnerebbe, se certamente il sei capitasse sempre come primo esito in un gruppo di tre lanci; in nessun caso però vincerebbe più di quanto perderebbe: da ciò, va da sé che nessuno può giocare a questa condizione senza subirne un danno. ([4], p. 27)

La motivazione qualitativa per preferire di giocare sull'uscita di un sei in 4 anziché in 3 lanci riflette l'idea di aspettazione come risultato intermedio tra ciò che è meglio e ciò che è peggio. Questa *Annotatio* di Bernoulli mostra la necessità di una riflessione sul significato operativo del concetto di aspettazione che va applicato con attenzione quando un esperimento viene condotto tante volte, circostanza non contemplata da Huygens. L'argomento di Bernoulli sembra suggerire anche la necessità di stare attenti nel leggere i dati accorpandoli in modo poco accorto.

Bernoulli risolse i cinque problemi che concludevano il *De Ratiociniis*, in particolare, la soluzione del problema della rovina del giocatore è incompleta.

Nella *Annotatio* alla Prop. IV del *De Ratiociniis* di Huygens Bernoulli aveva introdotto la parola probabilità, senza darvi una connotazione matematica. Vi tornerà nella Sezione IV quando, abbandonata la palestra dei giochi d'azzardo, Bernoulli sentirà il bisogno di fissare i termini su cui fondare le applicazioni della probabilità alle questioni più disparate.

### 5.3 La Sezione IV dell'*Ars Conjectandi*

Bernoulli specifica anzitutto come intendere la *certezza* (*certitudo*) di qualcosa, distinguendo, secondo lo schema aristotelico, tra *certezza oggettiva* e *certezza soggettiva*, in quanto a noi (*in ordine ad nos*). Nel primo caso la certezza di qualcosa indica la verità della sua esistenza presente o l'assicurazione del suo realizzarsi in futuro (*futuritio*). La certezza di un evento passato o presente non è motivo di sorpresa mentre potrebbe rivestirlo il fatto che si parli di certezza riguardo ad eventi che debbono ancora accadere. Ciò si spiega non ricorrendo al fatalismo quanto alla prescienza divina, di fronte alla quale tutto è eternamente presente in un modo che resta misterioso agli uomini, soprattutto quando si tratta di mettere d'accordo la *certitudo futuritionis* con la contingenza o la libertà delle cause seconde, problema di cui Bernoulli non intende trattare.

La certezza soggettiva di qualcosa è, di per sé, variabile con le persone:

Tutte le altre cose ottengono di sé presso le nostre menti una misura [di certezza] più imperfetta, maggiore o minore, a seconda che siano di più o di meno le probabilità che ci persuadono che qualcosa è, sarà o fu.<sup>5</sup> ([4], p. 211)

Ancora una volta la parola *probabilità* appare senza che la si sia definita, ma Bernoulli vi pone subito rimedio:

la *probabilità* è infatti un grado di certezza e ne differisce come la parte dal tutto. Pertanto, se supponiamo che la certezza piena ed assoluta—che indichiamo con la lettera *a* o con l'unità 1—consti, ad esempio, di cinque probabilità come di parti, tre delle quali militano a favore dell'esistenza attuale o futura di un certo evento, le altre contro: si dirà che quell'evento ha  $\frac{3}{5}a$ , ovvero  $\frac{3}{5}$  di certezza.<sup>6</sup> ([4], p. 211)

Dunque, se rapportata a noi, la certezza del realizzarsi di un evento non è monolitica, ma può spezzarsi in parti la cui maggior o minore consistenza dipende da quante di esse possano essere ascritte a favore o meno del realizzarsi dell'evento. In questo passo si è voluto vedere una anticipazione della concezione *soggettiva* del concetto di probabilità ma, se procediamo nella lettura dell'*Ars conjectandi*, vediamo come altre interpretazioni siano possibili. Nel Capitolo II, Bernoulli specifica come stimare la probabilità di un evento:

Le probabilità sono stimate sia dal *numero* che dal *peso degli argomenti* che in qualunque modo dimostrano ed indicano che qualcosa è, sarà o fu.<sup>7</sup> ([4], p. 214)

È qui il caso di richiamare un passo della *De partitione oratoria* di Cicerone in cui l'argomentazione era definita come *probabile inventum ad faciendam fidem*, un passo che Étienne Chauvin citò per definire la voce *argumentum* nel *Lexicon Rationale*, una sorta di dizionario dei termini filosofici, pubblicato nel 1692 [3].

Il peso (*pondus*) di un argomento è la sua capacità di costituire una prova, (*vis probandi*), la robustezza della logica con cui induce a dare il proprio assenso. In questo senso, il grado di certezza che costituisce la probabilità avvicina Bernoulli ad una concezione *logicista* delle probabilità, ripresa a partire dalla metà del XIX secolo da diversi studiosi, soprattutto di estrazione anglosassone. Quanto agli argomenti, Bernoulli distingue tra argomenti *intrinseci* o artificiali ed *estrinseci* o non artificiali, seguendo anche qui categorie filosofiche ben colaudate. Sono argomenti intrinseci quelli desunti da alcuni luoghi comuni (*loci communes*) della logica quali causa, effetto, circostanza. Argomenti estrinseci

<sup>5</sup>Caetera omnia imperfectiorem ejus mensuram in mentibus nostris obtinent, majorem minoremve, prout plures vel pauciores sunt probabilitates, quae suadent rem aliquam esse, fore aut fuisse.

<sup>6</sup>*Probabilitas* enim est gradus certitudinis, et ab hac differt ut pars a toto. Nimirum, si certitudo integra et absoluta, quam litera *a* vel unitate 1 designamus, quinque verb. gr. probabilitatibus ceu partibus constare supponatur, quarum tres militent pro existentia aut futuritione alicujus eventus, reliquae contra: eventus ille dicitur habere  $\frac{3}{5}a$ , seu  $\frac{3}{5}$  certitudinis.

<sup>7</sup>Probabilitates aestimantur ex *numero* simul et *pondere argumentorum*, quae quoquo modo probant et indicant, rem aliquam esse, fore aut fuisse.

sono quelli dovuti alla testimonianza ed alla autorità degli uomini. L'esempio addotto da Bernoulli è rimasto celebre perché tratto dalla giurisprudenza e sottolinea l'influenza di questa disciplina sul nascente calcolo delle probabilità:

Tizio viene trovato senza vita per strada e Mevio viene accusato dell'omicidio; Gli argomenti dell'accusa sono, 1. risulta che egli odiasse Tizio (ecco un argomento *dalla causa*, l'odio ha potuto spingerlo ad uccidere). 2. una volta interrogato [Mevio] impallidì e rispose in modo incerto (ecco un argomento *dall'effetto*; è infatti possibile che il pallore e la paura derivassero dalla consapevolezza di aver commesso il crimine). 3. nella casa di Mevio è stato ritrovato un pugnale sporco di sangue (ecco un *segno*). 4. Nel giorno in cui Tizio venne ucciso, Mevio era passato sulla stessa strada dove avvenne l'omicidio (ecco la *circostanza* di tempo e luogo). 5. Infine, Caio ha depono affermando che, il giorno prima dell'omicidio, Tizio e Mevio avevano litigato (ecco un *testimone*).<sup>8</sup> ([4], p. 214)

Bernoulli distingue tra eventi probabili e possibili

È *possibile* ciò che ottiene una piccola parte di certezza, impossibile ciò che non ne ottiene alcuna o solo una infinitamente piccola. È possibile, ad esempio, ciò che ha  $\frac{1}{20}$  o  $\frac{1}{30}$  di certezza.<sup>9</sup> ([4], p. 211)

Un evento può però essere più probabile (*probabilius*) di un altro se gli compete una porzione maggiore di certezza, anche se in sé l'evento in sé potrebbe essere solo possibile.

A questo punto Bernoulli introduce il celebre concetto di *certezza morale* di un evento, già considerato da Nicola di Oresme:

È *moralmente certo* ciò la cui probabilità è circa eguale alla piena certezza, al punto che non se ne percepisce la differenza: al contrario, è *moralmente impossibile* ciò che possiede una probabilità pari alla distanza di un evento moralmente certo dalla certezza piena. Per esempio, se si ritiene moralmente certo ciò che possiede  $\frac{999}{1000}$  di certezza, sarà moralmente impossibile ciò che ne possiede soltanto  $\frac{1}{1000}$ .<sup>10</sup> ([4], pp. 211-212)

<sup>8</sup>Titius occisus reperitur in via, Maevius commissi homicidii accusatur; Argumenta accusationis sunt, 1. quod constet illum odio habuisse Titium (en argumentum a *causa*, potuit enim odium hoc ipsum impulsisse ad occidendum). 2. quod examinatus palluerit timideque responderit (en argumentum ab *effectu*; potest enim pallor et metus iste ex conscientia patratu criminis profluxisse). 3. quod in aedibus Maevii repertus mucro sanguine tinctus (en *signum*). 4. quod quo die occisus in via Titius, eodem illac transierit Maevius (en *circumstantiam* loci et temporis). 5. quod denique Cajus deponat, pridie commissi homicidii Titio cum Maevio lites intercessisse (en *testimonium*).

<sup>9</sup>Possibile est, quod vel tantillam certitudinis partem obtinet: impossibile, quod nullam aut infinite exiguam. Ita possibile est, quod habet  $\frac{1}{20}$  aut  $\frac{1}{30}$  certitudinis.

<sup>10</sup>Moraliter certum est, cujus probabilitas fere aequatur integrae certitudini, sic un defectus sentiri non possit: *Moraliter impossibile* contra, quod tantum duntaxat probabilitatis habet, quantum moraliter certo ad omnimodam certitudinem deest. Ita si pro moraliter certo habeatur, quod  $\frac{999}{1000}$  certitudinis possidet, erit moraliter impossibile, quod ejus tantum habet  $\frac{1}{1000}$ .

Un'ultima distinzione operata da Bernoulli è tra ciò che è *necessario* e ciò che è *contingente*. Necessario è ciò che

non può non essere, avvenire o essere stato<sup>11</sup> ([4], p. 212)

e Bernoulli distingue tre generi di necessità: fisica, ipotetica e convenzionale. Nel primo genere rientrano affermazioni quali: è necessario che il fuoco bruci, che un triangolo abbia angoli interni la cui somma è pari a due retti, che il plenilunio si verifichi quando la luna passa per i nodi dell'eclittica. Necessità ipotetica è quella in base alla quale

mentre si suppone che una cosa qualunque sia o sia stata o si sa che è o è stata, non è possibile che non sia o non sia stata.<sup>12</sup> ([4], p. 212)

Se, ad esempio so o ipotizzo che Pietro sappia scrivere, è necessario, per coerenza, tenere questa ipotesi per vera in tutte le successive deduzioni. La necessità convenzionale è esemplificata in questo modo:

in forza di un patto il giocatore d'azzardo, che ottiene un sei lanciando un dado, si dice che vincerà necessariamente se prima del lancio si convenne tra i giocatori che la vittoria sarebbe stata raggiunta ottenendo sei con un lancio.<sup>13</sup> ([4], p. 212)

Ciò che è contingente, cioè ciò che può non essere, non essere stato o non essere in futuro, si distingue in quanto *libero*, cioè che deriva la propria contingenza dall'arbitrio di una creatura, *fortuito* ovvero *casuale* se dipende dalla fortuna o dal caso. La distinzione tra necessario e contingente non è qualcosa di rigidamente fissato ed anzi la contingenza *non* esclude una qualche forma di necessità. Bernoulli porta due esempi:

È fuori di ogni dubbio che, assegnata la posizione, la velocità e la distanza dal terreno di un dado nel momento in cui esso abbandona la mano di chi lo lancia, il dado non può cadere altrimenti che nel modo in cui cadrà effettivamente: similmente, data l'attuale costituzione dell'aria, nonché la quantità, la posizione, il moto, la direzione, la velocità e le leggi dei meccanismi con cui tutte queste cose interagiscono, la tempesta di domani non potrà essere altro che quella che sarà effettivamente; al punto che questi effetti discendono dalle loro cause prossime in modo non meno necessario della dipendenza delle eclissi dal movimento dei corpi celesti.<sup>14</sup> ([4], p. 212)

<sup>11</sup>non potest non esse, fore aut fuisse

<sup>12</sup>unumquodque dum est aut fuit, vel esse aut fuisse supponitur, non potest non esse aut fuisse.

<sup>13</sup>quo pacto aleator, qui tessera senarium jecerit, necessario vincere dicitur, si prius inter lusores ita conventum fuerit, ut jactu senarii victoria constet.

<sup>14</sup>Certissimum est, quod data tesserae positione, velocitate et distantia ab alveo, eo momento quo manum projicientis deserit, tessera non potest aliter cadere, quam uti revera cadit: item quod data aëris constitutione praesente, datisque ventorum, vaporum, nubium mole, situ, motu, directione, velocitate et mechanismi legibus, quibus haec omnia in se invicem agunt, tempestas crastinae diei non possit alia fore, quam qualis reapse futura est; adeo ut hi effectus ex suis causis proximis non minus necessario, atque Eclipsium phenomena ex luminarium motu sequantur.

L'accostamento con le eclissi non è casuale perché Bernoulli contrappone un fenomeno le cui regole sono prevedibili matematicamente da quelli in cui questa possibilità è preclusa *non* da ragioni intrinseche al fenomeno quanto dall'inadeguatezza delle conoscenze scientifiche o dei modelli matematici caratteristici del momento in cui si scrive:

Ne segue che, quanto può sembrare contingente a qualcuno qui ed in questo istante, a qualcun altro (in fondo, anche alla stessa persona) appaia necessario, dopo che è venuto a conoscenza delle sue cause; pertanto la contingenza riflette soprattutto la nostra cognizione, in quanto non scorgiamo alcun impedimento affinché qualcosa non sia o non sarà, benché ora ed in questo luogo, in forza di una causa, prossima ma a noi sconosciuta, avviene o avverrà necessariamente.<sup>15</sup> ([4], pp. 212-213)

Se, come vedremo, la probabilità esplora razionalmente quanto è contingente, ne emerge una visione negativa, nel senso che, laddove non è possibile mettere in piena luce l'evoluzione *deterministica* che i fenomeni naturali seguono, non resta che attenersi ai risultati ottenibili tramite il ricorso alla probabilità. Il capitolo si chiude con la definizione di *fortuna prospera* o *fortuna adversa*, come il realizzarsi di un evento a noi propizio o contrario la cui probabilità è inferiore ad  $\frac{1}{2}$ .

La distinzione tra ciò che è certo e ciò che è soggetto al dubbio si riflette sul diverso approccio gnoseologico perché se ciò che è certo si può conoscere (*scire*) e comprendere (*intelligere*), su ciò che è incerto si può solo congetturare (*conjectare*) od opinare (*opinari*). Per Bernoulli

*Congetturare* qualcosa significa misurarne la probabilità: pertanto definiamo *arte della congettura* o *Stocastica* l'arte di misurare nel modo più preciso possibile le probabilità delle cose.<sup>16</sup> ([4], p. 213)

L'arte della congettura è dunque il mezzo che permette di iniziare un discorso scientifico su ciò che non avviene necessariamente ma è solo contingente, superando lo sbarramento che aveva posto Aristotele, per il quale non si dava un discorso scientifico su ciò che non avviene sempre con certezza. L'arte di congetturare è la base razionale che deve servire da guida per potere scegliere sempre la soluzione migliore e più sicura nei giudizi e nelle azioni. Esposti i diversi tipi di argomento, Bernoulli elenca nove regole generali, dettate dal buon senso, cui occorre attenersi nell'esercizio dell'arte della congettura, illustrandole con esempi.

#### 1. Non bisogna mai formulare congetture in tutto ciò che è certo.

<sup>15</sup>Sequitur hinc, uni et uno tempore videri posse contingens, quod alii (imo et idem) alio tempore post cognitās ejus causas sit necessarium; adeo ut contingētia praecipue etiam respiciat cognitionem nostram, in quantum nos nullam videmus repugnantiam in objecto ad non esse vel fore, etiamsi hic et nunc vi causae proximae sed nobis ignotae necessario fit vel fiat.

<sup>16</sup>*Conjectare* rem aliquam est metiri illius probabilitatem: ideoque *Ars Conjectandi* sive *Stochastice* nobis definitur ars metiendi quam fieri potest exactissime probabilitates rerum.

Si tratta della prima occorrenza del termine "stocastico" in un contesto probabilistico.

2. Non è sufficiente soppesare questo o quell'argomento ma bisogna analizzare in dettaglio tutti gli indizi che possano aumentarne o diminuirne il peso.
3. Bisogna considerare sia gli indizi a favore che quelli contrari ad una certa affermazione, per consentire una valutazione equanime; così, per decidere se dichiarare la morte di un amico che manca da tantissimo tempo dalla propria patria, occorre considerare argomenti a favore di questa decisione (molti di coloro che effettuano viaggi frequenti sono esposti a pericoli sia nei viaggi in mare che per le malattie che possono contrarre o le battaglie in cui possono incappare) ed argomenti contrari (l'amico non poté scrivere perché fatto prigioniero; forse ha scritto ma le sue lettere sono andate perdute; vi sono casi di persone che effettivamente hanno fatto ritorno a casa dopo tempo lunghissimo).
4. Giudizi che riguardano affermazioni universali richiedono argomenti generali ma se la congettura è formulata su individui particolari occorre, se possibile, aggiungere altri argomenti di natura speciale; così, per rispondere alla domanda se sia più probabile che viva più a lungo un ventenne che un sessantenne, occorre basarsi solo su questo dato anagrafico. Se però dobbiamo stimare la speranza di vita del ventenne Pietro, che conduce una vita intemperante, e del sessantenne Paolo che conduce una vita tranquilla, la nostra conclusione può essere diversa.
5. Se occorre prendere una decisione senza avere il tempo di soppesare i dubbi, occorre prendere la decisione che appare più sicura, conveniente e probabile, anche se fosse, in casi estremi, un male minore.
6. Occorre preferire ciò che giova in qualche caso e mai nuoce a quanto né giova né nuoce.
7. Il metro con cui giudicare le azioni umane non può essere solo quello del successo; ad esempio, è da ritenere uno stolto chi scommette sul realizzarsi di un triplo 6 con il lancio di tre dadi, anche quando egli riuscisse a vincere.
8. La fiducia che occorre prestare nel realizzarsi di qualcosa deve essere proporzionata alla sua probabilità (*gradus certitudinis*).
9. Poiché è difficile raccogliere una certezza assoluta intorno al realizzarsi di un evento, si può adottare la certezza morale come garanzia del realizzarsi dell'evento medesimo.

Con il capitolo III Bernoulli esamina ancora la natura degli argomenti usati per elevare una congettura, dividendoli in tre tipologie:

1. Argomenti che esistono necessariamente ma indicano in modo contingente (*necessario existunt et contingentiter indicant*)
2. Argomenti che esistono in modo contingente ma indicano necessariamente (*contingenter existunt et necessario indicant*)

3. Argomenti che esistono ed indicano in modo contingente (*contingenter existunt et simul indicant*).

Anche in questo caso Bernoulli vuole chiarire con esempi la classificazione proposta.

Non ricevo da molto tempo lettere da parte di mio fratello; dubito se ciò sia dovuto alla sua pigrizia o alle occupazioni; temo che sia morto.<sup>17</sup> ([4], p. 217)

Gli argomenti da valutare come possibili cause dell'interruzione della corrispondenza sono dunque tre: la pigrizia, la morte e le occupazioni del fratello. Il primo argomento esiste necessariamente perché, conoscendo il fratello, Bernoulli sa che egli è pigro; tuttavia questo argomento indica in modo contingente perché la pigrizia non impedisce di scrivere. Al contrario la morte è un argomento che esiste in modo contingente (il fratello potrebbe ancora essere in vita) ma indica in modo necessario: i morti non possono scrivere. Infine le occupazioni esistono ed indicano in modo contingente perché il fratello potrebbe avere o meno delle occupazioni e queste possono essere tali da impedirgli di scrivere oppure no.

Anche se la concezione di argomento di Bernoulli è piuttosto ampia, la distinzione tra la sua esistenza e quanto esso indica mostra che si deve intendere un argomento come una affermazione in cui vi è una premessa  $A$  ed una conclusione  $C$ : se la premessa non si verifica necessariamente, si ha un'esistenza contingente dell'argomento mentre se la premessa non implica necessariamente la conclusione, l'argomento indica in modo contingente [19]. Occorre dunque valutare se la premessa  $S$  è vera o meno e se l'implicazione  $A \rightarrow C$  sia o meno stringente.

Una distinzione ulteriore è quella tra argomenti *puri* e *misti*. Puri sono quegli argomenti che in qualche circostanza provano qualcosa (*rem probant*) mentre in altri casi non conducono a nulla di positivo; misti sono detti gli argomenti che talvolta provano qualcosa e talvolta il contrario della stessa cosa. Anche qui Bernoulli fornisce un esempio tratto dalla giurisprudenza. In un tumulto viene ucciso di spada un uomo. Un testimone degno di fede afferma che l'omicida indossava un mantello nero. Tra i partecipanti al tumulto quattro indossavano un mantello nero e tra questi Gracco. Il colore del mantello è un argomento misto per la colpevolezza di Gracco perché in un caso ne prova la colpevolezza mentre in tre casi ne prova l'innocenza, vista la indiscutibile autorità del testimone. Interrogato, Gracco impallidisce. Il pallore è un argomento puro perché, se proviene da cattiva coscienza, prova la colpevolezza di Gracco ma non ne dimostra l'innocenza se proviene da un altro motivo dal momento che egli potrebbe benissimo impallidire per un qualsiasi motivo e ciononostante essere l'omicida. Notiamo che solo gli argomenti misti provano qualcosa in un certo numero di casi ed il contrario di quella cosa nei casi rimanenti.

<sup>17</sup>Frater meus diu nihil ad me literarum dedit; dubito, an ejus segnitias aut negotia in culpa sint; vereor etiam ne plane fato concesserit.

Completata la classificazione degli argomenti, Bernoulli è pronto per applicare il calcolo delle probabilità a situazioni complesse in cui vi sono svariate circostanze di cui tenere conto e che generano diversi gradi di certezza (*gradus certitudinis*) nella verità o meno di una certa affermazione. Per raggiungere l'obiettivo, Bernoulli considera anzitutto un argomento che esiste in modo contingente ma indica in modo necessario e denota con  $b$  e  $c$  rispettivamente il numero di casi in cui questo argomento esiste o non esiste in modo contingente e con  $a := b + c$  i casi possibili. L'ipotesi fondamentale è

Suppongo che tutti i casi siano ugualmente possibili, cioè che possano presentarsi con la stessa facilità; occorre però una certa attenzione, ed i casi che possono presentarsi con più facilità vanno contati tante volte quanto maggiore è la loro facilità a presentarsi.<sup>18</sup> ([4], p. 219)

Bernoulli si muove nell'ambito della misurazione della probabilità come rapporto tra casi favorevoli e casi possibili. Per misurare il grado di probabilità con cui un argomento—che esiste in modo contingente ed indica in modo necessario—dimostra la conclusione, ([11], p. 56) Bernoulli osserva che, nei  $b$  casi favorevoli all'esistenza dell'argomento, la cosa da provare sarà indicata in modo necessario, cioè l'argomento conduce alla certezza piena, cui si attribuisce valore unitario. Nei  $c$  nei casi restanti l'argomento non esiste e dunque non può indicare nulla. La *vis probandi* dell'argomento è

$$\frac{b \cdot 1 + c \cdot 0}{a} = \frac{b}{a}.$$

Quando un argomento esiste necessariamente ma indica in modo contingente, Bernoulli chiama  $\beta$  e  $\gamma$  i casi in cui esso indica o meno la validità della conclusione e pone  $\alpha := \beta + \gamma$ . La probabilità dell'evento indicato da questo argomento è

$$\frac{\beta \cdot 1 + \gamma \cdot 0}{\alpha} = \frac{\beta}{\alpha}$$

e, solo se l'argomento è misto,  $\frac{\gamma}{\alpha} = 1 - \frac{\beta}{\alpha}$  indicherà il grado di certezza dell'evento complementare. Se infine l'argomento è contingente sia quanto ad esistenza sia quanto a ciò che indica, la probabilità che si trae da esso sul realizzarsi di un evento è

$$\frac{b \cdot \frac{\beta}{\alpha} + c \cdot 0}{a} = \frac{b\beta}{a\alpha}$$

mentre, se l'argomento è misto,  $\frac{b\gamma}{a\alpha} = 1 - \frac{b\beta}{a\alpha}$  è la probabilità che esso dimostri il contrario. In ogni caso, Bernoulli, indicata con  $p(A)$  la probabilità che la premessa sia vera e con  $p(A \rightarrow C)$  la probabilità che la premessa, una volta verificata, implichi la conclusione, per valutare la probabilità che l'argomento implichi la conclusione si serve della formula

$$p(A \cdot (A \rightarrow C)) = p(A)p(A \rightarrow C).$$

<sup>18</sup>Pono autem, omnes casus aequae possibile esse, seu pari facilitate evenire posse; alias enim moderatio est adhibenda, et pro quovis casu faciliori tot alii casus numerandi sunt, quoties in caeteris facilius evenit.

Bernoulli insegnò anche a comporre tra loro le forze persuasive dei singoli argomenti, a partire dal caso in cui sono coinvolti solo argomenti puri. Al primo argomento puro sarà associata la probabilità<sup>19</sup>  $\frac{b}{a} = \frac{a-c}{a}$ . Si affianchi ora un secondo argomento puro che dimostra la cosa da provare in  $e$  casi e non dimostra nulla in  $f$  casi e sia  $d := e + f$  il numero di casi possibili. Questo secondo argomento è risolutivo in  $e = d - f$  casi ma negli  $f$  casi restanti rimarrà solo la forza persuasiva del primo argomento  $\frac{a-c}{a}$ . Il peso dei due argomenti è allora

$$\frac{(d-f) \cdot 1 + f \frac{(a-c)}{a}}{d} = \frac{ad - cf}{ad} = 1 - \frac{cf}{ad}.$$

Se si aggiunge un terzo argomento che indichi la validità di qualcosa in  $h = g - i$  casi mentre esso non dimostra nulla in  $i$  casi, poiché in quest'ultimo caso la *vis probandi* resta quella  $\frac{ad-cf}{ad}$  dovuta ai primi due argomenti, avremo che la *vis probandi* dei tre argomenti puri è

$$\frac{(g-i) \cdot 1 + i \cdot \frac{ad-cf}{ad}}{g} = \frac{adg - cfi}{adg} = 1 - \frac{cfi}{adg} \quad (5.1)$$

da cui si evince una plausibile regola generale, quando gli argomenti puri sono in numero di  $n$ . Osserviamo che le probabilità che i singoli argomenti dimostrino la conclusione vengono moltiplicate tra loro, assumendoli implicitamente indipendenti uno dall'altro.

Quando tutti gli argomenti sono misti, ferme restando le notazioni del caso precedente, ciascuno di essi può o provare una cosa o il suo contrario. Allora il *rapporto* tra le probabilità che l'argomento dimostri qualcosa o il suo contrario, desunte solo dal primo argomento è  $\frac{b}{c}$ ; dal solo secondo argomento è  $\frac{e}{f}$  e del solo terzo argomento è  $\frac{h}{i}$ : componendo i tre quozienti, Bernoulli conclude che la probabilità che gli argomenti provino la verità di una conclusione sta a quella che essi ne dimostrino la falsità come  $\frac{beh}{cfi}$  e quindi la probabilità che l'affermazione sia dimostrata dall'argomento è  $\frac{beh}{beh+cfi}$ , quella del suo contrario è  $\frac{cfi}{beh+cfi}$ .

Quando, infine, occorre combinare argomenti puri ad altri misti, Bernoulli riparte dall'equazione (5.1) osservando che  $adg - cfi$  casi provano la verità della conclusione mentre i rimanenti  $cfi$  casi non provano nulla e lasciano l'onere della prova agli argomenti misti. Se si considerano due di questi argomenti che, nel loro insieme, forniscono la probabilità  $\frac{qs}{qs+ru}$  che la conclusione sia vera e la frazione  $\frac{ru}{qs+ru}$  che non lo sia, la probabilità che tutti gli argomenti, sia puri che misti, hanno di dimostrare la conclusione si può scrivere nella forma

$$\frac{(adg - cfi) \cdot 1 + cfi \cdot \frac{qs}{qs+ru}}{adg} = 1 - \frac{cfiru}{adg(qs + ru)}$$

<sup>19</sup>La notazione non deve far intendere che Bernoulli limiti la sua analisi agli argomenti che esistono in modo contingente. In qualunque delle tre categorie introdotte in precedenza cada l'argomento in questione, la sua forza persuasiva sarà sempre misurata quantitativamente da una frazione il cui numeratore indica i casi che militano a favore della capacità dell'argomento di dimostrare la conclusione, mentre il denominatore indica tutti i casi possibili.

che fornisce la misura di allontanamento dalla certezza come prodotto tra l'allontanamento dovuto ai soli argomenti puri  $\left(\frac{cfi}{adg}\right)$  e di quello imputabile ai soli argomenti misti  $\left(\frac{ru}{qs+ru}\right)$ .

Osserviamo come le probabilità che gli argomenti dimostrino una conclusione o il suo contrario possono entrambe essere maggiori di  $\frac{1}{2}$ . In questo caso, per soppesare l'importanza delle conclusioni, Bernoulli ritiene si debba calcolare il rapporto delle probabilità a favore o contrarie alla conclusione. Così, se gli argomenti a favore di una certa conclusione hanno probabilità  $\frac{2}{3}$  e quelli contrari alla stessa conclusione hanno probabilità  $\frac{3}{4}$ , i primi hanno un peso minore, essendo  $\frac{8}{9}$  il rapporto tra le due probabilità ottenute. Bernoulli chiude il Capitolo III con due esempi che vogliono mettere in guardia dall'applicazione superficiale dei metodi probabilistici. Il primo esempio riprende quello già visto in precedenza in cui Gracco è sospettato dell'omicidio di Mevio perché testimoni autorevoli affermano che l'omicida indossava un mantello di colore nero, come Gracco ed altri tre individui presenti sulla scena del delitto. Secondo altre testimonianze degne di fede, continua Bernoulli, si può concludere che l'assassino aveva i capelli ricci, come Gracco ed altri due individui, nessuno dei quali però indossava un mantello nero. Un calcolo superficiale porterebbe a concludere che il rapporto tra le probabilità che Gracco sia colpevole a quelle di innocenza per l'indizio del mantello sia  $\frac{1}{3}$  e che per l'indizio del tipo di capelli sia  $\frac{1}{2}$  per cui si potrebbe a prima vista concludere che la probabilità di innocenza stia a quella di colpevolezza nel rapporto  $\frac{1}{6} = \frac{1}{3} \times \frac{1}{2}$ , suggerendo l'innocenza di Gracco, che invece risulta essere inchiodato come colpevole perché è l'unico ad appartenere ai due gruppi di indiziati. Un secondo esempio interessante riguarda la possibilità che un documento ufficiale sia stato antedatato in modo fraudolento da un notaio ([4], pp. 222-223). La veridicità di questa conclusione può essere valutata sotto due punti di vista. Infatti, essendo il notaio di una figura molto rispettata, Bernoulli stima che se ne troverebbe forse 1 su 50 disposto a macchiarsi di una frode per cui la probabilità che il documento sia autentico si può stimare in  $\frac{49}{50}$ . D'altro canto se sappiamo che il notaio che ha redatto l'atto ha una pessima reputazione, Bernoulli ritiene moralmente certo che l'atto sia stato falsificato ed attribuisce a questa conclusione una probabilità di  $\frac{999}{1000}$ . Dunque sia una conclusione  $C$  (il documento è autentico) che la sua opposta  $\neg C$  (il documento è falsificato) hanno probabilità superiore ad  $\frac{1}{2}$ . Il punto è che le due stime sono ottenute a partire da ipotesi *differenti*, che possiamo chiamare [11]  $H_1$  o  $H_2$  e non vi è alcun motivo per cui si debba avere  $P(C|H_1) + P(\neg C|H_2) = 1$ , come dovremmo richiedere se le ipotesi fossero le stesse nei due casi.

L'impresa di applicare il calcolo delle probabilità ai campi annunciati all'inizio della quarta parte dell'*Ars conjectandi* appare sempre più un atto temerario, per l'estrema difficoltà pratica di determinare con precisione l'insieme di tutti i casi possibili. Già all'inizio del Cap. IV questa appare la distinzione capitale tra le applicazioni classiche della probabilità ai giochi d'azzardo e quelle alla nuova classe di problemi. In questa discussione viene formulato il principio di ragione insufficiente che giustifica la scelta della distribuzione uniforme di probabilità nei problemi classici:

I numeri di casi possibili sono noti, ad esempio, nei giochi con i dadi, dal momento che conosciamo il numero di facce di ciascuno di essi e che tutte hanno la stessa propensione a presentarsi, dal momento che non vi è alcuna ragione, vista la somiglianza delle facce e la distribuzione uniforme del peso del dado, perché una faccia debba essere più propensa a presentarsi rispetto ad un'altra, come potrebbe succedere se le facce avessero una forma diversa ovvero se il dado fosse di due materie diverse. Similmente sono noti i numeri di casi favorevoli all'estrazione di una scheda bianca o una nera da un'urna ed è noto che sono tutte egualmente possibili, dal momento che sono determinati e noti i numeri di schede di entrambi i tipi e non si vede alcun motivo per cui debba essere estratta una piuttosto che un'altra.<sup>20</sup> ([4], pp. 223-224)

La difficoltà di determinare chiaramente *a priori* i casi possibili non ferma l'analisi di Bernoulli che decide di procurarsi questa informazione *a posteriori*

cioè dall'osservazione ripetuta di un evento in situazioni simili, sarà possibile vincersela, dal momento che si deve presumere che qualcosa potrà avvenire o meno in tanti casi quanti si è già presentato in una situazione simile.<sup>21</sup> ([4], p. 224)

Questo passo contiene due osservazioni importanti. Anzitutto la necessaria fiducia nella omogeneità tra passato e presente, alla base di ogni procedimento inferenziale di indagine scientifica. Inoltre Bernoulli fa balenare un metodo alternativo, *per experimenta*, per stimare la probabilità che un certo evento si realizzi. Se vogliamo sapere quale sia la probabilità che Tizio sopravviva altri dieci anni, occorre aver considerato un insieme sufficientemente rappresentativo—300, nell'esempio di Bernoulli—di individui che abbiano la stessa età e corporatura: se si sa che 200 di loro sono morti nel giro di dieci anni, allora

si potrà con sufficiente sicurezza concludere che i casi nei quali Tizio pagherà il proprio debito alla natura entro un decennio sono il doppio di quelli in cui egli potrà superare questo limite.<sup>22</sup> ([4], p. 225)

Bernoulli non reclama diritti di priorità nel metodo proposto, già presentato nell'*Ars cogitandi*, meglio nota come la Logica di Port-Royal, che si deve in gran parte ad Antoine Arnauld e Pierre Nicole e che fu pubblicata per la prima volta

<sup>20</sup>Ita ex. gr. noti sunt numeri casuum in tesseris; in singulis enim tot manifeste sunt quot hedrae, iique omnes aequae proclives; cum propter similitudinem hedrarum et conforme tesserae pondus nulla sit ratio, cur una hedrarum pronior esset ad cadendum quam altera, quemadmodum fieret, si hedrae dissimilis forent figurae, aut tessera una in parte ex ponderiore materia constaret quam in altera. Sic itidem noti sunt numeri casuum ad educendam ex urna schedulam albam nigramve, et notum est omnes aequae possibiles esse; quia nimirum determinati notique sunt numeri schedarum utriusque generis, nullaque perspicitur ratio, cur haec vel illa potius exire debeat quam quaelibet alia.

<sup>21</sup>hoc est, ex eventu in similibus exemplis multoties observato eruere licebit; quandoquidem praesumi debet, tot casibus unumquodque post hac contingere et non contingere posse, quoties id antehac in simili rerum statu contigisse et non contigisse fuerit deprehensum.

<sup>22</sup>satis tuto colligere poteris, duplo plures casus esse, quibus et Tizio intra decennium proximum naturae debitum solvendum sit, quam quibus terminum hunc transgredi possit.

nel 1662. Egli sottolinea però la necessità di ricorrere ad un gran numero di esperimenti, se si vuole ridurre la possibilità di errori vistosi. Anche se tutto ciò appare ragionevole, occorre fornirne un risultato che chiarisca se, al crescere del numero di esperimenti, cresce la probabilità di trovare il valore dell'autentico rapporto (*genuinae rationis*) tra il numero di casi in cui un evento si realizza e quelli in cui esso non si realizza, al punto che questa probabilità possa superare qualsiasi grado di certezza morale, cioè avvicinarsi ad 1 quanto si voglia. In altri termini occorre sapere

se in realtà il problema ha un asintoto, per dire così, cioè se esiste un grado di certezza che non è mai possibile superare, per quanto si moltiplichino le osservazioni come, ad esempio, se non può mai superare la metà o i  $\frac{2}{3}$  o i  $\frac{3}{4}$  della certezza di aver individuato il vero rapporto dei casi.<sup>23</sup> ([4], p. 225)

Per essere più specifico, Bernoulli considera due esempi: nel primo vi è un'urna contenente 3000 palle bianche e 2000 palle nere e si chiede ad una persona che non conosce la composizione dell'urna, di compiere un'estrazione con reimpulso. È possibile ripetere l'esperimento un numero di volte tale che sia dieci o cento o mille volte più probabile che il numero di volte in cui si è estratta una palla bianca diviso per il numero di volte in cui si è estratta una palla nera raggiunga il valore  $\frac{3}{2}$  piuttosto che un qualunque altro rapporto? Bernoulli precisa che non si tratta tanto di trovare il valore esatto del rapporto ma delle limitazioni inferiori e superiori per questo valore, che rendano piccolo quanto si voglia l'intervallo entro cui il valore vero del rapporto può trovarsi. Se questo obiettivo è raggiungibile, avremo ottenuto la morale certezza di avere il corretto rapporto, una certezza che, pur non essendo quella assoluta, è sufficiente per dirimere le questioni tipiche della vita civile. Il secondo esempio è introdotto quasi di soppiatto:

se ora al posto dell'urna consideriamo l'atmosfera od il corpo umano, ad esempio, che contengono al loro interno un numero di mutazioni o di malattie quanti sassolini può contenere un'urna, potremo determinare allo stesso modo, grazie ad osservazioni, con quanta maggior facilità questo o quell'evento potrà realizzarsi.<sup>24</sup> ([4], p. 226)

Il passaggio da un esempio all'altro non è esente da difficoltà e Bernoulli riporta alcune obiezioni mosse da diversi studiosi (*viri quidam docti*) nel lungo periodo di gestazione dell'opera. Una prima obiezione è che il rapporto tra numero di casi favorevoli e sfavorevoli nell'esempio dell'urna è determinato, nell'altro esempio è

---

<sup>23</sup>an vero problema, ut sic dicam, suam habet asymptoton, hoc est an detur quidam certitudinis gradus quem nunquam excedere liceat, utcunque multiplicentur observationes, puta, ut nunquam ultra semissem, aut  $\frac{2}{3}$ , aut  $\frac{3}{4}$  certitudinis partes certi fieri possumus, nos veram casuum rationem detexisse.

<sup>24</sup>etenim si loco urnae substituamus aërem, exempli gratia sive corpus humanum, quae fomitum variarum mutationum atque morborum intra se, velut urna calculos, continent, poterimus utique eodem modo per observationes determinare, quanto facilius in istis subjectis hic vel ille eventus accidere possit.

vago ed indeterminato. La risposta di Bernoulli lascia trasparire ancora una fede incrollabile nel determinismo di tutti gli eventi che sono governati dalla mente onnisciente Dio, mentre noi possediamo una conoscenza parziale ed incerta. La seconda obiezione riguarda il fatto che, nel caso dell'urna, il numero di palle nell'urna è finito mentre quello delle malattie è infinito. Se Bernoulli dapprima osserva che sarebbe forse meglio dire che quest'ultimo numero è incredibilmente grande (*stupende vastum*) piuttosto che infinito ma, anche ammettendo che sia davvero infinito, non è ravvisabile una contraddizione dal momento che si possono trovare infiniti che hanno un rapporto ben determinato. Anche se il pensiero sembra correre al concetto di limite, Bernoulli ricorda l'esempio del numero  $\pi$  (*numerus Ludolphi*) che può essere approssimato da limiti razionali in modo preciso quanto si voglia. La terza obiezione riguarda il fatto che il numero delle malattie non resta costante nel tempo ed a questa obiezione, quella meno insidiosa, Bernoulli oppone il fatto che non occorre utilizzare osservazioni effettuate nel presente per trarre delle conclusioni su quanto accadeva nell'antichità.

Ancora una volta la probabilità appare come uno strumento di indagine da usare quando la nostra ignoranza circa le autentiche cause che determinano gli eventi, ci impedisce di scorgere i nessi causali nascosti efficacemente in tutti i fenomeni che cadono sotto i nostri sensi. Osserviamo infine Bernoulli si muove ancora nel caso di probabilità *discrete* in cui il numero di casi può confondersi con l'infinito solo al limite.

Bernoulli è ora pronto a passare a dimostrare il teorema che deve permettere di completare il progetto di applicazione a vasto raggio della probabilità alle diverse discipline economico-sociali, oggi noto come teorema di Bernoulli-De Moivre o legge debole dei grandi numeri. La dimostrazione vera e propria è preceduta da quelle di cinque lemmi, tutti centrati sullo sviluppo di una potenza del binomio.

Lemma 1 Si considerino le successioni di numeri consecutivi

$$0, 1, 2, 3, 4, \dots, r-1, r, r+1, \dots, r+s$$

e

$$0, 1, 2, 3, 4, \dots, n(r-1), \dots, nr, \dots, n(r+1), \dots, n(r+s) :$$

il numero di termini compreso tra  $n(r+1)$  ed  $n(r+s)$  non può superare di  $s-1$  volte quello dei termini compresi tra  $nr$  ed  $n(r-1)$  o tra  $nr$  ed  $n(r+1)$ ; similmente il numero di termini compreso tra 0 ed  $n(r-1)$  non può essere più di  $r-1$  volte quello dei termini compresi tra  $n(r-1)$  ed  $nr$  o tra  $nr$  ed  $n(r+1)$ , come si ottiene con una verifica diretta. Infatti,  $n(r+s) - n(r+1) = n(s-1)$  e  $n(r+1) - nr = nr - n(r-1) = n$  cosicché  $n(s-1) : n = s-1 : 1$  e  $n(r-1) : n = r-1 : 1$ . Il punto saliente del lemma è che, se anche il numero  $n$  diventa arbitrariamente grande, questi rapporti restano *finiti*.

Lemma 2 Il numero degli addendi che compongono lo sviluppo di  $(r+s)^n$  è  $n+1$ . Bernoulli non si sofferma a lungo su questo lemma la cui dimostrazione è semplice perché basta osservare che gli addendi debbono contenere potenze del tipo  $r^k s^{n-k}$ , per  $k = 0, 1, \dots, n$  e dunque sono  $n+1$ .

Lemma 3 Si consideri la potenza  $(r+s)^{n(r+s)} = (r+s)^{nt}$  e sia  $M$  il termine nel suo sviluppo tale che il rapporto tra il numero di termini che lo precedono e quelli che lo seguono è pari ad  $s/r$ , ovvero  $M$  è il termine che contiene  $r^{nr} s^{ns}$ . Allora  $M$  è il termine più grande nello sviluppo. Inoltre, il rapporto tra  $M$  ed il termine che lo precede è minore di quello tra quest'ultimo ed il suo precedente.

Per la dimostrazione, Bernoulli ricorda anzitutto come i coefficienti dello sviluppo della potenza di un binomio siano simmetrici, ovvero che

$$\binom{n}{k} = \binom{n}{n-k}.$$

Dunque il termine  $M$  ammette due scritture equivalenti:

$$M = \begin{cases} \frac{nt(nt-1)(nt-2)\cdots(nr+1)}{1\cdot 2\cdot 3\cdot 4\cdots ns} r^{nr} s^{ns} \\ \frac{nt(nt-1)(nt-2)\cdots(ns+1)}{1\cdot 2\cdot 3\cdot 4\cdots nr} r^{nr} s^{ns} \end{cases} \quad (5.2)$$

mentre i termini che lo precedono e seguono immediatamente nello sviluppo sono, rispettivamente,

$$M_- = \frac{nt(nt-1)(nt-2)\cdots(nr+2)}{1\cdot 2\cdot 3\cdot 4\cdots (ns-1)} r^{nr+1} s^{ns-1}$$

e

$$M_+ = \frac{nt(nt-1)(nt-2)\cdots(ns+2)}{1\cdot 2\cdot 3\cdot 4\cdots (nr-1)} r^{nr-1} s^{ns+1}.$$

Il termine che precede  $M_-$  è

$$M_{--} = \frac{nt(nt-1)(nt-2)\cdots(nr+3)}{1\cdot 2\cdot 3\cdot 4\cdots (ns-2)} r^{nr+2} s^{ns-2}$$

mentre quello che segue  $M_+$  è

$$M_{++} = \frac{nt(nt-1)(nt-2)\cdots(ns+3)}{1\cdot 2\cdot 3\cdot 4\cdots (nr-2)} r^{nr-2} s^{ns+2}.$$

Si ottiene allora

$$\frac{M}{M_-} = \frac{(nr+1)s}{nsr}, \quad \frac{M_-}{M_{--}} = \frac{(nr+2)s}{(ns-1)r}, \quad \frac{M}{M_+} = \frac{(ns+1)r}{nrs}, \quad \frac{M_+}{M_{++}} = \frac{(ns+2)r}{(nr-1)s}$$

da cui seguono immediatamente le disuguaglianze

$$\frac{M}{M_-} < \frac{M_-}{M_{--}} \quad \text{e} \quad \frac{M}{M_+} < \frac{M_+}{M_{++}} \quad (5.3)$$

che si possono proseguire considerando altri rapporti tra termini consecutivi.

Lemma 4 Fermo restando il significato di  $M$ , si considerino nello sviluppo di  $(r+s)^{n(r+s)} = (r+s)^{nt}$  i due termini  $L$  e  $\Lambda$  che sono  $n$  posizioni a sinistra e a

destra, rispettivamente, di  $M$  nello sviluppo. A patto di scegliere  $n$  sufficientemente grande si può fare assumere sia a  $\frac{M}{L}$  che ad  $\frac{M}{\Lambda}$  un valore arbitrariamente grande.

I termini considerati sono infatti

$$L = \frac{nt(nt-1)(nt-2)\cdots(nr+n+1)}{1\cdot 2\cdot 3\cdot 4\cdots(ns-n)} r^{nr+n} s^{ns-n}$$

e

$$\Lambda = \frac{nt(nt-1)(nt-2)\cdots(ns+n+1)}{1\cdot 2\cdot 3\cdot 4\cdots(nr-n)} r^{nr-n} s^{ns+n}$$

da cui si ottengono, dopo semplificazioni dirette, i rapporti

$$\frac{M}{L} = \frac{(nr+n)(nr+n-1)\cdots(nr+1)s^n}{(ns-n+1)(ns-n+2)\cdots(ns)r^n}$$

$$\frac{M}{\Lambda} = \frac{(ns+n)(ns+n-1)\cdots(ns+1)r^n}{(nr-n+1)(nr-n+2)\cdots(nr)s^n}$$

o, riassorbendo le potenze in ciascuno dei fattori che le precedono,

$$\frac{M}{L} = \frac{(nrs+ns)(nrs+ns-s)\cdots(nrs+s)}{(nsr-nr+r)(nsr-nr+2r)\cdots(nsr)}$$

$$\frac{M}{\Lambda} = \frac{(nsr+nr)(nsr+nr-r)\cdots(nsr+r)}{(nrs-nr+s)(nrs-nr+2s)\cdots(nrs)}.$$

Se si raccoglie  $n$  da ciascun fattore a numeratore e a denominatore e si fa tendere  $n \rightarrow \infty$ , si ottiene al limite

$$\frac{M}{L} = \frac{(rs+s)(rs+s)\cdots(rs)}{(sr-r)(sr-r)\cdots(sr)} \quad \frac{M}{\Lambda} = \frac{(sr+r)(sr+r)\cdots(sr)}{(rs-s)(rs-s)\cdots(rs)}$$

e dunque

$$\frac{M}{L} \simeq \left( \frac{s(r+1)}{r(s-1)} \right)^n \quad \frac{M}{\Lambda} \simeq \left( \frac{r(s+1)}{s(r-1)} \right)^n,$$

che mostrano quanto asserito. Pur avendo raggiunto l'obiettivo, Bernoulli volle fornire una ulteriore dimostrazione del Lemma 5, per convincere anche chi non era assuefatto all'uso dell'infinito. Per questo egli dimostrò che è possibile scegliere un numero finito  $n$  tale che la somma di tutti i termini compresi tra  $L$  e  $\Lambda$  abbia un rapporto, con la somma dei termini restanti, maggiore di una quantità grande a piacere  $c$ . Per far questo egli considerò la frazione  $\frac{r+1}{r}$  e cercò il numero  $m$  tale che

$$\left( \frac{r+1}{r} \right)^m \geq c(s-1) : \quad (5.4)$$

passando ai logaritmi

$$m \geq \frac{\log c(s-1)}{\log(r+1) - \log r}. \quad (5.5)$$

Preso un valore intero di  $m$  che verifichi (5.5), si considerino le frazioni

$$\frac{nrs+ns}{nsr-nr+r}, \quad \frac{nrs+ns-s}{nsr-nr+2r}, \quad \frac{nrs+ns-2s}{nsr-nr+3r}, \dots, \frac{nrs+s}{nsr}$$

che formano, moltiplicate fra loro, il rapporto  $\frac{M}{L}$ . Ora, se supponiamo  $n > m$ , nel rapporto  $\frac{M}{L}$  compare la frazione

$$\frac{nrs + ns - ms + s}{nsr - nr + mr}.$$

Bernoulli intende scegliere  $n$  in modo che essa coincida con  $\frac{r+1}{r}$ . Anche qui, si verifica direttamente che

$$n = m + \frac{s(m-1)}{r+1} \quad \text{ovvero} \quad nt = mt + \frac{st(m-1)}{r+1} :$$

questo valore di  $nt$  è quello da scegliere per far in modo che nello sviluppo di  $(r+s)^{nt}$  sia  $M \geq Lc(s-1)$ . Infatti, la scelta di  $m$  effettuata in (5.5) garantisce la validità della disuguaglianza (5.4). Ora, siccome  $\frac{M}{L}$  è dato dal prodotto

$$\frac{nrs + ns}{nsr - nr + r} \frac{nrs + ns - s}{nsr - nr + 2r} \frac{nrs + ns - 2s}{nsr - nr + 3r} \dots \frac{nrs + ns - ms + s}{nsr - nr + mr} \dots \frac{nrs + s}{nsr}$$

e tutte le  $m$  frazioni  $\frac{nrs+ns-ks+s}{nsr-nr+kr}$  con  $k \leq m$  non sono inferiori a  $\frac{nrs+ns-ms+s}{nsr-nr+mr} = \frac{r+1}{r}$  mentre quelle con  $k > m$  sono comunque maggiori di 1, si conclude che  $\frac{M}{L} > \left(\frac{r+1}{r}\right)^m > c(s-1)$ . Una simile proposizione viene ottenuta per  $\frac{M}{\Lambda}$  e Bernoulli conclude che, elevato  $(r+s)$  ad un esponente

$$nt = \max \left\{ mt + \frac{st(m-1)}{r+1}, mt + \frac{rt(m-1)}{s+1} \right\}$$

la somma dei termini compresi tra  $L$  e  $\Lambda$  nello sviluppo di  $(r+s)^{nt}$  supererà di  $c$  volte la somma dei termini restanti dello sviluppo, sia inferiori ad  $L$  che superiori a  $\Lambda$ .

Lemma 5 Fermi restando i significati di  $M$ ,  $L$  e  $\Lambda$ , è sempre possibile trovare un numero  $n$  tale che il rapporto tra la somma di tutti i termini compresi tra  $L$  e  $\Lambda$  e la somma dei termini restanti nello sviluppo sia maggiore di ogni quantità assegnata.

Bernoulli chiama  $F, G, H, \dots$  i termini compresi tra  $M$  ed  $L$ , con  $F$  il precedente di  $M$  e  $P, Q, R$ , i termini che precedono  $L$ , essendo  $P$  quello immediatamente precedente. Per quanto dimostrato nel Lemma 3 si ha

$$R < Q < P < L < \dots < H < G < F < M$$

e

$$\frac{M}{F} < \frac{L}{P}, \quad \frac{F}{G} < \frac{P}{Q}, \quad \frac{G}{H} < \frac{Q}{R}$$

e dunque

$$\frac{M}{L} < \frac{F}{P} < \frac{G}{Q} < \frac{H}{R} < \dots \quad (5.6)$$

per cui, essendo in virtù del Lemma 4  $\frac{M}{L} \geq c(s-1)$ , *a fortiori* questa disuguaglianza è valida per gli altri rapporti. Resta il fatto che il numero di termini

tra  $L$  ed  $M$  e quelli inferiori ad  $L$  non sono uguali ma il secondo è  $r - 1$  volte il primo, per il Lemma 1. Poiché Bernoulli è interessato al valore del limite

$$\lim_{n \rightarrow \infty} \frac{F + G + H + \dots}{P + Q + R + \dots}$$

possiamo raggruppare i termini che compongono il denominatore in  $(r - 1)$  gruppi, ciascuno formato da tanti termini quanti figurano al numeratore. Inoltre, poiché i termini inferiori ad  $L$  *diminuiscono* sempre, man mano che ci si allontana da  $L$ , possiamo concludere che

$$\frac{F + G + H + \dots}{P + Q + R + \dots} > \frac{F + G + H + \dots}{(r - 1)(P + Q + R + \dots)}$$

dove la parentesi  $(P + Q + R + \dots)$  contiene ora tanti termini quanto ne figurano al numeratore. Poiché le disuguaglianze (5.6) permettono di scrivere

$$F > c(s - 1)P \quad G > c(s - 1)Q \quad H > c(s - 1)R \dots$$

avremo

$$\frac{F + G + H + \dots}{P + Q + R + \dots} > \frac{F + G + H + \dots}{(r - 1)(P + Q + R + \dots)} > \frac{c(s - 1)}{r - 1}$$

e pertanto il Lemma 5 è dimostrato per i termini a sinistra di  $L$ : similmente si opera a destra di  $\Lambda$ .

Terminata questa sezione tecnica, Bernoulli si concentra sull'obiettivo dichiarato all'inizio che in sostanza comporta la traduzione probabilistica dei lemmi precedenti,  $r$  rappresentando il numero di casi favorevoli (*faecundi* o *fertiles*) al realizzarsi di un evento ed  $s$  quello dei casi sfavorevoli (*steriles*) al realizzarsi dello stesso evento:

Sia dunque il rapporto tra il numero di casi favorevoli e quello dei casi sfavorevoli o esattamente o circa uguale al rapporto  $\frac{r}{s}$ , cosicché il rapporto con il numero totale di casi sia  $\frac{r}{(r+s)}$  o  $\frac{r}{t}$ , rapporto che determina i limiti  $\frac{r+1}{t}$  et  $\frac{r-1}{t}$ . Occorre dimostrare che è possibile prendere un numero sufficientemente grande di esperimenti affinché risulti più verosimile di un assegnato numero di volte (diciamo  $c$ ) che il numero di osservazioni con esito favorevole cada all'interno di questi limiti, piuttosto che fuori; cioè che le osservazioni con esito favorevole stiano al numero complessivo di osservazioni in un rapporto che non è né maggiore di  $\frac{r+1}{t}$  né minore di  $\frac{r-1}{t}$ .<sup>25</sup> ([4], p. 236)

<sup>25</sup>Sit igitur numerus casuum fertilium ad numerum sterilium vel praecise vel proxime in ratione  $\frac{r}{s}$ , adeoque ad numerum omnium in ratione  $\frac{r}{(r+s)}$  seu  $\frac{r}{t}$ , quam rationem terminent limites  $\frac{r+1}{t}$  et  $\frac{r-1}{t}$ . Ostendendum est, tot posse capi experimenta, ut datis quodlibet (puta  $c$ ) vicibus verisimilius evadat, numerum fertilium observationum intra hos limites quam extra casurum esse, hoc est, numerum fertilium ad numerum omnium observationum rationem habiturum nec majorem quam  $\frac{r+1}{t}$  nec minorem quam  $\frac{r-1}{t}$ .

Osserviamo esplicitamente che in ogni osservazione il numero di casi favorevoli e contrari all'evento che interessa, restano *costanti*. Per la dimostrazione, Bernoulli indica con  $nt$  il numero di osservazioni e nota che la probabilità che non vi siano casi contrari è  $\left(\frac{r}{t}\right)^{nt}$ , quella che ve ne sia uno solo è  $\frac{nt}{1} \frac{r^{nt-1}s}{t^{nt}}$  e così via. In questo modo il termine massimo  $M$  rappresenta la situazione in cui vi sono  $nr$  casi favorevoli ed  $ns$  contrari, mentre i coefficienti  $L$  e  $\Lambda$  contengono un numero di casi favorevoli pari ad  $nr - n$  e ad  $nr + n$ , rispettivamente: in questo modo il numero di casi in cui gli eventi favorevoli sono compresi tra  $n(r - 1)$  ed  $n(r + 1)$  è dato dalla somma di tutti i termini compresi tra  $L$  e  $\Lambda$ , mentre i termini restanti rappresentano nel loro insieme tutti i casi in cui gli eventi favorevoli sono in numero maggiore o minore. Grazie ai Lemmi 4 e 5, è possibile scegliere il numero  $n$  di osservazioni in modo che la somma dei termini compresi tra  $L$  e  $\Lambda$  sia almeno  $c$  volte più grande della somma dei termini restanti. In altre parole, si può scegliere il numero  $n$  di osservazioni in modo che sia  $c$  volte più probabile che il rapporto tra numero di successi e numero di prove sia compreso nell'intervallo  $\left[\frac{(r-1)}{t}, \frac{(r+1)}{t}\right]$  piuttosto che non vi appartenga.

Per illustrare il risultato, Bernoulli ritornò all'esempio dell'urna contenente solo palle bianche o nere in proporzione  $\frac{3}{2}$ . Prendendo  $r = 30$  ed  $s = 20$ , si vuole ottenere  $c = 1000$  e avere così una probabilità mille volte più grande che la frazione di successi, cioè di estrazioni di una pallina bianca, si trovi nell'intervallo  $\left[\frac{29}{50}, \frac{31}{50}\right]$  piuttosto che fuori da quest'intervallo. Il numero di esperimenti da effettuare per raggiungere questo grado di certezza morale è pari a 25550. Dopo questo esempio l'*Ars Conjectandi* si chiude bruscamente, forse [22] per il disappunto di Bernoulli circa l'entità di questo numero, un numero "astronomico" di prove, ben superiore al numero delle stelle classificate nel catalogo di Flamsteed del 1725, che ne annovera 3000, e superiore alla stessa popolazione di Basilea, la città dove Bernoulli viveva. Inoltre, come abbiamo già visto in precedenza, esperti giocatori d'azzardo erano in grado di elevare congetture sulle probabilità degli esiti nel lancio dei dadi con un numero inferiore di prove. Dunque è possibile che Bernoulli si sia reso conto che qualcosa nelle sue approssimazioni fosse sfuggito, portando a stime eccessivamente grossolane. Se, in questo senso, il risultato di Bernoulli era deludente, esso servì di stimolo a generazioni di matematici a venire, partendo da Abraham de Moivre.

## 5.4 Reazioni al lavoro di Bernoulli

Un primo cambiamento di prospettiva al lavoro di Jakob Bernoulli fu apportato dal nipote Nicolaus [10] che, anziché imporre il grado di morale certezza  $c$  e determinare il numero di prove necessarie per ottenerlo, fissò il numero di prove cercando quale fosse il corrispondente grado di certezza. Una indagine più approfondita e sistematica dei risultati di Jakob Bernoulli fu offerta da Abraham De Moivre in più riprese, nella *Miscellanea Analytica* pubblicata nel 1730 e nella seconda e terza edizione del *The Doctrine of Chances* pubblicate nel 1738 e 1756, rispettivamente, mentre nella prima edizione del 1718, pur citando con

ammirazione i risultati di Jakob Bernoulli, affermava di non essere in grado di migliorarli. In realtà, fin dal 1721 egli si occupò di un problema propostogli da Alexander Cumings, attinente allo schema di Bernoulli:

Due giocatori  $A$  e  $B$ , le cui abilità stanno tra loro nel rapporto  $a$  su  $b$ , si vincolano tramite questo accordo stipulato con un certo spettatore: al termine di un numero  $n$  di partite che è un multiplo di  $a + b$ ,  $A$  darà allo spettatore tante monete quanto è il numero di partite che egli avrà vinto in eccesso rispetto  $\frac{a}{a+b}n$ : similmente  $B$  gli darà tante monete quanto è il numero di partite che egli avrà vinto in eccesso rispetto  $\frac{b}{a+b}n$ . Si domanda l'aspettazione dello spettatore.<sup>26</sup> ([7], p. 99)

La soluzione del problema, quando  $a = b (= 1)$  coinvolge il rapporto tra il coefficiente massimo  $M$  nello sviluppo di  $(1+1)^n$  e la somma di tutti i coefficienti, cioè  $2^n$ . Nella *Miscellanea Analytica* (Libro V, Problema III) De Moivre presentò come valore approssimato di questo rapporto, quando  $n \gg 1$ ,

$$\frac{M}{2^n} \simeq \frac{2 \frac{21}{125} \left(1 - \frac{1}{n}\right)^n}{\sqrt{n-1}},$$

dove la scrittura  $2 \frac{21}{125}$  è da intendersi come  $2 + \frac{21}{125} = \frac{271}{125}$ . Nel Problema IV De Moivre considerò, quando  $n = 2m \gg 1$ , il rapporto tra  $M$  ed il termine  $Q$  che dista  $p$  posizioni da  $M$  nello sviluppo, trovando

$$\frac{M}{Q} \simeq \frac{(m-p-1)^{m+p-\frac{1}{2}} (m-p+1)^{m-p+\frac{1}{2}} \binom{m+p}{m}}{m^{2m}}.$$

Egli propose un esempio numerico del calcolo di  $\frac{M}{2^n}$  nel caso  $n = 900$  e di  $\frac{M}{Q}$  nel caso di  $p = 30$ . In entrambi i casi egli fece uso dei logaritmi ed in particolare di una tavola contenente  $\log_{10} n!$ .

Nel Capitolo III della *Miscellanea*, De Moivre considerò potenze del tipo  $(a+b)^n$  con  $a \neq b$  e dimostrò, in modo diverso da Bernoulli, quale fosse la posizione  $\ell$  del termine massimo nello sviluppo del binomio. Detto  $M$  il termine che occupa la posizione  $\ell$ -esima nello sviluppo, i termini precedente e successivo sono, rispettivamente,

$$\frac{\ell-1}{n-\ell+2} M \frac{a}{b} \quad \text{e} \quad \frac{n-\ell+1}{\ell} M \frac{b}{a}$$

per cui, imponendo che  $M$  sia maggiore di entrambe queste quantità si trova che  $\ell$  deve essere compreso tra  $\frac{(n+1)b}{a+b}$  e  $\frac{(n+1)b}{a+b} + 1$  e dunque risulta pienamente determinato se questi estremi non sono interi; in caso contrario, vi sono due termini che corrispondono al valore massimo. I dettagli dei metodi seguiti da

<sup>26</sup>Collusores duo  $A$  et  $B$  quorum dexteritates sint in ratione data  $a$  ad  $b$ , spectatores cuidam ita se obstringant, ut post conclusum quemlibet ludorum numerum  $n$  quem  $a + b$  metitur,  $A$  sit ei tot nummos largiturus quot ipse plures ludos vicerit quam qui designetur per  $\frac{a}{a+b}n$ : similiter  $B$  sit ei largiturus tot nummos quot ipse plures ludos vicerit quam qui designentur per  $\frac{b}{a+b}n$ , quaeritur sors spectatoris.

De Moivre per stimare il coefficiente massimo  $M$  nello sviluppo di un binomio  $(a+b)^{2m}$ , con  $a=b$ , cioè

$$M = \binom{2m}{m}$$

sono precisati nel Capitolo II del Libro VI della *Miscellanea* dove si comprende come egli abbia utilizzato una distribuzione continua, oggi nota come distribuzione normale, per approssimare la distribuzione binomiale *discreta*. De Moivre ne riordinò i termini

$$M := \binom{2m}{m} = \frac{m+1}{m-1} \times \frac{m+2}{m-2} \times \frac{m+3}{m-3} \times \cdots \times \frac{m+(m-1)}{m-(m-1)} \times \frac{2m}{m} \quad (5.7)$$

e passò al logaritmo naturale per servirsi della serie di Newton-Gregory

$$\ln \frac{1+x}{1-x} = 2 \sum_{k=1}^{\infty} \frac{x^{2k-1}}{2k-1}, \quad (5.8)$$

convergente per  $|x| < 1$ : in questo passaggio l'ultimo quoziente nell'espressione di  $\ln M$  contribuirà con  $\ln 2$ . Poiché De Moivre è interessato al limite  $m \gg 1$  egli applicò ripetutamente la (5.8), ottenendo

$$\ln \frac{1 + \frac{1}{m}}{1 - \frac{1}{m}} = 2 \left( \frac{1}{m} + \frac{1}{3m^3} + \frac{1}{5m^5} + \frac{1}{7m^7} + \cdots \right)$$

$$\ln \frac{1 + \frac{2}{m}}{1 - \frac{2}{m}} = 2 \left( \frac{2}{m} + \frac{8}{3m^3} + \frac{32}{5m^5} + \frac{128}{7m^7} + \cdots \right)$$

$$\ln \frac{1 + \frac{3}{m}}{1 - \frac{3}{m}} = 2 \left( \frac{3}{m} + \frac{27}{3m^3} + \frac{243}{5m^5} + \frac{2187}{7m^7} + \cdots \right)$$

e così via, fino ad esaurire le frazioni in cui è stato scomposto  $\ln E$ . De Moivre procedette sommando *in colonna* (*in columnas perpendiculares*) ottenendo

$$\frac{2}{m} [1 + 2 + 3 + \cdots + m - 1] = \frac{2}{m} S_1,$$

$$\frac{2}{3m^3} [1 + 2^3 + 3^3 + \cdots + (m-1)^3] = \frac{2}{3m^3} S_3,$$

$$\frac{2}{5m^5} [1 + 2^5 + 3^5 + \cdots + (m-1)^5] = \frac{2}{5m^5} S_5$$

e così via, dove  $S_k$  è la somma delle potenze  $k$ -esime dei primi  $m-1$  interi consecutivi a partire da 1, ottenute da Jakob Bernoulli nell'*Ars Conjectandi*, senza indicazione sul modo di dimostrarle.<sup>27</sup> De Moivre espresse le somme

<sup>27</sup>Si veda l'appendice 2 di questo capitolo per una dimostrazione ottenuta con il ricorso al calcolo simbolico.

trovate in termini di  $s = m - 1$  come

$$\begin{aligned} \frac{2}{m} S_1 &= \frac{s(s+1)}{m} = \frac{s^2}{m} + \frac{s}{m} \\ \frac{2}{3m^3} S_3 &= \frac{\frac{1}{2}s^4 + s^3 + \frac{1}{2}s^2}{3m^3} = \frac{s^4}{6m^3} + \frac{s^3}{3m^3} + \frac{s^2}{6m^3} \\ \frac{2}{5m^5} S_5 &= \frac{\frac{1}{3}s^6 + s^5 + \frac{5}{6}s^4 - \frac{1}{6}s^2}{5m^5} = \frac{s^6}{15m^5} + \frac{s^5}{5m^5} + \frac{s^4}{6m^5} - \frac{s^2}{30m^5} \end{aligned} \quad (5.9)$$

.....

in generale,  $S_k$  è un polinomio in  $s$  di grado  $k + 1$ . Posto ancora  $\frac{s}{m} = x$ , De Moivre ripeté con (5.9) l'artificio di sommare i termini colonna per colonna e ottenne, per la prima colonna, la somma  $\Sigma_1$  data da

$$\begin{aligned} \Sigma_1 &= s \left[ x + \frac{1}{2 \times 3} x^3 + \frac{1}{3 \times 5} x^5 + \frac{1}{4 \times 7} x^7 + \dots \right] = \\ &= s \left[ \frac{2}{1 \times 2} x + \frac{2}{3 \times 4} x^3 + \frac{2}{5 \times 6} x^5 + \frac{2}{7 \times 8} x^7 + \dots \right] = \\ &= m \left[ \frac{2}{1 \times 2} x^2 + \frac{2}{3 \times 4} x^4 + \frac{2}{5 \times 6} x^6 + \frac{2}{7 \times 8} x^8 + \dots \right] = mx \ln \frac{1+x}{1-x} - m \ln \frac{1}{1-x^2}, \end{aligned}$$

dove De Moivre osservò che la serie nell'ultima equazione coincide con quella che si ottiene integrando termine a termine la serie (5.8)<sup>28</sup> Poiché  $x = \frac{s}{m} = \frac{m-1}{m}$ ,

$$\Sigma_1 = (2m - 1) \ln(2m - 1) - 2m \ln m. \quad (5.10)$$

De Moivre sommò i termini della seconda colonna a destra delle equazioni (5.9)

$$\frac{s}{m} + \frac{s^3}{3m^3} + \frac{s^5}{5m^5} + \frac{s^7}{7m^7} + \dots = \frac{1}{2} \ln \frac{m+s}{m-s} = \frac{1}{2} \ln(2m-1), \quad (5.11)$$

dove è stato ancora utilizzato lo sviluppo di Newton-Gregory ed il significato di  $s$  in termini di  $m$ . Per ottenere una prima approssimazione di  $\ln \frac{M}{2^{2m}}$ —logaritmo della probabilità di ottenere  $m$  successi in  $n = 2m$  prove indipendenti di un evento avente probabilità di successo  $\frac{1}{2}$  in ciascuna prova—occorre sommare (5.10) e (5.11) e sottrarre  $2m \ln 2$ , che tiene conto del denominatore  $2^{2m}$ , e aggiungere il termine  $\ln 2$  derivante dal rapporto  $\frac{2m}{m}$  che figurava in  $M$ , ricavando

$$\begin{aligned} \ln \frac{M}{2^{2m}} &\simeq \left(2m - \frac{1}{2}\right) \ln(2m - 1) - 2m \ln 2m + \ln 2 = \\ &= \left(n - \frac{1}{2}\right) \ln(n - 1) - n \ln n + \ln 2. \end{aligned}$$

Occorre a questo punto sommare le altre colonne e De Moivre trova il valore della terza colonna, che costituisce una progressione geometrica, oltre ad un valore approssimato della quarta colonna. Anziché proseguire, però, egli stimò queste somme nel limite  $m \gg 1$ , ottenendo

$$\ln \frac{M}{2^{2m}} = \log \ln \frac{M}{2^n} \simeq \left(n - \frac{1}{2}\right) \ln(n - 1) - n \ln n + \ln 2 + \ln A \quad (5.12)$$

<sup>28</sup>Si integri il membro di sinistra della (5.8) per parti, prendendo  $1dx$  come fattore finito.

dove

$$\ln A := \frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680} + \dots$$

che rappresenta una serie numerica a segni alterni, ciò che De Moivre non dimostrò. È possibile allora ottenere il termine da aggiungere per migliorare l'approssimazione di  $\frac{M}{2^n}$ . Osservando ora che, nel limite in cui  $n \gg 1$ ,

$$\frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n \simeq \frac{1}{e}$$

si può anche concludere che

$$\ln \frac{M}{2^n} \simeq \ln A - \ln \sqrt{n-1} + \ln 2 - 1 \quad (5.13)$$

e, introdotto  $\ln B = -1 + \ln A$ , si ha

$$\frac{M}{2^n} \simeq \frac{2B}{\sqrt{n-1}} \simeq \frac{2B}{\sqrt{n}}.$$

Nel 1732, il matematico scozzese James Stirling (1692-1770) dimostrerà che  $B = \frac{1}{\sqrt{2\pi}}$  cosicché

$$\frac{M}{2^n} \simeq \frac{2}{\sqrt{2\pi n}}. \quad (5.14)$$

Osserviamo che con l'uso dello sviluppo di Stirling, si stanno utilizzando numeri irrazionali per approssimare numeri razionali.<sup>29</sup> Sempre servendosi dello sviluppo di Newton-Gregory, De Moivre ottenne il logaritmo del rapporto tra il termine  $L$  che dista  $\ell$  posti da quello centrale  $M$  nello sviluppo ad una potenza pari  $n = 2m \gg 1$  del binomio  $(1+1)^{2m}$ , giungendo a

$$\ln \frac{L}{M} = \begin{cases} (m + \ell - \frac{1}{2}) \ln(m + \ell - 1) + (m - \ell + \frac{1}{2}) \ln(m - \ell + 1) - \\ -2m \ln m + \ln \frac{m+\ell}{m} \end{cases}$$

che, nel limite in cui  $m \gg 1$ , si riduce al termine dominante  $-\frac{\ell^2}{m} = -\frac{2\ell^2}{n}$  per cui, ripassando dai logaritmi ai numeri,

$$\frac{L}{M} = e^{-\frac{2\ell^2}{n}} = 1 - \frac{2\ell^2}{n} + \frac{4\ell^4}{2n^2} - \frac{8\ell^6}{6n^3} + \frac{16\ell^8}{24n^4} - \dots \quad (5.15)$$

Ricordiamo che questo rapporto esprime il rapporto tra due probabilità: quella— $P\left(\frac{n}{2} \pm \ell\right)$ —che un evento di probabilità  $\frac{1}{2}$  si presenti in un numero  $n$  pari di esperimenti ripetuti,  $\frac{n}{2} - \ell$  o  $\frac{n}{2} + \ell$  volte e quella— $P\left(\frac{n}{2}\right)$ —che lo stesso evento si presenti  $\frac{n}{2}$  volte. La probabilità che l'evento si presenti un numero qualsiasi di volte compreso tra  $\frac{n}{2}$  e  $\frac{n}{2} + \ell$  è

$$\sum_{k=0}^{\ell} P\left(\frac{n}{2} + k\right).$$

<sup>29</sup>In appendice 1 viene proposta una dimostrazione della formula di Stirling di natura elementare

Per stimare questa somma, De Moivre *integrò* termine a termine lo sviluppo in serie (5.15) il risultato ottenuto in precedenza e calcolando termine a termine l'integrale  $\int_0^\ell e^{-\frac{2k^2}{n}} dk$ , egli ottenne, servendosi anche dell'equazione (5.14), la probabilità che un evento—avente probabilità costante pari ad  $\frac{1}{2}$  di verificarsi in ciascuna prova—si ripeta, in  $n$  prove indipendenti, un numero di volte compreso tra  $\frac{n}{2}$  e  $\frac{n}{2} + \ell$ , come

$$\sum_{k=0}^{\ell} P\left(\frac{n}{2} + k\right) = P\left(\frac{n}{2}\right) \sum_{k=0}^{\ell} \frac{P\left(\frac{n}{2} + k\right)}{P\left(\frac{n}{2}\right)} \simeq \frac{2}{\sqrt{2\pi n}} \left[ \ell - \frac{2\ell^3}{3n} + \frac{4\ell^5}{2 \times 5n^2} - \frac{8\ell^7}{6 \times 7n^3} + \dots \right]$$

o, posto  $\ell = s\sqrt{n}$ ,

$$\frac{2}{\sqrt{2\pi}} \left[ s - \frac{2s^3}{3} + \frac{4s^5}{2 \times 5} - \frac{2s^7}{6 \times 7} + \dots \right].$$

In particolare, De Moivre era interessato al caso  $s = \frac{1}{2}$  che fornisce per la probabilità il valore di 0.341344.... Grazie alla simmetria della distribuzione binomiale poté concludere che

E quindi, se fosse possibile prendere un numero infinito di esperimenti, la probabilità che un evento, il quale ha un ugual numero di possibilità di realizzarsi come di non realizzarsi, non si verifichi né più frequentemente di  $\frac{1}{2}n + \frac{1}{2}\sqrt{n}$  volte, né più raramente di  $\frac{1}{2}n - \frac{1}{2}\sqrt{n}$  volte sarà data dal doppio del numero trovato (...), 0.682688 e, di conseguenza, la probabilità del contrario, cioè che l'evento si verifichi più frequentemente o più raramente della proporzione assegnata sopra, sarà 0.317312.<sup>30</sup> ([8], p. 246)

Un aspetto che De Moivre sottolinea è che, a dispetto del fatto che  $\frac{1}{2}\sqrt{n}$  diventi rapidamente trascurabile a fronte di  $n$  quando quest'ultimo cresce all'infinito, vi sia una probabilità piuttosto alta,  $0.682688 \simeq \frac{28}{41}$  che i dati siano lontani non più di  $\frac{\sqrt{n}}{2}$  dal valor medio  $\frac{n}{2}$ . De Moivre considerò anche il caso in cui l'evento ha probabilità diversa da  $\frac{1}{2}$  ma quanto esposto dovrebbe essere sufficiente a comprendere come il ruolo centrale della distribuzione oggi detta normale o gaussiana nel calcolo delle probabilità affondi le sue radici nel passaggio al continuo operato da De Moivre. Bisogna però avvertire che non si tratta di un caso isolato, dal momento che nel XVIII secolo era prassi comune quella di utilizzare integrali per calcolare in modo approssimato delle serie.

## 5.5 Appendice 1: la formula di Stirling

In questa appendice presentiamo una dimostrazione della formula di Stirling, proposta da Francesco Giacomo Tricomi (1897-1978) nel 1938 [23]. Il punto di

<sup>30</sup>And therefore, if it was possible to take an infinite number of experiments, the probability that an event which has an equal number of chances to happen or fail, shall neither appear more frequently than  $\frac{1}{2}n + \frac{1}{2}\sqrt{n}$ , nor more rarely than  $\frac{1}{2}n - \frac{1}{2}\sqrt{n}$  will be expressed by the double sum of the number exhibited (...), 0.682688, and consequently the probability of the contrary, which is that of happening more frequently or more rarely than in the proportion above assigned will be 0.317312.

partenza è l'identità

$$\int \ln x dx = x \ln x - x \quad (5.16)$$

che, preso un numero naturale  $m$  qualsiasi, permette di scrivere

$$\int_m^{m+1} \ln x dx = (m+1) \ln(m+1) - m \ln m - 1. \quad (5.17)$$

Ora, essendo la funzione  $\ln x$  concava possiamo concludere che  $\int_m^{m+1} \ln x dx$  è maggiore dell'area del trapezio individuato dai punti  $A \equiv (m, 0)$ ,  $B \equiv (m+1, 0)$ ,  $C \equiv (m+1, \ln(m+1))$ ,  $D \equiv (m, \ln m)$ , per cui

$$(m+1) \ln(m+1) - m \ln m - 1 > \frac{1}{2} [\ln m + \ln(m+1)]. \quad (5.18)$$

Sommando queste diseguaglianze per tutti i valori interi di  $m$  da 1 a  $n-1$  si ricava

$$n \ln n - (n-1) > \ln 2 + \ln 3 + \ln(n-1) + \frac{1}{2} \ln n = \ln n! - \ln \sqrt{n} \quad (5.19)$$

da cui segue

$$n! < e n^n e^{-n} \sqrt{n} \quad (5.20)$$

che dimostra come i termini della successione

$$f_n := \frac{n!}{n^n e^{-n} \sqrt{n}}$$

appartengano sempre all'intervallo  $[0, e]$  dell'asse reale. Dal momento che

$$\frac{f_n}{f_{n+1}} = \frac{(n+1)^{n+1}}{n^n} \frac{e^{-1}}{\sqrt{n(n+1)}}, \quad (5.21)$$

passando i logaritmi si ha

$$\ln \frac{f_n}{f_{n+1}} = (n+1) \ln(n+1) - n \ln n - 1 - \frac{1}{2} [\ln n + \ln(n+1)] \quad (5.22)$$

che è positivo, grazie alla (5.18). Di conseguenza  $\frac{f_n}{f_{n+1}} > 1$  e quindi la successione  $f_n$  è monotona decrescente. Poiché sappiamo che è anche limitata inferiormente, essa ammette un limite  $k \in [0, e]$ . Per trovarne il valore, si consideri la famiglia di integrali dipendenti da  $n \in \mathbb{N}$

$$I_n := \int_0^{\frac{\pi}{2}} \sin^n x dx.$$

Si dimostra senza problemi che  $I_0 = \frac{\pi}{2}$  ed  $I_1 = 1$  mentre, per valori di  $n \geq 2$  bisogna scrivere

$$I_n = I_{n-2} - \int_0^{\frac{\pi}{2}} \cos^2 x \sin^{n-2} x dx$$

ed integrare per parti, assumendo  $\sin^{n-2} x \cos x dx = \sin^{n-2} x d \sin x$  come fattore finito e  $\cos x$  come fattore differenziale. Si ottiene così

$$I_n = \frac{n-1}{n} I_{n-2} \quad (5.23)$$

da cui si deduce

$$\begin{cases} I_{2m} = \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2 \cdot 4 \cdot 6 \cdots (2m)} \frac{\pi}{2} = \frac{(2m)!}{(2^m m!)^2} \frac{\pi}{2} \\ I_{2m+1} = \frac{2 \cdot 4 \cdot 6 \cdots (2m)}{1 \cdot 3 \cdot 5 \cdots (2m+1)} = \frac{(2^m m!)^2}{(2m+1)(2m)!} \end{cases} \quad (5.24)$$

Poiché  $\sin x \in [0, 1]$  nell'intervallo di integrazione, abbiamo

$$I_n = \int_0^{\frac{\pi}{2}} \sin^n x dx = \int_0^{\frac{\pi}{2}} \sin^{n-1} x \sin x dx < \int_0^{\frac{\pi}{2}} \sin^{n-1} x dx = I_{n-1}$$

e quindi

$$I_{2m+1} < I_{2m} < I_{2m-1}. \quad (5.25)$$

Queste disuguaglianze possono essere tradotte in termini degli elementi della successione  $f_n$ , dal momento che

$$\frac{f_m^2}{f_{2m}} = \frac{(2m+1)\sqrt{2m}I_{2m+1}}{m} \quad \frac{f_{2m}}{f_m^2} = \frac{\sqrt{2m}}{\pi} I_{2m} \quad \frac{f_m^2}{f_{2m}} = 2\sqrt{2m}I_{2m-1}$$

per cui

$$\frac{m}{2m+1} \frac{1}{\sqrt{2m}} \frac{f_m^2}{f_{2m}} < \frac{\pi}{\sqrt{2m}} \frac{f_{2m}}{f_m^2} < \frac{1}{2} \frac{1}{\sqrt{2m}} \frac{f_m^2}{f_{2m}}$$

ovvero, essendo  $f_{2m} > 0$ ,

$$\frac{m}{2m+1} < \pi \frac{f_{2m}^2}{f_m^4} < \frac{1}{2}$$

dal teorema del confronto si ottiene

$$\lim_{m \rightarrow \infty} \pi \frac{f_{2m}^2}{f_m^4} = \frac{\pi}{k^2} = \frac{1}{2}$$

da cui segue che  $k = \sqrt{2\pi}$  e quindi la formula di Stirling

$$n! \simeq n^n e^{-n} \sqrt{2\pi n} \quad n \gg 1.$$

## 5.6 Appendice 2: i numeri di Bernoulli

Diamo una dimostrazione delle identità di Bernoulli, che forniscono le somme

$$S_k(N) = \sum_{n=1}^{N-1} n^k,$$

servendoci del calcolo simbolico, nella versione proposta da John Blissard e Édouard Lucas (1842-1891) nella seconda metà del XIX secolo. consideriamo due polinomi  $f(x, b)$  e  $g(x, b)$  in  $b$  con coefficienti che dipendono da  $x$ :

$$f(x, b) = f_0(x)b^n + f_1(x)b^{n-1} + \dots + f_n(x) \quad g(x, b) = g_0(x)b^n + g_1(x)b^{n-1} + \dots + g_n(x)$$

se  $f(x, b) \equiv g(x, b)$ , deve essere vero, in particolare, che

$$f_i(x) = g_i(x) \quad \forall i = 1, \dots, n.$$

Ora, prese delle quantità  $b_0, b_1, b_2, \dots, b_n$  arbitrarie, sarà anche vero che

$$f_0(x)b_n + f_1(x)b_{n-1} + \dots + f_n(x) = g_0(x)b_n + g_1(x)b_{n-1} + \dots + g_n(x)$$

in cui cioè, alle potenze  $b^k$  sono state sostituite le quantità  $b_k$ . Consideriamo ora l'identità simbolica

$$(b+1)^n - b^n = 0 \tag{5.26}$$

dove l'intero  $n$  è  $n > 1$ . Passiamo alla forma non simbolica: per  $n = 2$  abbiamo

$$b_2 + 2b_1 + 1 - b_2 = 0$$

da cui otteniamo  $b_1 = -\frac{1}{2}$ ; poniamo ora  $n = 3$ ; avremo

$$b_3 + 3b_2 + 3b_1 + 1 - b_3 = 0$$

da cui segue

$$3b_2 = -1 - 3b_1 \quad b_2 = \frac{1}{6}.$$

Se  $n = 4$  abbiamo

$$b_4 + 4b_3 + 6b_2 + 4b_1 + 1 - b_4 = 0$$

da cui otteniamo  $b_3 = 0$  e così via. Consideriamo ora la formula di Taylor per un polinomio  $f(x)$  ed osserviamo che vale l'identità

$$f(x+b+1) - f(x+b) = f'(x) + \frac{f''(x)}{2}[(b+1)^2 - b^2] + \frac{f'''(x)}{6}[(b+1)^3 - b^3] + \dots$$

che continua a valere anche in termini simbolici, sostituendo l'interpretazione simbolica delle potenze di  $b$  data da (5.26). In questo caso si ha la formula simbolica

$$f(x+b+1) - f(x+b) = f'(x) \tag{5.27}$$

che possiamo valutare in  $x = -1$  per il polinomio  $f(x) = x^n$ , ottenendo una nuova identità simbolica

$$b^n - (b-1)^n = (-1)^{n-1}n \tag{5.28}$$

che, sommata a (5.26), fornisce

$$(b+1)^n - (b-1)^n = (-1)^{n-1}n \quad n > 1. \tag{5.29}$$

Nel caso di  $n$  pari,  $n = 2k$ , questa relazione diventa

$$2 \sum_{\ell=1}^k \binom{2k}{2\ell-1} b_{2\ell-1} = -2k$$

e siccome  $b_1 = -\frac{1}{2}$ ,  $\binom{2k}{1} = 2k$  si può scrivere

$$\sum_{\ell=2}^k \binom{2k}{2\ell-1} b_{2\ell-1} = 0.$$

Se si pone successivamente  $k = 2, 3, 4, \dots$  si ottiene

$$4b_3 = 0 \quad 6b_5 + 20b_3 = 0 \quad 8b_7 + 56b_5 + 56b_3 = 0$$

da cui si vede che tutti i numeri  $b_{2\ell-1}$  sono nulli. È possibile dimostrare ([24], pp. 254-257) che tutti gli altri numeri  $b_{2k}$  sono a segni alterni e questa proprietà suggerisce di introdurre i numeri

$$B_k := (-1)^{k-1} b_{2k}$$

che sono detti numeri di Bernoulli. Se ora poniamo  $n = 2k - 1$  in (5.29), otteniamo

$$\sum_{\ell=1}^k \binom{2k-1}{2\ell} B_\ell = k - \frac{1}{2}$$

da cui, aumentando gradualmente il valore di  $k$ , possiamo ottenere una tabella dei numeri di Bernoulli. Consideriamo l'identità simbolica (5.27) ed applichiamo ripetutamente, prendendo per argomenti  $x = 0, 1, \dots, N-1$ . Sommando le identità così ottenute, avremo

$$f'(0) + f'(1) + \dots + f'(N-1) = f(N+b) - f(b) : \quad (5.30)$$

se

$$f(x) = \frac{x^{k+1}}{k+1}$$

giungiamo a

$$S_k(N) = \sum_{n=1}^{N-1} n^k = \frac{(N+b)^{k+1} - b^{k+1}}{k+1}$$

che possiamo porre nella forma usata da De Moivre sviluppando le potenze e sostituendo a  $b^k$  le quantità  $b_k$  che soddisfano (5.26).

# Bibliografia

- [1] N. Bernoulli I: *De usu artis conjectandi in jure*. Conrari, Basel, (1709). In *Die Werke von Jakob Bernoulli*. Band 3, pp. 287-326.
- [2] J. Bernoulli: *Ars Conjectandi, opus posthumum*. Tournes, Basel, (1713).
- [3] É. Chauvin: *Lexicon Rationale sive Thesaurus Philosophicum*. van der Slaart, Rotterdam, (1692).
- [4] F.N. David: *Games, Gods and Gambling. A history of probability and statistical ideas*. General Publishing Company, Toronto, (1962). Ristampato da Dover, New York, (1998).
- [5] M. Dresher: *Games of strategy. Theory and applications*. Prentice-Hall, Englewood Cliffs, New Jersey, U. S. A., (1961).
- [6] A. De Moivre: De mensura sortis seu; de probabilitate eventuum in ludiis a caso fortuito pendentibus. *Transactions of the Royal Society of London*, **27**, 213-264, (1710-1712).
- [7] A. De Moivre: *Miscellanea analytica de seriebus et quadraturis*. Tonson & Watts, London, (1730).
- [8] A. De Moivre: *The Doctrine of Chances*. Millar, London, (1756).
- [9] J. De Witt: *Treatise on Life annuities*. In: R.G. Barnwell: *A sketch of the life and times of John De Witt, Grand Pensionary of Holland*. Pudney & Russel, New York, (1856).
- [10] L. Euler: Calcul de la probabilité dans le jeu de rencontre. *Mémoires de l'Académie des Sciences de Berlin*, **7**, (1753), 255-270.
- [11] T. Hailperin: *Sentential Probability Logic*. Lehigh University Press, Bethlehem, Pennsylvania, U.S.A. (1996).
- [12] A. Hald: *History of probability and Statistics and their applications before 1750*. Wiley-Interscience, Hoboken, New-Jersey, U.S.A. (1988).
- [13] J. Henny: Niklaus und Johann Bernoullis Forschungen auf den Gebiet der Wahrscheinlichkeitsrechnung in ihrem Briefwechsel mit Pierre Rémond de Montmort. In *Die Werke von Jakob Bernoulli*. Band 3, pp. 457-507, (1975).

- [14] M.G. Kendall: Studies in the history of probability and statistics. XVIII. Thomas Young on coincidences. *Biometrika*, **55**, (1968), 249-250.
- [15] J.-H. Lambert: Examen d'une espece de superstition ramenée au calcul des probabilités. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Berlin*, (1771), 411-420.
- [16] P.-S. de Laplace: *Théorie analytique des probabilités*. II Édition. Courcier, Paris, (1814).
- [17] J. von Neumann: Zur Theorie der Gesellschaftsspiele *Mathematische Annalen*, **100**, 295-320, (1928).
- [18] J. von Neumann, O. Morgenstern: *The theory of games and economic behavior*. Princeton University Press, Princeton, U.S.A., (1944).
- [19] P. Rémond de Montmort: *Essay d'Analyse sur les Jeux de Hazard*. II Édition. Quillau, Paris, (1713).
- [20] I. Schneider: Der Mathematiker Abraham De Moivre (1667-1754). *Archive for History of exact Sciences*, **5**, (1968), 177-317.
- [21] G. Shafer: Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of exact Sciences*, **19**, (1978), 309-370.
- [22] S.M. Stigler: *The History of Statistics*. Belknap Press, Cambridge (U.S.A.), (1986).
- [23] F.G. Tricomi: Dimostrazione della formula di Stirling relativa ad  $n!$ , per via del tutto elementare. *Atti della Reale Accademia delle Scienze di Torino*, **74**, (1938-39), 105-109.
- [24] J.V. Uspensky, M.A. Heaslet: *Elementary Number Theory*. Mc-Graw-Hill, New York and London, (1939).
- [25] T. Young: Remarks on the probabilities of error in physical observations, and on the density of the Earth, considered, especially with regard to the reduction of experiments on the pendulum. *Philosophical Transactions of the Royal Society of London*, **109**, (1819), 70-95.

## Capitolo 6

# Il paradosso di San Pietroburgo

### 6.1 Le origini

Il paradosso di San Pietroburgo rappresentò la prima *crisi* del calcolo delle probabilità ed obbligò gli studiosi ad un ripensamento dei fondamenti della disciplina. Inoltre, come era successo con il problema della ripartizione della posta, anche il paradosso di S. Pietroburgo abbandonò presto l'ambito dei giochi d'azzardo per fornire ulteriori occasioni di applicazione del calcolo delle probabilità. Il problema che diede origine al paradosso fu formulato per la prima volta da Nicolaus Bernoulli in una lettera indirizzata a Rémond de Montmort il 9 settembre 1713 pubblicata, con la risposta di Montmort nell'appendice alla seconda edizione dell'*Analyse*. Bernoulli propose la risoluzione di cinque problemi, gli ultimi due dei quali sono rilevanti per il nostro studio. Il quarto problema fu formulato in questi termini:

*A* promette di dare a *B* uno scudo se ottiene sei al primo lancio con un dado ordinario, due scudi se ottiene sei al secondo lancio, tre scudi se lo ottiene al terzo, quattro se lo ottiene al quarto lancio, e così via. Si chiede quale sia la speranza di *B*.<sup>1</sup>

Il successivo problema varia, in modo significativo, la legge con cui viene assegnata la posta ad ogni partita:

Stessa domanda se *A* promette di dare a *B* gli scudi nella progressione 1, 2, 4, 8, 16 ecc. o 1, 3, 9, 27, ecc. o 1, 4, 9, 16, 25 ecc. o 1, 8, 27, 64 al posto di 1, 2, 3, 4, 5, ecc. come prima. Benché la maggior

---

<sup>1</sup>*A* promet de donner un écu à *B*, si avec un dé ordinaire il amene au premier coup 6 points, deux écus s'il amene le 6 au second coup, 3 écus s'il amene ce point au troisième coup, 4 écus s'il l'amene au quatrième et ainsi de suite; on demande quelle est l'esperance de *B*?

parte di questi problemi non siano difficili, vi troverete qualcosa di assai intrigante.<sup>2</sup> ([31], p. 402)

La risposta di Montmort lascia trapelare il fatto che egli non avesse colto il punto delicato che si celava nella formulazione del problema:

Gli ultimi due dei vostri cinque problemi non offrono alcuna difficoltà, non trattandosi altro che di trovare la somma di successioni i cui numeratori sono progressioni di quadrati, cubi, ecc. mentre i denominatori sono in progressione geometrica. Il vostro compianto zio fornì un metodo per trovare la somma di queste serie.<sup>3</sup> ([31], p.407)

Bernoulli richiamò in seguito Montmort a prestare maggior attenzione alla soluzione, in una lettera del 20 febbraio 1714:

Quanto dite sulla facilità degli ultimi due problemi è vero ma avreste fatto bene a cercarne la soluzione perché vi avrebbe dato occasione di un'osservazione molto curiosa. Detta  $x$  la speranza di  $B$  nel caso del quarto problema, avrete  $x = (1 + 5y)/6$  (chiamo  $y$  la speranza di  $B$  nel caso in cui non ottenga 6 al primo lancio); ora,  $y$  è necessariamente  $= x + 1$  perché, dopo aver mancato il successo al primo colpo, egli spera di ricevere gli scudi secondo la progressione 2, 3, 4, 5, 6, nella quale ogni termine supera di una unità il termine corrispondente della progressione 1, 2, 3, 4. Sostituite quindi  $x + 1$  al posto di  $y$  ed otterrete  $x = (5x + 6)/6$ , e dunque  $x = 6$ , come avreste anche ottenuto ricorrendo alle successioni infinite.

Se però seguite la stessa analisi negli esempi del problema 5 come nell'esempio della progressione 1, 2, 4, 8, ecc., dove avrete  $y = 2x$ , troverete  $x = (1 + 10x)/6 = -1/4$ , che è una contraddizione. Per rispondere a questa contraddizione si potrebbe dire che questa frazione, considerata come avente il denominatore negativo e pertanto più piccola di zero, è più grande di  $1/0$  per cui la speranza di  $B$  è più che infinita, come si ottiene adoperando le successioni infinite. Da ciò seguirebbe che  $B$  dovrebbe dare ad  $A$  una somma infinita o meglio più che infinita (se possiamo esprimerci così), perché  $A$  gli possa fornire il vantaggio di dargli gli scudi secondo la progressione 1, 2, 4, 8, 16, etc. Ora è chiaro che se  $B$  desse questa cifra, la perderebbe sempre perché è moralmente impossibile che  $B$  non ottenga sei entro un numero finito di lanci.<sup>4</sup> ([33], p. 558)

<sup>2</sup>On demande la meme chose si  $A$  promet à  $B$  de luy donner des écus en cette progression 1, 2, 4, 8, 16 etc. ou 1, 3, 9, 27, etc. ou 1, 4, 9, 16, 25 etc. ou 1, 8, 27, 64 etc. au lieu de 1, 2, 3, 4, 5, etc. comme auparavant. Quoique ces Problemes pour la plûpart ne sont pas difficiles, vous y trouverez pourtant quelque chose de fort curieux.

<sup>3</sup>Les deux derniers de vos cinq Problemès n'ont aucune difficulté, il ne s'agit que de trouver les sommes des suites dont les numerateurs étant en progression des quarrés, cubes, etc. les denominateurs soient en progression géométrique: feu M. votre Oncle a donné la methode de trouver la somme de ces suites.

<sup>4</sup>Il est vrai ce que vous dites que les deux derniers de mes Problemes n'ont aucune difficulté, cependant vous auries bien fait d'en ehereher la solution, car elle Vous auroit fourni l'occasion de faire une remarque tres curieuse. Soit appelle  $x$  l'esperance de  $B$  dans le cas du 4me

Il calcolo della prima speranza matematica è corretto e fa uso del metodo di Huygens. Per utilizzare le serie infinite, osserviamo che la probabilità che  $B$  vinca al primo lancio è  $\frac{1}{6}$ , al secondo è  $\frac{5}{6} \cdot \frac{1}{6}$ , all' $n$ -esimo lancio è  $\left(\frac{5}{6}\right)^{n-1} \frac{1}{6}$ . Pertanto la speranza matematica è

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 3 \cdot \left(\frac{5}{6}\right)^2 \frac{1}{6} + \cdots + n \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} + \cdots = \frac{1}{6} \sum_{n=1}^{\infty} n \left(\frac{5}{6}\right)^{n-1}.$$

Ora, siccome

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{d}{dx} \sum_{n=1}^{\infty} x^n = \frac{d}{dx} \left( \frac{x}{1-x} \right) = \frac{1}{(1-x)^2}$$

la speranza matematica risulta pari a 6, come asserito succintamente da Bernoulli. Nel quinto problema, al contrario, la speranza matematica è

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 2^2 \cdot \left(\frac{5}{6}\right)^2 \frac{1}{6} + \cdots + 2^n \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} + \cdots = \frac{1}{6} \sum_{n=0}^{\infty} n \left(\frac{5}{3}\right)^n$$

e dunque è infinita.

Il riferimento ai numeri negativi come maggiori di infinito, riecheggia una interpretazione che risale al matematico inglese John Wallis (1616-1703), in opposizione a quella più diffusa che li riteneva essere “meno del nulla”.

Dall'analisi dei risultati proposta a Montmort, Bernoulli conclude che la misura della speranza matematica adottata sino a quel momento non è adatta al problema:

Il motivo di questo è che i casi che hanno una probabilità molto piccola debbono essere trascurati e ritenuti nulli anche se possono far conseguire una speranza molto grande.<sup>5</sup>

---

Probleme, vous aures  $x = (1 + 5y) / 6$  (je nomme  $y$  l'esperance de  $B$  apres qu'il a manque le six au premier coup) ; or  $y$  est necessairement  $= x + 1$ ; car apres qu'il a manque le six au premier coup, il espere de recevoir des ecus en cette progression 2, 3, 4, 5, 6, dont chaque terme est d'une unite plus grand que le terme correspondant de celle ci 1, 2, 3, 4. Substitues donc  $x + 1$  au lieu de  $y$ , et Vous aures  $x = (5x + 6) / 6$ , et partant  $x = 6$ . Ce que Vous auries aussi trouve par la voye des suites infinies. Mais si Vous suives la meme analyse dans les exemples du 5me Probleme comme dans l'exemple de cette progression 1, 2, 4, 8, etc., ou Vous aures  $y = 2x$ , Vous trouveres  $x = (1 + 10x) / 6 = -1/4$ , ce qui est une contradiction. Pour repondre à cette contradiction, on pourroit dire que cette fraction regardee comme ayant le denominateur negatif et par consequent plus petit que zero, est plus grande que  $1/0$ , et qu'ainsi le sort de  $B$  est plus qu'infini, ce qu'on trouve aussi effectivement par la voye des suites infinies. Mais il suivroit de là que  $B$  devoit donner à  $A$  une somme infinie et meme plus qu'infinie (s'il est permis de parler ainsi) pour qu'il luy puisse faire l'avantage de lui donner des ecus en cette progression 1, 2, 4, 8, 16, etc. Or il est certain que  $B$  en donnant une telle somme perdroit toujours, puisqu'il est moralement impossible que  $B$  n'amene pas le six dans un nombre de coups fini.

<sup>5</sup>La raison de ceci est que les cas qui ont une tres petite probabilité doivent être negligés et censés pour nuls, quoiqu'ils puissent apporter une tres grande esperance.

Montmort, pur avendo effettuato dei calcoli riguardanti il problema, non ebbe più occasione di scrivere a Nicolaus Bernoulli sull'argomento e, morto nel 1719, lo lasciò isolato, dal momento che né lo zio Johann né gli allievi di quest'ultimo mostrarono interessi per il calcolo delle probabilità, fino a Gabriel Cramer (1704-1752) che si occupò della questione all'inizio del 1727. Cramer, in una lettera indirizzata a Nicolaus Bernoulli il 21 maggio del 1728 da Londra, oltre a proporre la variante del paradosso che conosciamo, basata cioè sul lancio di una moneta e non di un dado, cercò una via d'uscita distinguendo tra il valore *quantitativo* attribuito ad una certa somma di denaro sulla base della sua entità ed il valore *qualitativo*, che gli compete in virtù di ciò che con quel denaro si vuol fare: se i matematici si erano occupati fino a quel momento del valore quantitativo, la gente assennata (*les hommes de bon sens*) si interessava della seconda accezione. Cramer immagina come ragionerebbe un uomo assennato, trovandosi a giocare secondo le regole stabilite:

Ciò che rende la speranza matematica infinita è la somma prodigiosa che posso ricevere se “testa” non avviene che molto tardi, al 100° o al 1000° lancio. Ora questa cifra non ha per me più valore, non mi arreca più soddisfazione, non mi spinge ad accettare la scommessa, più di quanto farebbe la somma di 10 o 20 milioni di scudi.<sup>6</sup> ([33], pp. 560-561)

Cramer introdusse un troncamento superiore alla cifra massima che  $B$  può ottenere:  $2^{24}$  scudi, superiore a 16 milioni. Il valore della speranza matematica così modificata diventa, in scudi,

$$\sum_{n=1}^{24} \frac{1}{2^n} \times 2^{n-1} + \sum_{n=25}^{\infty} \frac{1}{2^n} 2^{24} = \frac{1}{2} \times 24 + \sum_{n=1}^{\infty} \frac{1}{2^n} = 12 + 1 = 13.$$

Questa somma rappresenta ciò cui si riduce la speranza matematica, dal punto di vista *morale* (*moralement parlant*). Cramer ottenne un valore ancora più esiguo con un altro procedimento:

È ben vero che 100 milioni fanno più piacere di 10 milioni: non però nella misura di dieci volte tanto. P. es., se si facesse l'ipotesi che il *valore morale* dei beni fosse come la radice quadrata della loro quantità matematica, cioè, che la soddisfazione che mi arreca ricevere la cifra di 40000000 fosse il doppio di quella (...) che mi reca ricevere 10000000.<sup>7</sup> ([33], p. 561)

In questo caso la *speranza morale* è

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \times \sqrt{2^{n-1}} = \frac{1}{\sqrt{2}} \sum_{n=1}^{\infty} \left( \frac{1}{\sqrt{2}} \right)^n = \frac{1}{2} \frac{\sqrt{2}}{\sqrt{2}-1}.$$

<sup>6</sup>Ce qui rend l'Esperance Mathematique infinie, c'est la somme prodigieuse que je peux recevoir, si le côté de la Croix ne tombe que bien tard, le 100<sup>e</sup> ou le 1000<sup>e</sup> coup. Or cette somme, si je raisonne en homme sensé, n'est pas plus pour moi, ne me fait pas plus de plaisir, ne m'engage pas plus à accepter le parti, que si elle n'etoit que 10 ou 20 Millions d'Ecus.

<sup>7</sup>il sera vrai que 100 Millions font plus de plaisir que 10 Millions, quoiqu'ils n'en fassent pas 10 fois plus. P. E. Si l'on vouloit supposer que la valeur Morale des Biens fut comme la racine quarrée de leurs quantites Mathematiques, C. à D. que le plaisir que me fait 40000000 fut double du plaisir que (...) me fait m'en fait 10000000.

Questo valore non è ancora la somma che sono disposto a scommettere che invece deve essere

tale che il dispiacere di perderla sia uguale alla *speranza morale* del piacere che spero di ricevere vincendo.<sup>8</sup> ([33], p. 561)

Essa è dunque pari a

$$\left(\frac{1 + \sqrt{2}}{2\sqrt{2} - 1}\right)^2 = \frac{1}{2(3 - 2\sqrt{2})}$$

che è poco meno di 3 scudi e quindi vicina alla stima comune (*estime vulgaire*) di 13 scudi ottenuta poco prima. L'importanza del paradosso di San Pietroburgo—che in questo momento non ha ancora un nome—risiede tutta nel disallineamento tra le previsioni basate sul concetto di aspettazione o, meglio, di speranza matematica, ed il comune sentire o la speranza morale (*esperance morale*): se si vuole che le tecniche della probabilità non si esauriscano nel mondo futile dei giochi d'azzardo ma incidano sulla vita reale, occorre superare questa dicotomia. Se Cramer limitò la somma massima ottenibile per rendere finita l'aspettazione del gioco, Nicolaus Bernoulli—forse influenzato dall'idea di *certezza morale* dello zio Jakob seguì—la via opposta, sostenendo che la spinta a non puntare una cifra eccessiva derivava dal fatto che la probabilità di ottenere una vincita considerevole fosse trascurabile. Per questo riteneva che la risposta di Cramer non cogliesse il vero motivo del dissidio tra speranza matematica e stima comune:

Nel caso di testa o croce, non c'è chi, dotato di buon senso, vorrebbe dare 20 scudi non tanto perché l'utilizzo o il piacere che si può trarre da una somma infinita quasi non superi quelli che si possono ottenere da una somma di 10 o 20 o 100 milioni, ma perché puntando per esempio 20 scudi si ha una probabilità molto piccola di vincere qualcosa e si crede che la perdita sia moralmente certa. La persona comune non tiene in considerazione né i milioni né le centinaia di scudi, non prestando alcuna attenzione al fatto che i termini della progressione geometrica 1, 2, 4, 8, 16, etc. una volta cresciuti si possano considerare uguali tra loro: non è per questo che egli è spinto ad accettare o a rifiutare la scommessa; si decide solo a seconda dei gradi di probabilità che egli ha di vincere o perdere; per lui una probabilità molto piccola di vincere una grande somma non controbilancia una probabilità molto grande di perdere una somma piccola: egli considera impossibile l'avverarsi del primo caso e certo quello del secondo. Occorre dunque, per fissare correttamente l'equivalente, determinare sino a dove la grandezza di una probabilità debba diminuire per poter essere considerata nulla; ecco però che questa determinazione è impossibile: quale che sia l'ipotesi che si faccia, si incontrano sempre delle

<sup>8</sup>tel que le chagrin de sa perte soit egal à l'*Esperance morale* du plaisir que j'espere de recevoir en gagnant.

difficoltà; le limitazioni di queste piccole probabilità non sono precise ma hanno una certa ampiezza che non si riesce a determinare facilmente; per esempio, una probabilità che possiede  $1/100$  di certezza non deve essere considerata nulla più di quanto non lo sia quella che possiede  $1/99$  di certezza. Mi sembra dunque che, supponendo che un uomo di buon senso non voglia giocare 20 scudi, perché tiene per certo che la somma che otterrà sarà minore di 20 scudi, sia possibile trovare l'equivalente cercato grazie al seguente ragionamento: per ipotesi è moralmente impossibile che egli ottenga 20 scudi; sarà dunque anche moralmente impossibile che ne ottenga 32 o qualche altro numero di scudi nella progressione: 32, 64, 128, etc.; ora, la probabilità di ottenere un numero di questa progressione è  $1/64 + 1/128 + 1/256 + \text{etc.} = 1/32$ ; dunque quest'uomo di buon senso reputa come nulla una probabilità che non superi il valore  $1/32$ , ed una probabilità di avere  $31/32$  come certezza intera, pertanto la sua speranza ubbidisce alla regola  $1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 4 + 1/16 \cdot 8 + 0 \cdot 16 + 0 \cdot 32 + \text{etc.} = 2$ . Non so però se sia più corretto quest'altro ragionamento: Un uomo che non voglia giocare 20 scudi ritiene tutti i casi che gli fanno avere una somma inferiore a 20 scudi come possibili ed ognuno degli altri, che gli possono dare una somma maggiore come impossibili; egli ritiene dunque solo le probabilità inferiori ad  $1/32$  come nulle e pertanto la sua speranza varrà  $1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 4 + 1/16 \cdot 8 + 0 \cdot 16 + 0 \cdot 32 + \text{etc.} = 2$ .<sup>9</sup> ([33],

---

<sup>9</sup>dans le Cas de Croix ou Pile il n'y a personne de bon sens qui voulut donner 20 Ecus, non par cette raison que l'usage ou le plaisir qu'on peut tirer d'une somme infinie n'est guère plus grand que celui qu'on peut tirer d'une somme de 10, ou 20, ou 100 Millions, mais parce qu'en donnant par ex. 20 Ecus on a une très petite probabilité de gagner quelque chose, et que l'on croit la perte moralement certaine. Le vulgaire ne met ici en ligne de compte ni des millions, ni des centaines d'Ecus, ne faisant point attention à ce que les termes de la progression geometrique 1, 2, 4, 8, 16, etc. devenus asses grands peuvent être censés égaux, il n'est engagé par là ni à accepter ni à refuser le parti, il se determine seulement selon les degrés de probabilité qu'il a de gagner ou de perdre; chez lui une très petite probabilité de gagner une grande somme ne contrebalance pas une très grande probabilité de perdre une petite somme, il regarde l'évenement du premier cas comme impossible, et l'évenement du second comme certain. Il faut donc, pour regler au juste l'Equivalent, déterminer jusqu'où la quantité d'une probabilité doit diminuer, afin qu'elle puisse être censée nulle; mais voilà ce qui est impossible de déterminer, quelque supposition que l'on fasse, on rencontre toujours des difficultés; les limites de ces petites probabilités ne sont pas precises, mais elles ont une certaine latitude que l'on ne peut pas fixer aisément; une probabilité qui par ex. a  $1/100$  de certitude ne doit pas être reputée nulle plustôt que celle qui a  $1/99$  de certitude. Il me semble pourtant qu'en admettant cette supposition qu'un homme de bon sens ne veuille pas donner 20 Ecus, parce qu'il tient pour certain que la somme qui lui echerra sera moindre que 20 Ecus, on puisse trouver l'equivalent cherché par le raisonnement suivant: par hypoth. il est moralement impossible qu'il obtienne 20 Ecus; il sera donc aussi moralement impossible qu'il obtienne 32 Ecus ou quelque autre nombre d'Ecus dans cette progression: 32, 64, 128, etc.; or la probabilité pour obtenir un nombre de cette progression est  $1/64 + 1/128 + 1/256 + \text{etc.} = 1/32$ ; donc cet homme de bon sens reputa une probabilité qui ne surpasse pas  $1/32$  comme nulle, et une probabilité qui a  $31/32$  comme une certitude entière, par conséquent son esperance vaudra par la regle  $1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 4 + 1/16 \cdot 8 + 0 \cdot 16 + 0 \cdot 32 + \text{etc.} = 2$ . Mais je ne sai si cet autre raisonnement seroit plus juste: Un homme qui ne veut pas donner 20 Ecus estime tous les cas qui lui donnent une moindre somme que 20 Ecus possibles, et chacun des autres, qui lui peuvent donner une somme plus grande, impossible; il regarde donc seulement les probabilités moindres que  $1/32$  comme nulles, par consequent son esperance

pp. 562-563)

La difficoltà risiede nell'arbitrarietà del criterio in base al quale ritenere nulla una probabilità bassa che introduce un elemento di soggettività difficile da valutare correttamente. Il 27 settembre Cramer rispose a Nicolaus Bernoulli dichiarandogli che l'obiettivo della sua soluzione non era quello di indovinare il motivo che spinga un uomo a non accettare un'aspettazione infinita, quanto quello di cercare un motivo per convincere *se stesso* a non concedere un'aspettazione infinita e mantenne il punto nel ritenere più vantaggioso il proprio criterio, che operava sulla somma massima ritenuta appetibile che non quello proposto da Bernoulli, basato sulla difficile stima del grado minimo di probabilità che si può confondere con zero. In una lettera del 27 agosto del 1728, Nicolaus Bernoulli comunicò il problema al cugino Daniel che si trovava a San Pietroburgo già dal 1724. La prima reazione di Daniel al problema fu molto simile a quella di Montmort ed egli si limitò a comunicargli che il paradosso per lui si fondava sulla scarsa probabilità che il gioco ha di durare più di 20 o 30 partite. Nicolaus, come aveva già fatto con Montmort, invitò Daniel a non sottovalutare la portata del problema e questi tornò sull'argomento proponendo una via per uscire dal paradosso che comunicò al padre, Johann, in una lettera del tardo autunno del 1729, purtroppo perduta. Comunque, la soluzione di Daniel giunse al cugino che non se ne dichiarò soddisfatto. L'argomento di Daniel, a quanto possiamo ricostruire dalla risposta di Nicolaus del 4 febbraio 1730, si basava sulla considerazione dei beni posseduti dal giocatore intenzionato a scommettere; al contrario, Nicolaus poneva un quesito provocatorio:

Chiedo come bisogna comportarsi *qui ed adesso* se *A* volesse liberarsi dall'impegno preso senza giocare: quale somma dovrà corrispondere a *B*?<sup>10</sup> ([33], p. 565)

Daniel continuò ad occuparsi del problema e il 5 aprile 1732 inviò copia della memoria *Specimen theoriae novae metiendi sortem pecuniariam* a Nicolaus che, oltre a rammentare come Cramer fosse partito da principii simili, non trovava convincente la soluzione di Daniel dal momento che

Non si tratta di misurare l'uso od il piacere che si trae da una somma di denaro guadagnata, né la privazione od il dolore che deriva dalla perdita di una somma. Non si tratta neppure di cercare un equivalente tra queste cose; si tratta invece di trovare quanto un giocatore sia tenuto, secondo giustizia ed equità, a concedere all'altro per il vantaggio che quest'ultimo gli concede nel gioco d'azzardo in questione od in altri giochi in generale, affinché possa essere ritenuto equo, come lo è per esempio un gioco nel momento in cui i due giocatori puntano la stessa somma sotto condizioni uguali, mentre secondo la Vostra teoria e prestando attenzione ai loro beni,

---

vaudrà  $1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 4 + 1/16 \cdot 8 + 0 \cdot 16 + 0 \cdot 32 + \text{etc.} = 2$ .

<sup>10</sup>Je demande ce qu'il faut faire *hic et nunc*, si *A* vouloit se degager de son obligation sans jouer, quelle somme devra-t-il payer à *B*?

il piacere od il vantaggio del guadagno nel caso favorevole non bilanciano il dolore o lo svantaggio che si soffre in caso contrario.<sup>11</sup> ([33], p. 566)

Siamo così arrivati alla pubblicazione dello *Specimen theoriae novae de mensura sortis* avvenuta nel 1738, benché Daniel avesse inviato il manoscritto all'Accademia delle Scienze di San Pietroburgo fin dal 1731 [2]. Daniel, richiamato il concetto di *valore dell'aspettazione* (*valor expectationis*), cioè a dire la speranza matematica, osserva che la teoria della probabilità si riduce ad enumerare tutti i casi possibili, risolvendoli in quelli che hanno egual facilità di manifestarsi (*aeque proclives*) per disporli poi nella classe opportuna. L'affermazione:

il valore dell'aspettazione è ottenuto quando i singoli valori che si ottengono vengono moltiplicati per il numero di casi nei quali possono essere ottenuti e la somma dei prodotti è divisa per il numero di tutti i casi: [i matematici] impongono però di considerare casi che hanno la stessa facilità a presentarsi<sup>12</sup> ([2], p. 175)

è giustificata—*demonstrata*—dice Bernoulli, con il ricorso al principio di ragione insufficiente.

dal momento che non vi è alcun motivo perché qualcosa in più sia da attribuire ad uno piuttosto che all'altro dei contraenti, a ciascuno va attribuita la stessa quantità.<sup>13</sup> ([2], p. 175)

Se questo punto di vista può soddisfare un giudice che debba dirimere una lite, per Bernoulli occorre qui fornire dei consigli in base ai quali

ciascuno debba stimare per suo conto la propria sorte in virtù della diversa consistenza delle sue sostanze.<sup>14</sup> ([2], p. 176)

Come biasimare un uomo di modeste condizioni economiche se, avendo uguali possibilità di ottenere 20000 ducati oppure nulla, si risolvesse a cedere ad altri questa possibilità a prezzo di 9000 ducati, inferiore alla speranza matematica di 10000 ducati? Al contrario, un uomo ricco che non si risolvesse ad acquistare questa possibilità a 9000 ducati sbaglierebbe, per Bernoulli. La classica definizione di *aspettazione matematica* non può essere presa come misura *universale*

<sup>11</sup>Il ne s'agit pas de mesurer l'usage ou le plaisir qu'on tire d'une somme que l'on gagne, ni le défaut d'usage ou le chagrin qu'on a de la perte d'une somme; il ne s'agit non plus de chercher un equivalent entre ces choses là; mais il s'agit de trouver combien un joueur est obligé selon la justice ou selon l'équité de donner à l'autre pour l'avantage que celui ci lui accorde dans le jeu de hazard en question, ou dans d'autres jeux en general, afin que le jeu puisse être censé égal, comme par exemple un jeu est censé égal, lorsque les deux joueurs mettent une somme égale au jeu sous des conditions égales, quoique selon Vôte theorie, et en faisant attention à leurs richesses, le plaisir ou l'avantage du gain dans le cas favorable n'égale pas le chagrin ou le desavantage qu'on souffre dans le cas contraire.

<sup>12</sup>valorem expectationis obtineri, cum valores singuli expectati multiplicentur per numerum casuum quibus obtingere possunt, aggregatumque productorum dividatur per summam omnium casuum: casus autem considerare [geometrae] iubent, qui sunt inter se aequae proclives

<sup>13</sup>quod cum nulla sit ratio, cur expectanti plus tribui debeat uni quam alteri, unicuique aequae sint adiudicandae partes.

<sup>14</sup>quisque suam sibimet aestimare debeat sortem pro diversa rerum suarum constitutione.

delle possibilità (*mensura sortis*). Di un bene qualunque occorre distinguere tra il prezzo (*Pretium*), su cui tutti sono chiamati a concordare, ed il valore (*valor*) che invece è soggettivo (*aestimatur ex conditione personae*) dipendendo dall'utile (*emolumentum*) che se ne ricava. Nell'esempio precedente il *pretium* era lo stesso per il povero ed il ricco ma la prospettiva di guadagno ha più attrattiva per il povero che per il ricco. L'utilità media (*emolumentum medium*) è così definita da Bernoulli:

quando le *utilità* delle singole aspettative sono moltiplicate per il numero di casi nei quali vengono ottenute e la somma dei prodotti viene divisa per il numero di tutti i casi, si ottiene l'*utilità* media ed il guadagno corrispondente a questa *utilità* media equivarrà alla sorte richiesta.<sup>15</sup>

L'utilità media ha la stessa definizione formale dell'aspettazione ma per renderla operativa occorre stabilire quale utilità corrisponda ad un certo guadagno—*lucrum*—formulando delle ipotesi a partire dalla considerazione che un guadagno interessa maggiormente a chi è in condizioni economiche peggiori. Per questo motivo Bernoulli riteneva molto probabile (*Ita vero valde probabile est*) che

un piccolo guadagno qualsiasi porti sempre un'utilità inversamente proporzionale alla somma delle ricchezze.<sup>16</sup> ([2], p. 178)

L'ammontare delle ricchezze non comprende soltanto i beni materiali ma anche la propria forza lavoro, il proprio ingegno che possono essere utilizzati per svolgere attività che portano guadagni. In questo senso, Bernoulli può affermare che tutti possiedono qualcosa, a meno che non siano ridotti alla fame. Egli vuole giustificare l'ipotesi fondamentale con una semplice considerazione: due persone possiedono, rispettivamente centomila ducati e centomila semi-ducato e se ciascuno riceve annualmente 5000 ducati o 5000 semiducati, rispettivamente, un ducato per la prima persona ha lo stesso significato di un semiducato per la seconda e pertanto, se *entrambi* ricevono un ducato, l'utilità (*emolumentum*) è maggiore per la seconda persona (1 ducato = 2 semiducati) che per la prima, mentre il guadagno (*lucrum*) è lo stesso per entrambi.<sup>17</sup>

Convinto il lettore della ragionevolezza dell'ipotesi di fondo, Bernoulli inizia a svilupparne le conseguenze quantitative, a partire dalla determinazione della curva che rappresenta l'utilità  $y$  in funzione della ricchezza  $x$  posseduta da un individuo. Sull'asse delle ascisse, con origine  $A$ , si introduce dapprima il segmento  $AB = \alpha$  che rappresenta le ricchezze di un individuo, *prima* di affrontare un evento aleatorio che può aumentarle o diminuirle. Siano  $BC$ ,  $BD$ ,  $BE$ ,  $BF$  gli incrementi di ricchezza o guadagni valutati a partire dalle ricchezze iniziali  $\alpha$ . A ciascuno di questi incrementi corrisponderà un valore dell'utilità, rappresentato

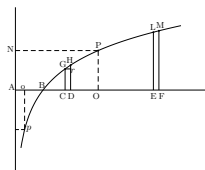
<sup>15</sup>Cum *emolumenta* singula expectata multiplicantur per numerum casuum, quibus obtinebuntur aggregatumque productorum dividitur per numerum omnium casuum, obtinebitur *emolumentum* medium, et *lucrum* huic *emolumento* respondens aequivalebit sorti quaesitae.

<sup>16</sup>*lucrum* quodvis semper *emolumentum* afferre summae bonorum reciproce proportionale.

<sup>17</sup>Questo argomento, all'apparenza inattaccabile, nasconde un'insidia perché dà per scontata la possibilità che due soggetti confrontino le rispettive utilità. Questo confronto intersoggettivo delle utilità è un punto piuttosto delicato (si veda [27], p. 34)

sull'asse delle ordinate. I segmenti  $CG$ ,  $DH$ ,  $EL$ ,  $FM$  sono le utilità associate ai guadagni  $BC$ ,  $BD$ ,  $BE$ ,  $BF$ . Se gli incrementi di ricchezza  $BC$ ,  $BD$ ,  $BE$ ,  $BF$  si possono verificare, rispettivamente, in  $m$ ,  $n$ ,  $p$ ,  $q$ ,... casi, l'utilità media sarà, per definizione, rappresentata da un punto  $P$  corrispondente all'ascissa  $O$  tale che

$$PO = \frac{m \cdot CG + n \cdot DH + p \cdot EL + q \cdot FM + \dots}{m + n + p + q + \dots} : \quad (6.1)$$



Se sull'asse delle ordinate si riporta il segmento  $AN = PO$ , allora  $BO = AO - AB$  rappresenta il guadagno (*lucrum*) che ci si può attendere legittimamente. Una prima domanda che Bernoulli affronta è: quanto è bene scommettere per ottenere questo guadagno? I punti  $G$ ,  $H$ ,  $L$ ,  $M$  che rappresentano le coppie ordinate (ricchezza o guadagno, utilità) si dispongono su una curva, la cui equazione  $y(x)$  dovrà essere determinata. Per il momento Bernoulli si basa solo sulla definizione di utilità media e per rispondere alla domanda egli considera il punto  $o$  sull'asse delle ascisse, a sinistra di  $B$ , cui corrisponda il punto  $p$  sulla curva tale che  $op = OP$ . Bernoulli dà per scontata la monotonia della curva per cui  $y(o) < 0$  e  $op$  rappresenta una perdita. La regola da seguire è di non rischiare una cifra che darebbe un danno morale (cioè un'utilità negativa) superiore al guadagno che legittimamente ci si può attendere. Il segmento  $oB$  rappresenta la cifra massima che posso rischiare per ottenere il guadagno aleatorio  $BO$ .

Per trovare l'equazione  $y = y(x)$  della curva che traduce la dipendenza funzionale tra ricchezze ed utilità occorre formalizzare l'ipotesi iniziale: un incremento infinitesimo di utilità  $dy$  corrispondente ad un guadagno infinitesimo  $dx$  è inversamente proporzionale alle ricchezze possedute  $x$ . Dunque, se  $AC = x$ ,  $CG = y$ ,  $CD = dx$  ed  $rH = dy$ , deve essere

$$dy = \frac{b dx}{x}, \quad (6.2)$$

dove  $b$  è una costante di proporzionalità per cui, presa la ricchezza iniziale  $\alpha$  come livello di riferimento per valutare gli incrementi o decrementi di utilità, cosicché  $y(\alpha) = 0$ , si ha

$$y(x) = b \ln \frac{x}{\alpha}. \quad (6.3)$$

La definizione (6.1) permette di trovare il punto  $P$  sulla curva (6.3) cui corrisponde un punto  $O$  sull'asse delle ascisse:  $P = y(O)$ . Ciò che distingue l'utilità dall'aspettazione è il suo legame *non* lineare con le ricchezze possedute. Ottenuta l'espressione analitica della curva di utilità  $y(x)$ , Bernoulli ritorna alla interpretazione geometrica dell'utilità media che, giova ricordarlo, è un concetto più generale dell'ipotesi (6.2) che conduce alla curva (6.3). Grazie alla (6.3) si ottiene

$$PO = b \ln \frac{AO}{AB} \quad CG = b \ln \frac{AC}{AB} \quad DH = b \ln \frac{AD}{AB} \dots \quad (6.4)$$

che, inseriti in (6.1) conducono a

$$AO = (AC^m \cdot AD^n \cdot AE^p \cdot AF^q \dots)^{1/(m+n+p+q+\dots)} \quad (6.5)$$

da cui, sottratta  $AB = \alpha$ , si giunge al guadagno atteso (*sortem quaesitam*).

$$BO = (AC^m \cdot AD^n \cdot AE^p \cdot AF^q \dots)^{1/(m+n+p+q+\dots)} - \alpha. \quad (6.6)$$

Come prima applicazione, Bernoulli mostra che anche nei giochi che possano ritenersi equi, sulla base della definizione tradizionale, chi gioca deve essere pronto a subire una perdita, risultato da cui trae una "morale":

certamente [è] un notevole ammonimento della natura ad evitare l'azzardo.<sup>18</sup> ([2], p. 183)

La giustificazione matematica sta sempre nella concavità della curva di utilità  $y(x)$  per cui, per attendersi un *danno* (*detrimentum*) che coincida con l'utilità attesa, si avrà  $Bo < BO$ , dove  $o$  è il punto tale che  $op = OP$ . Consideriamo due giocatori, ciascuno in possesso di 100 ducati, 50 dei quali vengono scommessi in un gioco dove entrambi hanno la stessa probabilità di successo. Se poniamo  $m = n = 1$  nella (6.5), allora la ricchezza attesa è  $AO = \sqrt{(100 - 50)(100 + 50)} \simeq 86.6$  ducati per cui la *sortem* (6.6)  $BO$  è inferiore di oltre 13 ducati alla ricchezza iniziale di 100 ducati e il punto  $O$  è a sinistra

<sup>18</sup>egregium profecto naturae documentum vitandae aleae.

di  $B$ , indicando che si è subito un danno.<sup>19</sup> Bernoulli calcolò anche quale fosse la puntata massima  $x$  da effettuare in un gioco che permetta di guadagnare  $a$  ovvero di perdere  $x$ , con pari probabilità. Se  $AB = \alpha$  e  $BC = a$ , l'utilità corrispondente al guadagno in caso di vittoria è

$$CG = b \ln \frac{AC}{AB} = b \ln \frac{AC + BC}{AB} = b \ln \frac{a + \alpha}{\alpha}.$$

Il danno (*detrimentum*) che avremo perdendo  $x$  è misurato da un punto  $g$  della curva corrispondente ad una riduzione  $Bc = x$  delle ricchezze, cioè

$$cg = \left| b \ln \frac{Ac}{AB} \right| = b \left| \ln \frac{AB - Bc}{AB} \right| = \left| b \ln \frac{\alpha - x}{\alpha} \right| = b \ln \frac{\alpha}{\alpha - x}$$

e quindi l'uguaglianza  $CG = cg$  diventa

$$b \ln \frac{\alpha}{\alpha - x} = b \ln \frac{a + \alpha}{\alpha}$$

da cui segue

$$x = a \frac{\alpha}{\alpha + a} < a :$$

si ricava pertanto da tutto questo che agirebbe stoltamente chi mettesse in pericolo tutte le sue ricchezze, nella speranza di un guadagno grande quanto si voglia, ciò di cui nessuno faticherà a persuadersi, se si soppesano correttamente le definizioni che abbiamo premesso.<sup>20</sup> ([2], p. 184)

Una ulteriore applicazione riguarda le assicurazioni sulle merci che viaggiano via nave. Caio, mercante di S. Pietroburgo, acquista ad Amsterdam delle merci il cui valore, se le vendesse in quel momento a S. Pietroburgo, sarebbe di 10000 rubli. Decide dunque di farsele inviare via mare ma è in dubbio se fare assicurare o meno il carico. I dati sulla base dei quali prendere una decisione sono:

- su 100 navi che annualmente compiono la rotta dai Paesi Bassi alla Russia, 5 fanno naufragio;
- il prezzo minimo di una assicurazione è di 800 rubli, che Caio ritiene una enormità.

Bernoulli vuole determinare l'ammontare  $x$  delle ricchezze di Caio, oltre ai 10000 rubli, per poter *ragionevolmente* sostenere la spesa. Egli applica la formula (6.5)

<sup>19</sup>Nel modello di Bernoulli, questo risultato non dipende dal fatto che la cifra scommessa sia troppo alta rispetto al valore delle ricchezze di cui i giocatori sono inizialmente in possesso. Se ogni giocatore scommette  $x$  ducati, avremo  $BO = \sqrt{(100 - x)(100 + x)}$  che è sempre minore di 100. Infatti il radicando è la media geometrica di  $100 - x$  e  $100 + x$  mentre 100 ne rappresenta la media aritmetica, che è maggiore o almeno uguale a quella geometrica. L'uguaglianza si ottiene per  $x = 0$ .

<sup>20</sup>exinde etiam deducitur, stulte hunc agere qui omnia sua bona periclitatur spe lucris quantumvis magni, quod nemo difficulter sibi persuadebit, si recte perpenderit definitiones nostras praemissas.

con  $m = 95$  ed  $n = 5$ , ottenendo, nel caso in cui *non* si stipuli l'assicurazione una ricchezza:

$$\sqrt[100]{(x + 10000)^{95}x^5} = \sqrt[20]{(x + 10000)^{19}x}$$

mentre, stipulando l'assicurazione, la ricchezza sarà  $(x + 9200)$ . Pertanto  $x$  deve risolvere l'equazione  $(x + 10000)^{19}x = (x + 9200)^{20}$  ed è pari a circa 5043 rubli: Caio agirà bene non assicurando il carico se possiede una cifra superiore a quella appena determinata, ma farà bene ad assicurarlo in caso contrario. Anche l'assicuratore deve però premunirsi e dunque Bernoulli calcola quale debba essere il minimo valore  $y$  delle sue ricchezze per poter offrire un contratto di assicurazione al prezzo di 800 rubli. Per questo occorre risolvere l'equazione

$$y = \sqrt[100]{(y + 800)^{95}(y - 9200)^5} = \sqrt[20]{(y + 800)^{19}(y - 9200)}$$

che ha come soluzione  $y \simeq 14243$  rubli: l'assicuratore può ragionevolmente offrire il contratto di assicurazione al prezzo di 800 rubli se possiede almeno questa somma ma non in caso contrario. Un altro esempio di natura commerciale riguarda un altro mercante, Sempronio, che possiede certamente 4000 ducati oltre ad 8000 ducati in merci che però si trovano in terre esotiche e che occorre trasportare via mare: è opportuno affidare il carico ad un'unica nave oppure ripartirlo su due navi, sapendo dall'esperienza (*ex diuturno rerum usu*) che una nave su dieci fa naufragio sulla rotta che occorre far seguire alle merci. Queste sono le alternative considerate:

- se affida le merci ad una sola nave, in 9 casi la sua ricchezza salirà a 12000 ducati ed in un caso rimarrà di 4000 ducati. Sempronio può attendersi un guadagno di

$$\sqrt[10]{12000^9 \cdot 4000^1} - 4000 \simeq 6751 \quad \text{ducato}$$

- se affida la merce a due navi si considerano  $100 = 10 \times 10$  casi possibili in  $81 = 9 \times 9$  dei quali entrambe le navi arriveranno a destinazione, in  $18 = 2 \times 9 \times 1$  solo una delle due arriverà a destinazione mentre nel caso restante entrambe naufragheranno, perdendo tutto il carico. Se ciascuna delle navi porta merci per 4000 ducati, il guadagno atteso sarà

$$\sqrt[100]{12000^{81} \cdot 8000^{18} \cdot 4000^1} - 4000 \simeq 7033 \quad \text{ducato}$$

È solo a questo punto, esaurite le applicazioni più importanti, che Bernoulli si dedica al paradosso (di S. Pietroburgo), ripercorrendone la storia—in particolare pubblicò le considerazioni di Cramer—per poi proporre una soluzione basata sulla teoria dell'utilità appena elaborata. Detto  $N$  il numero di partite che si vogliono giocare, ve ne sono  $\frac{N}{2}$  in cui il giocatore vince un ducato,  $\frac{N}{4}$  in cui ne guadagna 2,  $\frac{N}{8}$  in cui ne guadagna 4 e così di seguito. Se i beni posseduti inizialmente dal giocatore ammontano ad  $\alpha$  ducati, il suo guadagno atteso (*sors*) è dato da

$$\sqrt[N]{(\alpha + 1)^{\frac{N}{2}}(\alpha + 2)^{\frac{N}{4}}(\alpha + 4)^{\frac{N}{8}} \cdots} - \alpha = \sqrt{(\alpha + 1)^4} \sqrt[4]{(\alpha + 2)^8} \sqrt[8]{(\alpha + 4)} \cdots - \alpha \quad (6.7)$$

che cresce con  $\alpha$ , rimanendo però sempre *finita* se  $\alpha$  è finito. Nel caso particolare<sup>21</sup> in cui  $\alpha = 0$  questa *sors* vale 2 ducati, cresce attorno a 3 ducati se  $\alpha = 10$ , a 6 ducati se  $\alpha = 1000$ . Il paradosso è evitato ricorrendo all'utilità anziché all'aspettazione ma, come è stato giustamente notato [24], la memoria di Bernoulli *non* è centrata sul paradosso, che diventa una illustrazione di una teoria dall'ampio spettro di possibili applicazioni. Sul fatto che Bernoulli avesse *risolto* il paradosso, il consenso non fu unanime: d'altra parte Cramer aveva proposto una funzione di utilità diversa, proporzionale alla radice quadrata e non al logaritmo delle ricchezze. Lo *Specimen* ebbe comunque il merito di aprire un nuovo campo di applicazioni della probabilità, anche se non si deve pensare ad una rapida diffusione della teoria di Bernoulli che sarà però esposta nella *Théorie analytique* di Laplace, all'inizio dell'Ottocento.

Tra i matematici che seguirono la via di troncatura della serie che esprime l'aspettazione, quando la probabilità di un evento è sotto ad una certa soglia [9], citiamo il francese Georges-Louis Leclerc, conte di Buffon (1707-1788) che fu naturalista e matematico e che ebbe notizia del problema di S. Pietroburgo da Cramer, nel 1730. Egli si servì delle tavole di mortalità per stimare in  $\frac{1}{10190}$  la probabilità che un uomo di 56 anni avesse di morire entro un giorno.<sup>22</sup> Ora, un uomo di 56 anni ed in salute non ha alcun timore di morire entro 24 ore e dunque si può ritenere come nulla una probabilità di  $\frac{1}{10000}$  che rappresenta per Buffon la soglia di impossibilità *morale* della realizzazione di un evento. Buffon comunicò le sue idee a Daniel Bernoulli che gli rispose in questi termini, il 19 marzo 1762:

Approvo pienamente il modo in cui stimate le probabilità morali; voi indagate la natura dell'uomo a partire dalle sue azioni e supponete in effetti che nessuno si preoccupi al mattino se morirà in quel giorno; detto questo, poiché secondo voi muore una persona su diecimila [entro un giorno], concludete che uno su diecimila sia da ritenere come il nulla assoluto. È un modo di ragionare degno di un matematico-filosofo; questo principio ingegnoso però sembra condurre ad una quantità inferiore perché l'assenza di paura non è certa in chi è già ammalato. Non combatto il vostro principio che però sembra condurre piuttosto a  $\frac{1}{100000}$  che a  $\frac{1}{10000}$ .<sup>23</sup> ([9],

<sup>21</sup>A stretto rigore,  $\alpha = 0$  farebbe perdere di significato a (6.3) ma l'utilità media resta calcolabile. Il radicando è:

$$2^{\frac{1}{4}} 4^{\frac{1}{8}} 8^{\frac{1}{16}} 16^{\frac{1}{32}} \dots = 2^{\sum_{n=1}^{\infty} n(\frac{1}{2})^{n+1}}.$$

La serie ad esponente è del tipo

$$\sum_{n=1}^{\infty} nx^{n+1} = x^2 \sum_{n=1}^{\infty} nx^{n-1} = x^2 \frac{d}{dx} \left( \sum_{n=1}^{\infty} x^n \right) = x^2 \frac{d}{dx} \left( \frac{x}{1-x} \right) = \frac{x^2}{(1-x)^2}$$

che, quando  $x = \frac{1}{2}$ , ha somma 1.

<sup>22</sup>Buffon afferma che si può scommettere 10189 contro 1 che l'individuo sopravviverà ancora un giorno.

<sup>23</sup>J'approuve fort, monsieur, votre manière d'estimer les limites des probabilités morales; vous consultez la nature de l'homme par ses actions, et vous supposez en fait, que personne ne s'inquiète le matin s'il mourra ce jour-là; cela étant, comme il meurt, selon vous, un sur dix

p. 350)

L'osservazione di Buffon che, come egli stesso afferma, si riferisce all'uomo medio—*l'homme moyen*, ([9], p. 350)—resta fondata, anche con le cautele segnalate da Bernoulli. Molto tempo più tardi Émile Borel (1871-1956) tratterà delle soglie al di sotto delle quali una probabilità può considerarsi nulla, distinguendo tra probabilità nulle su scala umana, terrestre e cosmica ([2], pp. 6-7). Per Borel, le probabilità inferiori ad  $10^{-6}$  sono trascurabili su scala umana; trascurabili su scala terrestre o planetaria se inferiori a  $10^{-15} = 10^{-6}/10^9$ , dove  $10^9$  rappresenta una stima della popolazione umana di inizio Novecento; trascurabili su scala cosmica se inferiore a  $10^{-50}$ : si tratta comunque di stime che non debbono essere prese alla lettera. Prendendole troppo sul serio si rischierebbe di imbattersi nel cosiddetto paradosso del sorite ([11], p. 230), o del mucchio, attribuito ad Eubulide di Mileto (vissuto nel IV secolo avanti Cristo): se accettiamo che un milione di granelli di sabbia formino un mucchio di sabbia e che togliendo un granello di sabbia si ottiene ancora un mucchio di sabbia, ripetendo il ragionamento più volte, giungeremo alla conclusione che un granello di sabbia è ancora un mucchio di sabbia.<sup>24</sup>

Buffon era arrivato, indipendentemente da Cramer e Daniel Bernoulli, ad individuare nel paradosso di S. Pietroburgo le diverse accezioni, quantitativa e qualitativa, di una certa quantità di denaro, come causa del dissidio tra buon senso e aspettazione matematica:

Il motivo di questo dissidio tra il calcolo matematico ed il buon senso mi sembra consistere nella sproporzione esistente tra il denaro e il vantaggio che ne risulta. Nei suoi calcoli il matematico non valuta il denaro se non secondo la sua quantità, cioè per il suo valore numerico; ma l'uomo morale lo deve stimare diversamente, dai vantaggi o dal piacere che può arrecargli; è certo che debba muoversi in quest'ottica e non valutare il denaro altrimenti che in proporzione ai vantaggi che gli arreca e non solo rispetto alla quantità che, superati certi limiti, non potrà aumentare in alcun modo la sua felicità.<sup>25</sup> ([9], p. 370)

---

mille, vous concluez qu'un dix-millième de probabilité ne doit faire aucune impression dans l'esprit de l'homme, et par conséquent que ce dix-millième doit être regardé comme un rien absolu. C'est sans doute raisonner en mathématicien philosophe; mais ce principe ingénieux semble conduire à une quantité plus petite, car l'exemption de frayeur n'est assurément pas dans ceux qui sont déjà malades. Je ne combats pas votre principe, mais il paraît plutôt conduire à  $\frac{1}{100000}$  qu'à  $\frac{1}{10000}$ .

<sup>24</sup>Sulla effettiva trascurabilità di una probabilità di  $\frac{1}{10000}$ , nel suo trattato di probabilità ([28], p. 322) John Maynard Keynes riportò una frase dello storico Edward Gibbon [22]:

Se si istituisse una lotteria pubblica per scegliere una vittima da sacrificare, e se il tuo nome fosse scritto su uno dei diecimila biglietti, ti sentiresti perfettamente tranquillo?[If a public lottery were drawn for the choice of an immediate victim, and if your name were inscribed on one of the ten thousand tickets, should we be perfectly easy?]

<sup>25</sup>La raison de cette contrariété entre le calcul mathématique et le bon sens, me semble consister dans le peu de proportion qu'il y a entre l'argent et l'avantage qui en résulte. Un mathématicien dans son calcul, n'estime l'argent que par sa quantité, c'est-à-dire par sa valeur

Il valore infinito dell'aspettazione è legato a diversi fattori: la cifra messa in palio dopo una trentina di lanci era più grande del denaro che probabilmente circolava in Francia in quel tempo. C'è dunque un problema di insolvenza che rende del tutto irrealistico ed inutile procedere a sommare altri termini nella serie.<sup>26</sup>

Inoltre, l'utilità del denaro vinto consente un'ulteriore riduzione dell'aspettazione morale ma, anziché accettare le proposte di Cramer o Daniel Bernoulli oppure formulare un modello alternativo, Buffon riteneva troppo vago il problema perché ne fosse possibile una formulazione matematica rigorosa e per questo cambiò strategia di attacco:

il primo strumento che si presenta è di confrontare il calcolo matematico con l'esperimento perché, in molti casi, noi possiamo, come dissi, conoscere l'effetto del caso tramite esperimenti ripetuti in modo tanto sicuro quanto sarebbe deducendolo immediatamente dalle cause.<sup>27</sup> ([9], p. 378)

È un inciso importante perché raccorda la probabilità con la frequenza in un esperimento ripetuto nelle stesse condizioni, sulla scia delle considerazioni di Jakob Bernoulli. Operativamente, Buffon fece eseguire ad un bambino (*par un enfant*)  $2048 = 2^{11}$  volte il “gioco di S. Pietroburgo”, ottenendo questi risultati

---

numérique; mais l'homme moral doit l'estimer autrement et uniquement par les avantages ou le plaisir qu'il peut procurer; il est certain qu'il doit se conduire dans cette vue, et n'estimer l'argent qu'à proportion des avantages qui en résultent, et non pas relativement à la quantité qui, passé de certaines bornes, ne pourrait nullement augmenter son bonheur.

<sup>26</sup>Il tema dell'insolvenza sarà ripreso più volte ma considerazioni di questo tipo non risolvono il paradosso, come osserveranno, tra gli altri, sia Bertrand ([3], p. 64) che Keynes ([28], p. 362)

Pietro prende degli impegni cui non può far fronte. Se testa non si presenta che al centesimo lancio, il guadagno di Paolo rappresenterà una massa d'oro più grande del Sole. Pietro sbaglia a promettergliela. L'osservazione è corretta ma non chiarisce nulla. Se si gioca scommettendo centesimi al posto di franchi, granelli di sabbia al posto di centesimi, molecole di idrogeno al posto di granelli di sabbia, il rischio di insolubilità diminuirà senza limite. La teoria non deve fare alcuna differenza. [Pierre prend des engagements qu'il ne peut tenir. Si face ne se présente qu'au centième coup, le gain de Paul représentera une masse d'or plus grosse que le soleil. Pierre le trompe en la lui promettant. L'observation est juste, mais n'éclaircit rien. Si l'on joue des centimes au lieu de francs, des grains de sable ou lieu de centimes, des molécules d'hydrogène au lieu de grains de sable, la crainte d'être insolvable peut diminuer, sans limite. La théorie ne doit pas faire la différence.] ([3], p. 64)

<sup>27</sup>le premier moyen qui se présente est de comparer le calcul mathématique avec l'expérience; car, dans bien des cas, nous pouvons, par des expériences répétées, arriver, comme je l'ai dit, à connaître l'effet du hasard, aussi sûrement que si nous le déduisions immédiatement des causes.

$N^\circ$ giochi	guadagno di $B$ in scudi
1061	1
494	2
232	4
137	8
56	16
29	32
25	64
8	128
6	256

La somma vinta in media è poco più di 5 scudi e, osserva Buffon:

ritengo valido questo risultato perché si basa su un gran numero di esperimenti e perché è anche in accordo con un altro ragionamento matematico e inconfutabile, grazie al quale si trova all'incirca lo stesso equivalente di cinque scudi.<sup>28</sup> ([9], p. 379)

Il ragionamento inattaccabile non differisce in sostanza da quello che condusse Daniel Bernoulli all'utilità espressa dalla (6.7): tra le 2048 partite ve ne saranno, in media, 1024 in cui Paolo vince uno scudo; 512 in cui ne vince 2; 256 in cui ne vince 4; 128 in cui ne vince 8; 64 in cui ne vince 16; 32 in cui ne vince 32 e così via. In questo modo si arriva ad un totale di 2047 partite,<sup>29</sup> per cui ne rimane una di cui non possiamo dire nulla ma che, per Buffon, è trascurabile senza errore sensibile. La vincita attesa, basata su questi valori teorici, è di 5 scudi e mezzo che salirebbe a circa 10 scudi, ripetendo il calcolo su un numero di giochi di S. Pietroburgo pari a  $2^{20} = 1048576$ . Tuttavia non ha alcun senso *pratico* ostinarsi in un numero tanto grande di ripetizioni, visto che per eseguire tanti lanci occorrerebbero più di 13 anni, giocando sei ore al giorno:

convenzione che è moralmente impossibile.<sup>30</sup> ([9], p. 382)

Limitandosi a ripetere la stessa stima teorica per numeri di giochi eseguibili in tempi ragionevoli e prendendo la media delle somme ottenute in ogni gioco, si ottiene un equivalente medio, una media di medie, ancora di cinque scudi. In modo sorprendente, almeno per noi, Buffon non si accontentò dell'accordo qualitativo tra calcolo ed esperienza ma volle servirsene per vedere se

non sia possibile dedurre il rapporto tra il valore del denaro e l'utilità che ne risulta.<sup>31</sup> ([9], p. 382)

Altrimenti detto, Buffon volle trovare una base empirica alla teoria dell'utilità che sostituisse il *modello teorico* di Daniel Bernoulli. Per far questo, convinto

<sup>28</sup>Je tiens ce résultat général pour bon, parce qu'il est fondé sur un grand nombre d'expériences, et que d'ailleurs il s'accorde avec un autre raisonnement mathématique et incontestable, par lequel on trouve à-peu-près ce même équivalent de cinq écus.

<sup>29</sup>Infatti  $\sum_{n=0}^{10} 2^n = 2^{11} - 1$ .

<sup>30</sup>ce qui est une convention moralement impossible.

<sup>31</sup>il ne serait pas possible de tirer la proportion de la valeur de l'argent par rapport aux avantages qui en résultent.

del ruolo *universale* del valore di cinque scudi appena trovato per la speranza matematica nel gioco di S. Pietroburgo, Buffon sostituì allo schema iniziale quello in cui le poste in palio ad ogni gioco formano la progressione geometrica di ragione  $\frac{9}{5}$ . L'aspettazione di questo *gioco efficace* è

$$E_B = \frac{1}{2} \times 1 + \frac{1}{2^2} \left(\frac{9}{5}\right) + \frac{1}{2^3} \left(\frac{9}{5}\right)^2 + \cdots = \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{9}{10}\right)^n = 5.$$

Il modello empirico di Buffon è artificiale, essendo basato sul ruolo cabalistico del valore di 5 scudi attorno al quale si accumulano i risultati sperimentali e le previsioni teoriche condotte su un numero molto grande, ma arbitrariamente troncato, di partite.

Osserviamo che il ruolo del buon senso come guida per condurre lo studio della probabilità su binari più vicini alla realtà fu criticato all'inizio del Novecento dal matematico francese Louis Bachelier (1870-1946), allievo di Poincaré e fondatore della moderna finanza matematica. Presentato il paradosso, così commentava:

Non si può invocare il buon senso quando si tratta di questioni delicate perché esso non permette di riconoscere se l'area compresa tra una curva ed il suo asintoto è finita o no, se una serie è convergente o divergente; gli indizi che può fornire non hanno spesso alcun valore quando si tratta di quantità che possono crescere indefinitamente come nel caso considerato. Il giocatore deve ricevere ad ogni partita quanto avrebbe ottenuto alla partita precedente, moltiplicato per due; se avesse ricevuto la stessa somma, moltiplicata per 1,999 la sua aspettazione sarebbe stata finita; il buon senso non può fare differenza tra i due casi.<sup>32</sup> ([1], pp. 26-27)

L'osservazione è corretta ma, a mio parere, non coglie il fatto che la divergenza della serie *amplifica* la paradossalità delle conclusioni ma non la genera. Infine, ricordiamo che la convergenza dell'utilità di Bernoulli nel problema di S. Pietroburgo non avviene sempre. Nel 1934, l'economista austriaco Carl Menger (1902-1985) mostrò [30] che se, quando testa si presenta per la prima volta all' $n$ -esimo lancio,  $B$  riceve da  $A$  la quantità  $\alpha(e^{2^n} - 1)$ , dove  $\alpha$  è la ricchezza iniziale di  $B$ , non solo la sua speranza matematica ma anche l'utilità attesa (*subjektive Hoffnung*) diverge con  $n$ , dal momento che essa riceve al passo  $n$  del gioco un contributo pari a

$$\frac{1}{2^n} b \ln \left( \frac{\alpha + \alpha(e^{2^n} - 1)}{\alpha} \right) = b,$$

dove il fattore  $\frac{1}{2^n}$  è la probabilità di vittoria al passo  $n$  e si è utilizzata la funzione di utilità (6.3). Poiché  $b > 0$  è finito, l'utilità attesa diverge e la

<sup>32</sup>Le bon sens ne peut être invoqué lorsqu'il s'agit de questions délicates, il ne permet pas de reconnaître si l'aire comprise entre une courbe et son asymptote est finie ou non, si une série est convergente ou divergente; les indices qu'il peut fournir n'ont souvent aucune valeur lorsqu'il s'agit de quantités qui peuvent croître indéfiniment comme dans le cas considéré. Le joueur doit recevoir à chaque partie ce qu'il eût reçu à la partie précédente multiplié par deux; s'il devait recevoir la même somme multipliée par 1,999 son espérance serait finie; le bon sens ne fait cependant aucune différence entre les deux cas.

risoluzione del paradosso resta apparente (*scheinbare*). Di più, anche servendosi di una funzione utilità  $f(w)$  diversa dal logaritmo, purché divergente all'infinito, è sempre possibile ottenere un problema di S. Pietroburgo per il quale l'utilità sia divergente. Infatti, se alla ricchezza  $w_n$  che si raggiungerebbe vincendo al passo  $n$ , corrispondesse un'utilità  $f(w_n)$  di almeno  $2^{n-1}$  ducati, l'utilità attesa sarà

$$\sum_{n=1}^{\infty} \frac{1}{2^n} f(w_n) \geq \sum_{n=1}^{\infty} \frac{1}{2^n} 2^{n-1},$$

che chiaramente diverge. Il merito delle considerazioni di Bernoulli non fu dunque quello di risolvere il paradosso di San Pietroburgo ma di introdurre un nuovo concetto, quello di utilità, che sarà ripreso molto tempo dopo nelle teorie economiche.

## 6.2 La prima critica dei fondamenti del calcolo delle probabilità

Critiche pungenti al calcolo delle probabilità furono mosse da Jean Baptiste Le Rond D'Alembert (1717-1783) in diverse circostanze, a partire dall'articolo *Croix ou pile* che comparve, nel 1754, sulla *Encyclopédie* di cui fu il curatore principale insieme a Denis Diderot (1713-1784). Fu il primo a chiamare problema di S. Pietroburgo, *le problème de Petersbourg* ([16], p. 78), il paradosso di cui ci stiamo occupando, prendendo ispirazione dalla città in cui era stato pubblicato lo *Specimen* di Daniel Bernoulli. Per D'Alembert, il paradosso di S. Pietroburgo era la prova lampante che il calcolo delle probabilità non riposava su principi abbastanza solidi. La sua critica ai fondamenti del calcolo delle probabilità era netta: parlando della successione ripetuta di più lanci consecutivi di una moneta equilibrata, D'Alembert riteneva estremamente improbabile l'avverarsi di una successione di sole teste, rispetto ad una successione in cui teste e croci fossero mescolate ma in un ordine prescritto, rinnegando in qualche modo la distinzione tra *cadentiae* e *punctaturae* introdotta molti secoli prima. Sono due i punti su cui D'Alembert portava l'attenzione dei lettori.

Anzitutto, occorre distinguere tra eventi che accadono *certamente* ed eventi che accadono *necessariamente* o, detto in altri termini ([13], p. 10), occorre distinguere tra possibilità *metafisica* e *fisica* di un evento: la prima riguarda eventi il cui realizzarsi non è assurdo, la seconda riguarda quegli eventi che, oltre a non essere assurdi, non sono neppure troppo straordinari. Per D'Alembert, la successione di  $n \gg 1$  teste consecutive nella ripetizione del lancio di una moneta rientra nella prima categoria di eventi ma non nella seconda. Al contrario, una successione di lanci in cui testa e croce sono intercalate tra loro, è per D'Alembert anche fisicamente possibile. Qui D'Alembert sembra confrontare una successione certamente singolare come quella in cui vi sono solo teste con *l'insieme* di successioni in cui teste e croci sono intercalate. La conseguenza che D'Alembert traeva non era lontana da quella di Nicolaus Bernoulli: occorre considerare come nulle le probabilità troppo prossime a zero. Come però

mettere in pratica questo criterio? Come è possibile che la probabilità passi a zero *bruscamente*, compiendo un salto? Questa difficoltà, che per D'Alembert denunciava l'insufficienza della teoria tradizionale, era aggirata in un modo che egli stesso definiva non scevro da dubbi e praticamente poco perseguibile, ma che contiene un'idea interessante:

Suppongo, ad esempio, che si lanci in aria una moneta quattro volte di seguito; ci saranno  $2^4$  o 16 combinazioni differenti di *testa* e *croce* prese a quattro a quattro. Se dunque si ricomincia questo gioco un numero di volte che è multiplo di 16 o, ciò che è lo stesso, se 32, 64, ecc. giocatori diversi fanno a turno questo gioco, ciascuno in particolare, gettando ognuno uno scudo in aria quattro volte di seguito, è evidente che qualcuna delle 16 combinazioni verrà ripetuta. Ora, credo che le combinazioni che saranno ripetute più raramente, e che può darsi non si presentano affatto in un gran numero di lanci, saranno quelle nelle quali *croce* si trova ripetuta quattro volte o *testa* è ripetuta quattro volte. Dopo questo esperimento, ripetuto un gran numero di volte di seguito, si potrebbe forse stimare il rapporto delle probabilità, attraverso il numero degli eventi realizzatisi. È vero che il risultato potrà lasciare dei dubbi, e che l'esperimento sarà forse irrealizzabile se il numero di lanci, anziché essere pari a quattro, come si è supposto, sarà molto maggiore, come cento; ecco però, così mi sembra, il solo modo di ottenere in questo caso un risultato che sia almeno prossimo al vero.<sup>33</sup> ([13], pp. 12-13)

Questo passo mi sembra significativo perché lascia intravedere quella che sarà l'impostazione di tipo frequentista al calcolo delle probabilità che però D'Alembert, più preoccupato a mostrare l'inconsistenza dell'approccio tradizionale, non sviluppò.

La dicotomia tra possibilità fisica e metafisica emerge chiaramente quando D'Alembert vuole giustificare perché, a suo dire, ottenuta testa una prima volta, al lancio successivo la probabilità di riottenere testa sia ridotta. D'Alembert suppone che, inizialmente, vi siano  $n$  modi per ottenere testa ed altrettanti per ottenere croce. Se si è ottenuto croce al primo lancio, D'Alembert si domanda:

è verosimile che l'impulso che mi fornirà ancora croce al secondo lancio sia *esattamente* lo stesso che me lo ha fornito al primo lancio? A me non

---

<sup>33</sup>Je suppose, par exemple, qu'on jette une pièce en l'air quatre fois de suite; on aura  $2^4$  ou 16 combinaisons différentes de *croix* & *pile* pris quatre à quatre. Si donc on recommence ce jeu un nombre de fois qui soit multiple de 16, ou, ce qui revient au même, si 32 ou 64 &c joueurs differens jouent à la fois ce jeu, chacun en particulier, en jettant chacun un écu en l'air quatre fois de suite; il est évident que quelqu'une ou quelques-unes des 16 combinaisons se trouveront répétées. Or je crois que les combinaisons qui seront répétées le plus rarement, & qui peut-être n'arriveront point du-tout dans un grand nombre de jets, seront celles dans lesquelles *croix* se trouve quatre fois de suite, ou *pile* quatre fois de suite. D'après cette expérience, répétée un grand nombre de fois de suite, on pourroit peut-être estimer le rapport des probabilités, par le nombre des évènements. Il est vrai que le résultat pourra laisser de doutes; & que d'ailleurs l'expérience seroit impraticable, si le nombre de jets, au lieu d'être de quatre, ainsi qu'on l'a supposé, étoit beaucoup plus grand, comment de cent; mais voilà, ce me semble, le seul moyen de parvenir en ce cas à un résultat qui soit au moins approchant du vrai.

## 6.2. LA PRIMA CRITICA DEI FONDAMENTI DEL CALCOLO DELLE PROBABILITÀ 173

sembra. Infatti, in questo caso, non resteranno che  $n - 1$  modi per ottenere *croce* al secondo tentativo mentre ve ne sono ancora  $n$  per ottenere *testa* al secondo lancio. Vi è dunque già un po' più di probabilità di ottenere *testa* rispetto a *croce* al secondo lancio.<sup>34</sup> ([18], p. 39)

D'Alembert, consapevole dell'impossibilità pratica di riprodurre due volte le stesse condizioni iniziali di lancio di una moneta, ritiene che ogni lancio che ha dato, ad esempio, *croce* come esito, riduca il numero di condizioni favorevoli a quell'evento. Se una moneta non è truccata, dopo avere ottenuto tante volte di seguito *croce* si è indotti a congetturare che, al lancio successivo, si otterrà *testa*. Alla base di questo ragionamento c'è la constatazione che

l'esperienza e la conoscenza che abbiamo delle leggi di natura ci insegnano che uno stesso evento non si presenta mai un gran numero di volte di seguito; è in virtù di questa *conoscenza acquisita* che noi dubitiamo della ripetizione di *testa* o *croce* un gran numero consecutivo di volte. Dal momento che tutto è legato nell'ordine delle cose, noi potremmo, conoscendo la concatenazione delle cause e degli effetti, divinare e predire l'esito di ogni lancio, se sarà *testa* o *croce*; nell'ignoranza in cui siamo circa i segreti della natura, non possiamo dire con certezza se otterremo *testa* o *croce*; poiché però l'esperienza ci ha insegnato che lo stesso effetto si ripete raramente, potremo almeno congetturare che, se *croce* si è verificata molte volte di seguito, allora si otterrà verosimilmente *testa* al prossimo lancio.<sup>35</sup> ([18], pp. 48-49)

In tutto questo, è bene ricordarlo, D'Alembert suppone che la moneta sia equilibrata perché, in caso contrario, la conclusione potrebbe essere molto diversa.

Strettamente legato a questa problematica è il secondo punto sollevato da D'Alembert, che è una proposta di revisione radicale dei concetti di probabilità ed aspettazione matematica. È possibile, si chiede D'Alembert, considerare equivalenti un gioco in cui vi sia certezza di ottenere 500 scudi ed un altro in cui vi sia probabilità  $\frac{1}{2}$  di ottenerne 1000 ed  $\frac{1}{2}$  di non ottenerne nulla?

La difficoltà proviene, se non mi sbaglio, dal fatto che l'idea di *speranza* comprende due cose; la somma che ci si attende e la probabilità di

---

<sup>34</sup>Est-il, vraisemblable que l'impulsion qui me donnera encore *croix* au second coup, sera précisément la même, que celle qui me-l'avoit donné au premier coup? Il me semble que non. Or en ce cas, il n'y aura plus que  $n - 1$  manieres d'amener *croix* au second coup, tandis qu'il y en a encore  $n$  d'amener *pile* à ce second coup. Il y a donc déjà un peu plus de probabilité pour *pile* au second coup, que pour *croix*.

<sup>35</sup>L'expérience & la connoissance que nous avons des loix de la nature, nous apprennent que le même événement n'arrive jamais un grand nombre de fois de suite; & c'est en vertu de cette *connoissance acquise*, que nous révoquons en doute la répétition de *croix* ou de *pile* un grand nombre de fois consécutives. Comme tout est lié dans l'ordre des choses, nous pourrions, si nous connoissions la loi de l'enchaînement des causes & des effets, deviner & prédire ce qui arrivera à chaque coup, si ce sera *croix* ou *pile*; dans l'ignorance où nous sommes du secret de la nature, nous ne pouvons dire précisément si ce sera *pile* ou *croix*; mais comme l'expérience nous a appris que le même effet se répète rarement, nous pouvons au moins, lorsque *croix* est arrivé plusieurs fois de suite, conjecturer avec vraisemblance que *pile* viendra.

guadagnarla. Ora, mi sembra che sia *soprattutto la probabilità* a dover stabilire la speranza e che la somma attesa non debba entrare (...) che in modo subordinato rispetto al grado di probabilità; tuttavia le si fa entrare entrambe e nella stessa misura nel calcolo.<sup>36</sup> ([16], pp. 82-83)

A questo proposito, osserviamo che il passaggio dalla aspettazione all'utilità operato da Daniel Bernoulli rompe questa simmetria grazie al legame non lineare tra ricchezza ed utilità. D'Alembert però si rifà alle riflessioni di Buffon, che aveva proposto una valutazione dei guadagni e delle perdite in un gioco, riferite a quanto posseduto da un giocatore: se egli ha beni pari ad  $a$  e può guadagnare  $x$ , i suoi beni dopo aver giocato diverranno  $a + x$  ed il guadagno reale sarà  $\frac{x}{a+x}$ ; al contrario, se egli perdesse  $x$ , allora la perdita è misurata da  $\frac{x}{a-x}$ . Poiché  $\frac{x}{a-x} > \frac{x}{a+x}$  D'Alembert conclude che il gioco d'azzardo è sempre svantaggioso.

Tornando al problema di San Pietroburgo, per D'Alembert ammetterne il carattere paradossale dimostrava il fatto che una sequenza estremamente lunga di lanci di una moneta che diano sempre lo stesso risultato era ritenuta impossibile e che il suo realizzarsi dovesse nascondere una causa che agisce a favore di un risultato tanto singolare. D'Alembert alimentava il suo argomento proprio con alcune considerazioni di Bernoulli, per il quale il fatto che tutti i pianeti allora noti del sistema solare giacessero in una porzione molto piccola della sfera celeste provava la presenza di una causa soggiacente, diversa dalla pura casualità. D'Alembert riteneva questo atteggiamento una prova che un'uniformità singolare fosse fisicamente impossibile, benché, dal punto di vista matematico, si tratta di una configurazione come tante. Egli non si limitava a proporre argomenti qualitativi ma mostrava che, se la probabilità di ottenere testa ad un lancio, sapendo che si era ottenuto testa in  $n$  lanci precedenti, è minore di  $\frac{1}{2}$  allora, anche l'usuale aspettazione era finita. Il suo ragionamento era il seguente:

- Si suppone che la probabilità che testa esca al secondo lancio, sapendo che è uscita al primo lancio è  $\frac{1-\varepsilon_1}{2}$ , con  $0 < \varepsilon_1 \ll 1$
- La probabilità che testa esca al terzo lancio, sapendo che è uscita ai primi due è  $\frac{1-\varepsilon_1-\varepsilon_2}{2}$ , con  $0 < \varepsilon_2 \ll 1$  e così via, supponendo che  $\sum_{n=1}^{\infty} \varepsilon_n < 1$ ;
- Si calcola, con le regole tradizionali la probabilità che testa esca in  $n - 1$  lanci consecutivi e che croce esca all' $n$ -esimo lancio:

$$\frac{1}{2} \times \frac{1-\varepsilon_1}{2} \times \frac{1-\varepsilon_1-\varepsilon_2}{2} \cdot \frac{1-\sum_{k=1}^{n-1} \varepsilon_k}{2} \times \frac{1+\sum_{k=1}^n \varepsilon_k}{2}$$

dove l'ultimo termine rappresenta la probabilità che croce esca, per la prima volta, dopo  $n$  lanci.

<sup>36</sup>La difficulté vient, si je ne me trompe, de ce que l'idée d'espérance enferme deux choses; la somme qu'on espère, et la probabilité qu'on gagnera cette somme. Or il me semble que c'est *principalement la probabilité* qui doit régler l'espérance; et que la somme espérée ne doit y entrer, si je puis parler de la sorte, que d'une manière subordonnée au degré de probabilité: cependant on les fait entrer toutes deux également et de la même manière dans le calcul.

- Si calcola l'aspettazione sulla base delle regole *tradizionali*, ottenendo il valore

$$\frac{1}{2}[1 + 1 + \varepsilon_1 + (1 - \varepsilon_1)(1 + \varepsilon_1 + \varepsilon_2) + (1 - \varepsilon_1)(1 - \varepsilon_2)(1 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3) + \dots].$$

- Si dimostra che, a patto di scegliere i valori di  $\varepsilon_n$  opportunamente, l'aspettazione resta *finita*.

L'obiettivo di D'Alembert non è tanto quello di proporre una teoria completa della probabilità che sia alternativa a quella accettata ma di mostrare l'opportunità di rivedere le basi del calcolo, visto che il contrasto con il buon senso può svanire se si considerano altri modelli per esso ed in effetti, anche se non sempre i ragionamenti di D'Alembert sono condivisibili, le sue argomentazioni mettono in luce come la validità dei fondamenti del calcolo delle probabilità non fosse scontata. Ribadiamo che D'Alembert non suppone la moneta truccata ma assumere  $\varepsilon_n \neq 0$  significa tenere conto del fatto che il presentarsi più volte della medesima faccia della moneta riduce l'insieme di condizioni favorevoli al suo realizzarsi.

### 6.3 L'approccio di Condorcet

Il ruolo del Marchese di Condorcet (Marie-Jean-Antoine-Nicholas de Caritat, 1743-1794) nella storia del calcolo delle probabilità è controverso. Il calcolo delle probabilità giocava un ruolo fondamentale nel suo programma di *mathématique sociale*, cioè di applicazione della matematica non solo alle scienze tradizionalmente affini come la fisica ma anche a discipline che coinvolgevano la vita sociale e politica. Questo programma era ereditato dalla visione di Anne-Robert-Jacques Turgot (1724-1781) che fu economista e filosofo oltre che ministro delle finanze sotto Luigi XVI, per poco meno di due anni. Come ricorda Condorcet nell'introduzione all'*Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Turgot

era persuaso che le verità delle scienze morali e politiche fossero suscettibili dello stesso grado di certezza di quelle che formano il sistema delle scienze fisiche e di quelle branche di queste scienze, come l'astronomia, che sembrano raggiungere la certezza matematica.<sup>37</sup> ([10], p. i)

Ora consideriamo l'approccio di Condorcet al problema di S. Pietroburgo che figura nella prima parte di una memoria del 1784, dedicata ad esaminare il concetto di speranza matematica, nell'*Essai* del 1785 e negli *Éléments de probabilité, et son application aux jeux d'hazard, a la loterie et aux jugemens des hommes*, pubblicati dopo la sua morte. Condorcet anzitutto sottolinea che l'aspettazione

<sup>37</sup>était persuadé que les vérités des Sciences morales et politiques, sont susceptibles de la même certitude que celles qui forment le système des Sciences physiques, et même que les branches de ces Sciences qui, comme l'Astronomie, paroissent approcher de la certitude mathématique.

non fornisce il reale vantaggio di un giocatore ma solamente un valore medio di questo vantaggio. L'argomento è semplice: se ho probabilità  $\frac{1}{3}$  di guadagnare 2 e  $\frac{2}{3}$  di guadagnare 1, il mio vantaggio, cioè l'aspettazione è  $\frac{4}{3}$ , valore che però non otterrò *mai*, dal momento che otterrò 2 oppure 1: l'aspettazione, qui calcolata su un singolo evento, fornisce una sorta di punto di equilibrio tra le parti. Il punto è di stabilire se la formula proposta per il calcolo dell'aspettazione sia attendibile o meno. Per Condorcet la risposta è affermativa dal momento che essa gode di tre proprietà:

1. La somma delle differenze positive e negative tra essa ed i valori dati dall'osservazione, ovvero la somma delle differenze tra questo valore ed i valori veri, ugualmente possibili, è uguale a zero.
2. La differenza tra il valore vero incognito, o un valore vero possibile qualunque ed il valore medio è uguale alla somma delle differenze tra ciascuno di questi valori veri e gli altri valori osservati, o i valori possibili, divisa per il loro numero.
3. Prendendo, in circostanze simili, questo valore medio come un valore vero, l'evento più probabile sarà quello nel quale le differenze in eccesso e in difetto tra realtà ed ipotesi si compenseranno; inoltre si avrà una probabilità sempre crescente che la loro somma non supererà che di una parte piccola a piacere la più grande somma possibile di queste differenze mentre, nello stesso tempo, non esiste alcuna altra legge che possa fornire una probabilità anch'essa sempre crescente di non scostarsi che di una quantità sempre costante. Infine si avrà una probabilità che si avvicina sempre ad  $\frac{1}{2}$  ovvero che tende sempre all'uguaglianza tra il fatto che questa somma sia positiva e che sia negativa, e viceversa.<sup>38</sup>

Questo estratto, che evidenzia lo stile piuttosto oscuro di Condorcet, è coerente con il fatto che egli, a differenza di Cramer o Daniel Bernoulli, si adoperò per dimostrare come fosse possibile neutralizzare gli aspetti paradossali del problema di San Pietroburgo senza introdurre un nuovo concetto di aspettazione. Condorcet era d'accordo con il fatto che il problema di S. Pietroburgo avesse messo in crisi l'accordo tra le conclusioni ottenute grazie all'adozione del concetto di

---

38

- (a) Que la somme de differences positive et négatives entre elle et les valeurs données par l'observation, ou bien la somme des différences entre cette valeur et les vrais valeurs également possibles, est égale à zero.
- (b) Que la différence entre la vraie valeur inconnue, ou une vraie valeur possible quelconque, et la valeur moyenne, est égale à la somme des différences entre chacune de ces valeurs vraies et les autres valeurs observées, ou les valeurs possibles, divisée par leur nombre.
- (c) Qu'en prenant, dans ces circonstances semblables, cette valeur moyenne pour une vraie valeur, l'événement le plus probable sera celui où les différences en plus ou en moins entre la réalité et l'hypothèse, se compenseront; qu'on aura une probabilité toujours croissante que leur somme n'excèdera pas une partie aussi petite qu'on voudra de la plus grande somme possible de ces différences, tandis qu'en même temps il n'existe aucune loi qui puisse donner une probabilité toujours croissante de ne s'écarter que d'une quantité toujours approchante de  $\frac{1}{2}$ , ou toujours tendante à l'égalité que cette somme sera positive plutôt que négative, ou réciproquement.

speranza matematica fu favorita dal fatto che le conclusioni cui esso conduceva erano conformi al buon senso (*conformes à la raison commune*). Per lui però questo problema *non* è un problema reale, nel senso che non potrà mai presentarsi nella realtà e dunque deve essere considerato solo come un caso limite e per Condorcet questo carattere asintotico gli fa perdere parte della forza contro la bontà del concetto di speranza matematica. Condorcet però non liquida il problema in modo sbrigativo ed egli stesso non ritiene soddisfacente questa spiegazione, perché il problema persiste anche limitando il numero di partite dal momento che:

la somma che  $B$  dovrebbe dare ad  $A$  per giocare a questo gioco senza questa ipotesi è ancora tale che, per poco che grande diventi il numero di lanci, nessuna persona dotata di senno vorrebbe rischiarla.<sup>39</sup> ([10], p. 714)

Negli *Éléments*, Condorcet osservò che, per decidersi a giocare quando si ha una piccola probabilità di vincere una grande somma, occorre una valutare *a priori* il proprio stato economico prima dell'inizio e dopo la conclusione del gioco. Condorcet suppose che il numero massimo di partite fosse  $n$  e introdusse una variante all'ultima partita: se testa *non* esce nemmeno all'ultimo lancio, il giocatore ( $B$ ) otterrà  $2^n$  denari mentre, se testa esce per la prima volta al lancio  $k$ -esimo,  $B$  riceverà al solito  $2^{k-1}$  denari. Le probabilità di successo di  $B$  sono:  $\frac{1}{2}$  al primo lancio,  $\frac{1}{2^2}$  al secondo lancio,  $\frac{1}{2^3}, \dots, \frac{1}{2^n}$ , se testa si presenta all'ultimo lancio e ancora  $\frac{1}{2^n}$ , se testa non si presenta mai. Per poter giocare,  $B$  deve versare una quota di

$$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + \dots + 2^{n-1} \frac{1}{2^n} + 2^n \frac{1}{2^n} = \frac{n}{2} + 1$$

denari. Ora,  $B$  inizierà a guadagnare quando testa compare, per la prima volta, al lancio  $p$  tale che

$$2^{p-1} > \frac{n}{2} + 1 \quad (6.8)$$

mentre al colpo in cui

$$2^{p-1} = \frac{n}{2} + 1, \quad (6.9)$$

ovvero quando  $n = 2^p - 2$ , si ha l'equilibrio tra vincite e perdite. La probabilità che  $B$  perda del denaro è la somma delle probabilità che egli vinca *prima* del lancio  $p$ -esimo nel quale si verifica la (6.9):

$$\frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{p-1}} = 1 - \frac{1}{2^{p-1}} \quad (6.10)$$

Condorcet considera l'esempio in cui  $p = 4$  e dunque  $n = 2^4 - 2 = 14$ . In questo caso la posta che  $B$  versa è  $\frac{n}{2} + 1 = 8$  denari. La probabilità che  $B$  perda del

<sup>39</sup>la somme que  $B$  devoit donner à  $A$  sans cette hypothèse pour jouer à jeu égal, est encore telle, pour peu que le nombre de coups soit grand, au'aucun homme raisonnable ne voudroit risquier de la donner.

denaro è, per quanto appena visto,  $1 - \frac{1}{2^3} = \frac{7}{8}$ ; quella che  $B$  concluda con un bilancio in pareggio è  $\frac{1}{16}$  così come quella che egli concluda il gioco realizzando un guadagno, benché sia possibile vincere, con probabilità  $\frac{1}{2^{14}} = \frac{1}{16384}$ , una somma massima di  $16376 = 16384 - 8$  denari. Al contrario,  $A$  ha probabilità  $\frac{15}{16}$  di non perdere denaro, esponendosi ad una perdita di 16376 denari che ha probabilità  $\frac{1}{16384}$  di verificarsi e ad un guadagno massimo di  $7 = 8 - 1$  denari, con probabilità  $\frac{1}{2}$ , che si realizza quando esce subito testa. Vi è dunque una grande differenza, una discriminazione, fra le posizioni di  $A$  e  $B$  che non dovrebbero vincolarsi a stipulare l'accordo implicato dal gioco se non modificando la cifra che viene considerata come unità di misura ed il numero  $N$  di partite da giocare, in modo che le probabilità di vittoria dei due contendenti siano confrontabili e che, con grande probabilità, entrambi non possano perdere più di una frazione assegnata del capitale o, con le parole di Condorcet,

vi sia una probabilità molto grande che né  $A$  né  $B$  perdano, in un numero  $m$  di partite, una cifra il cui valore non ecceda una frazione assegnata con  $m$ .<sup>40</sup> ([10], pp. 715-716)

La conclusione è la seguente:

Così, nell'esempio tratto dal gioco di testa o croce<sup>41</sup> (...) si vede che, colui che scommette  $\frac{n}{2}$  a fronte di una probabilità  $\frac{1}{2^n}$  di guadagnare  $2^{n-1}$  [denari] al lancio  $n$ -esimo, non deve risolversi a giocare se non quando potrà ripetere il gioco un numero di volte tale da conferirgli una probabilità pressoché uguale di vincere come di perdere. Similmente, colui che, al contrario, può trovarsi costretto a pagare  $2^{n-1}$  [denari], non avendone ricevuti che  $\frac{n}{2}$ , non deve giocare, a dispetto della grande probabilità che ha di vincere, se non quando egli possa considerare  $\frac{n}{2}$  o, meglio, le somme inferiori, che egli ha una fondata speranza di ottenere, come una compensazione del rischio molto piccolo di perdere la somma ben maggiore di  $2^{n-1}$  [denari], ciò che costringe a prendere  $n$  molto piccolo. Allora i giocatori potranno decidersi a giocare e il paradosso scomparirà.<sup>42</sup> ([11], pp. 113-114)

Se  $N$  è il numero di *giochi di San Pietroburgo* che si vuole effettuare, la probabilità che  $B$  perda *sempre* del denaro, supponendo i giochi indipendenti tra

<sup>40</sup>qu'on ait une assez grande probabilité que ni  $A$  ni  $B$  dans un nombre  $m$  de parties ne perdront au-delà d'une valeur qui ait une proportion donnée avec  $m$ .

<sup>41</sup>Cioè nel problema di San Pietroburgo.

<sup>42</sup>Ainsi, dans l'exemple tiré du jeu de *croix* ou *pile* (...) on voit que celui qui donne une mise  $\frac{n}{2}$ , et qui a probabilité  $\frac{1}{2^n}$  de gagner  $2^{n-1}$  au  $n^e$  coup, ne doit se déterminer à jouer, qu'autant qu'il pourra répéter le jeu assez souvent pour avoir une probabilité presque égale de gagner ou de perdre. De même celui qui, au contraire, peut être obligé de donner  $2^{n-1}$  après n'avoir reçu que  $\frac{n}{2}$ , ne doit jouer, malgré la grande probabilité qu'il a de gagner, que lorsqu'il peut regarder  $\frac{n}{2}$ , ou plutôt les sommes moindres qu'il a une espérance fondée de gagner, comme un dédommagement du risque très-petite de perdre la somme beaucoup plus grande  $2^{n-1}$ , ce qui oblige à faire  $n$  très-petit. Alors les joueurs pourront se déterminer à jouer le jeu, et le paradoxe disparaît.

loro, si ottiene da (6.10) ed è

$$\left(1 - \frac{1}{2^{p-1}}\right)^N$$

per cui la probabilità che  $B$  guadagni del denaro o comunque non ne perda *almeno* una volta è

$$P(N, p) = 1 - \left(1 - \frac{1}{2^{p-1}}\right)^N$$

La tabella seguente, tratta dal libro di Heinrich Bruns (1848-1919) ([9], p.71), contiene i valori di  $P(N, 10)$  per diversi valori di  $N$  e mostra come si debba prendere  $N$  molto grande affinché  $P(N, 10)$  raggiunga  $\frac{1}{2}$ .

$N$	500	600	700	800	900	1000
$P(N, 10)$	0.39	0.44	0.49	0.54	0.58	0.62

L'asimmetria nelle condizioni di  $A$  e  $B$  richiama il problema più generale della possibilità di derogare dalla condizione di gioco equo. L'opportunità di tale deroga è motivata da Condorcet osservando che, nei giochi dove il banco (*le banquier*) affronta più avversari (*les pontes*), egli va incontro ad una piccola probabilità di perdere una grossa cifra, mentre i giocatori hanno al contrario una ragionevole aspettazione di vincere somme non troppo grandi. Violare le regole del gioco equo a favore del banchiere è un sacrificio cui si espongono volontariamente i giocatori tentati di vincere una grossa somma. Similmente, le lotterie non vendono i biglietti al prezzo imposto dalle regole del gioco equo ma ad un costo superiore per compensare la necessità di rimborsare i premi, che deve essere garantita. Come ultimo esempio, non poteva mancare l'allusione alla pratica commerciale di applicare un interesse al prezzo da corrispondere per una merce il cui reperimento comporti un rischio concreto. Rinvio il lettore interessato ad ulteriori dettagli sul problema di San Pietroburgo ad alcuni dei lavori più significativi apparsi nella letteratura, come [22] e [25] o, seppure scritto molto tempo fa, [12].

## 6.4 Gestire il rischio in modo razionale: l'inoculazione del vaiolo

Gli esempi che Daniel Bernoulli aveva esposto a corredo della sua teoria dell'utilità nello *Specimen* [2] mostravano come il calcolo delle probabilità potesse costituire uno strumento razionale per decidere una strategia di comportamento di fronte a decisioni che comportavano un rischio. Una di queste situazioni si ricollegava ad un problema di grande impatto sociale: è opportuno o meno praticare l'inoculazione<sup>43</sup> per ridurre il rischio di mortalità dovuta al vaiolo? Nel XVIII secolo il vaiolo rappresentò la maggiore causa di mortalità in Europa ed

<sup>43</sup>In medicina l'inoculazione è l'introduzione nell'organismo di un agente patogeno tramite un'iniezione od un incisione cutanea.

il dibattito sull'opportunità di campagne di prevenzione fiorì in diversi paesi: in Francia esso coinvolse, tra gli altri, D'Alembert e Daniel Bernoulli. L'inoculazione comportava dei rischi poiché prevedeva il contatto diretto della pelle dell'individuo con pustole vaiolose attive. Si stimava che, per un parigino, la probabilità di morire per aver contratto il vaiolo fosse di  $\frac{1}{7}$  mentre la probabilità di morire a seguito dell'inoculazione era inferiore ma non nulla. Il medico e scienziato inglese James Jurin (1684-1750) riportò in [26] due dati: su una popolazione di 182 pazienti inglesi sottoposti all'inoculazione, 2 erano morti mentre un'inoculazione condotta a Boston, negli stati Uniti, aveva fatto registrare una mortalità di  $\frac{5}{300}$ . Si presentava dunque un dilemma: correre il rischio *certo* ed immediato legato all'inoculazione oppure non sottoporsi all'inoculazione e dunque diluire nel tempo il rischio [22]. Nelle parole di Jurin:

Veniamo ora, Signore, alla seconda parte del nostro progetto che è quella di ottenere una stima del rischio che tutti gli uomini, uno insieme all'altro, corrono di morire del vaiolo naturale e di confrontarla con il rischio dell'inoculazione, cosicché la gente possa formarsi un giudizio se la pratica dell'inoculazione tenda o meno a preservare l'umanità, riducendo il pericolo al quale altrimenti sono soggetti.<sup>44</sup> ([26], pp. 215-216, [32])

Nel 1760 Daniel Bernoulli presentò all'Accademia delle Scienze di Parigi un lavoro [3], in cui affrontava questo problema anche per rispondere ad un'obiezione piuttosto diffusa, non solo tra gli strati bassi della società: vale la pena sottoporsi all'inoculazione visto che, in media, si aggiunge una aspettativa di vita di soli due anni ai bambini nati da poco? Per rispondere in modo scientifico all'obiezione, Bernoulli, il cui lavoro sarà pubblicato solo nel 1766, riteneva che si dovessero confrontare due tavole di mortalità: una che tenesse conto delle morti per vaiolo; l'altra, ipotetica, corrispondente al caso in cui il vaiolo fosse stato debellato. A questo progetto si opponeva l'indisponibilità di un dato cruciale nelle tavole di mortalità esistenti: l'età delle persone decedute a causa del vaiolo. Per ovviare a questa carenza di informazioni Bernoulli svolse alcune considerazioni per stimare i due parametri ritenuti fondamentali:

1. la probabilità che un soggetto di una data età contragga la malattia in un certo anno;
2. la probabilità di guarigione per un soggetto che abbia contratto la malattia.

Anche questi parametri erano difficili da ottenere ma Bernoulli riteneva di poterne surrogare la conoscenza con quella di due altri dati. Egli anzitutto contestava l'idea che il vaiolo si diffondesse tra i più giovani per un problema di costituzione. Se i giovani si ammalavano di più era perché la malattia, salve rare eccezioni,

<sup>44</sup>We come now, Sir, to the second part of our design, which is to form an estimate of the hazard, which all mankind, one with another, are under of dying of the natural small pox, that, by comparing this with the hazard of inoculation, the publick may be enabled to form a judgment, whether or no the practice of inoculation tends to the preservation of mankind, by lessening the danger to which they are otherwise liable.

non veniva contratta più di una volta per cui molti adulti non contraevano la malattia perché erano diventati immuni. Secondo l'analisi di Bernoulli, la probabilità di contrarre il vaiolo entro i cinque anni di età era  $\frac{1}{2}$ ; entro i dieci anni era di  $\frac{3}{4}$ ; entro i 20 anni di età era  $\frac{15}{16}$  e  $\frac{4000}{4001}$  entro i 60 anni di età. L'ipotesi formulata da Bernoulli era la seguente:

finché non si è contratto il vaiolo, si corre sempre lo stesso rischio di contrarlo.<sup>45</sup> ([3], p. 4).

La seconda ipotesi riguardava la probabilità di morire, una volta contratta la malattia:

il rischio di morte a seguito del vaiolo potrà essere considerato lo stesso ad ogni età, in un anno qualsiasi.<sup>46</sup> ([3], p. 4).

Queste ipotesi di lavoro si traducono formalmente in queste richieste:

1. la probabilità che un individuo facente parte di un gruppo formato da un numero rilevante di persone contragga in un certo anno di vita il vaiolo è  $\frac{1}{n}$ , indipendentemente dalla sua età;
2. probabilità che un individuo muoia, una volta contratto il vaiolo, è  $\frac{1}{m}$ , indipendentemente dalla sua età.

Il modello epidemiologico di Bernoulli può ora essere costruito. Si indichi con  $x$  l'età di un individuo, espressa in anni; sia  $\xi = \xi(x)$  in numero di individui che raggiungono quell'età all'interno di una popolazione; sia  $s(x) \leq \xi(x)$  il numero di individui che non hanno ancora contratto il vaiolo all'età  $x$ . Allora  $-ds$  rappresenta il numero di individui che contraggono il vaiolo nel tempuscolo  $dx$  ed è  $-ds = \frac{sdx}{n}$  e quelli che ne muoiono sono  $\frac{sdx}{mn}$ . Gli individui che muoiono per altre cause, nello stesso tempuscolo  $dx$ , sono  $-d\xi - \frac{sdx}{mn}$ . Poiché questo numero va riferito al campione di  $s$  individui,<sup>47</sup> per avere l'espressione corretta di  $-ds$  occorre moltiplicare questo contributo per  $\frac{s}{\xi}$  arrivando a

$$-ds = \frac{sdx}{n} - \frac{s d\xi}{\xi} - \frac{s^2 dx}{mn\xi}. \quad (6.11)$$

Per integrare (6.11), Bernoulli ne riordinò i termini

$$\frac{s d\xi}{\xi} - ds = \frac{sdx}{n} - \frac{s^2 dx}{mn\xi}$$

e, moltiplicando ambo i membri per  $\frac{\xi}{s^2}$ , ottenne

$$\frac{s d\xi - \xi ds}{s^2} = \frac{\xi dx}{ns} - \frac{dx}{mn}. \quad (6.12)$$

<sup>45</sup>tant qu'on n's pas eu la petite vérole on court continuellement la même risque de l'avoir.

<sup>46</sup>le risque de mourir de la petite vérole, quand on en est attaqué, pourroit bien être, année commune, la même a tout âge.

<sup>47</sup>Come ripeterà più avanti, Bernoulli ipotizza che i decessi all'interno della popolazione dovuti a cause *diverse* dal vaiolo siano proporzionali al numero di individui che compongono la popolazione stessa.

Posto  $q := \frac{\xi}{s}$  si ha

$$mndq = m q dx - dx = (mq - 1)dx$$

ovvero

$$\frac{mndq}{mq - 1} = dx \quad (6.13)$$

da cui segue finalmente

$$n \ln(mq - 1) = x + C \quad (6.14)$$

per una opportuna costante di integrazione  $C$ . Ripristinando le variabili iniziali e passando alla forma esponenziale si ottiene

$$\left(\frac{m\xi}{s} - 1\right)^n = e^{x+C} \quad (6.15)$$

e quindi

$$s = \frac{m}{e^{\frac{x+C}{n}} + 1} \xi. \quad (6.16)$$

Per determinare il valore di  $C$ , Bernoulli suppose che all'età  $x = 0$  si avesse  $\xi(0) = s(0)$ , ricavando

$$e^{\frac{C}{n}} = m - 1$$

e, infine,

$$s(x) = \frac{m}{(m-1)e^{\frac{x}{n}} + 1} \xi(x) \quad (6.17)$$

che esprime  $s(x)$  in termini dei parametri di modello e della funzione  $\xi(x)$  che si suppone nota, almeno per punti, dalle tavole di mortalità che Bernoulli utilizza per costruirne a sua volta un'altra in cui vengano messi in luce gli effetti del vaiolo, prendendo  $m = n = 8$  cosicché

$$s(x) = \frac{8}{7e^{\frac{x}{8}} + 1} \xi(x). \quad (6.18)$$

Osserviamo come Bernoulli sia ben consapevole del fatto che supporre  $m$  ed  $n$  costanti sia solo una approssimazione, ragionevolmente plausibile per individui di età inferiore ai 20 anni perché la mortalità aumenta in individui di età maggiore. La teoria di Bernoulli consentiva di stimare in circa  $\frac{1}{13}$  la frazione dell'intera popolazione che muore di vaiolo, conformemente ai dati disponibili

6.4. GESTIRE IL RISCHIO IN MODO RAZIONALE: L'INOCULAZIONE DEL VAIOLO183

sulle cause di mortalità.

$x$	$\xi(x)$	$s(x)$	$\xi(x) - s(x)$	$I(x)$	$M(x)$	$\Sigma(x)$	$A(x)$
0	1300	1300	0				
1	1000	896	104	137	17,1	17,1	283
2	855	685	170	99	12,4	29,5	133
3	798	571	227	78	9,7	39,2	47
4	760	485	275	66	8,3	47,5	30
5	732	416	316	56	7,0	54,5	21
6	710	359	351	48	6,0	60,5	16
7	692	311	381	42	5,2	65,7	12,8
8	680	272	408	36	4,5	70,2	7,5
9	670	237	433	32	4,0	74,2	6
10	661	208	453	28	3,5	77,7	5,5
11	653	182	471	24,4	3,0	80,7	5
12	646	160	486	21,4	2,7	83,4	4,3
13	640	140	500	18,7	2,3	85,7	3,7
14	634	123	511	16,6	2,1	87,8	3,9
15	628	108	520	14,4	1,8	89,6	4,2
16	622	94	528	12,6	1,6	91,2	4,4
17	616	83	533	11,0	1,4	92,6	4,6
18	610	72	538	9,7	1,2	93,8	4,8
19	604	63	541	8,4	1,0	94,8	5
20	598	56	542	7,4	0,9	95,7	5,1
21	592	48,5	543	6,5	0,8	96,5	5,2
22	586	42,5	543	5,6	0,7	97,2	5,3
23	579	37	542	5,0	0,6	97,8	6,4
24	572	32,4	540	4,4	0,5	98,3	6,5

La prima colonna della tabella di Bernoulli rappresenta l'età degli individui di una coorte di 1300 individui che si suppongono essere nati tutti al medesimo istante; la seconda colonna contiene il numero  $\xi(x)$  di individui della coorte sopravvissuti all'età  $x$ ; la terza colonna contiene  $s(x)$ , gli individui che all'anno  $x$  di vita non hanno ancora contratto il vaiolo, secondo quanto ottenuto dalla (6.18). La quarta colonna presenta le differenze  $\xi(x) - s(x)$  che definiscono il numero di individui che, avendo raggiunto l'età  $x$ , hanno già contratto il vaiolo. La quinta colonna indica il numero  $I(x)$  di individui di età  $x$  che hanno probabilmente contratto il vaiolo nell'anno precedente e andrebbe descritto, secondo l'ipotesi di Bernoulli, come  $\frac{s}{8}$ . Qui però Bernoulli introdusse una correzione considerando  $I(x) = \frac{1}{8} \left( \frac{s(x)+s(x-1)}{2} \right)$ . La sesta colonna indica il numero  $M(x) = \frac{I(x)}{8}$  di morti di vaiolo in un anno mentre, nella colonna successiva,  $\Sigma(x)$  è numero di individui della coorte, morti di vaiolo con un'età non superiore ad  $x$  anni. Infine, nell'ultima colonna sono contati i morti  $A(x)$  per cause diverse dal vaiolo. È su questa colonna che occorre basarsi per le rendite vitalizie perché in alcuni contratti non era permesso stipulare una polizza a vantaggio di un giovane che ancora non avesse contratto (e superato) il vaiolo. Bernoulli non procede la ta-

bella oltre  $x = 24$  perché ritiene meno affidabili le ipotesi di modello per età più elevate. Per discutere quantitativamente i vantaggi della prevenzione del vaiolo, Bernoulli costruì una ulteriore tabella confrontando la mortalità nello stato naturale—in cui cioè il vaiolo è presente (*état naturel et variolique*)—con quella in un'ipotetica situazione dove il vaiolo è stato debellato (*état non variolique*).

$x$	$\xi(x)$	$z(x)$	$g(x)$
0	1300	1300	0
1	1000	1017,1	17,1
2	855	881,8	26,8
3	798	833,3	35,3
4	760	802,0	42,0
5	732	779,8	47,8
6	710	762,8	52,8
7	692	749,1	57,2
8	680	740,9	60,9
9	670	734,4	64,4
10	661	728,4	67,4
11	653	722,9	69,9
12	646	718,2	72,2
13	640	714,1	74,1
14	634	709,7	75,7
15	628	705,0	77,0
16	622	700,1	78,1
17	616	695,0	79,0
18	610	689,6	79,6
19	604	684,0	80,0
20	598	678,2	80,2
21	592	672,3	80,3
22	586	666,3	80,3
23	579	659,0	80,0
24	572	651,7	79,7
25	565	644,3	79,3

Le prime due colonne sono rimaste immutate mentre la terza colonna rappresenta la popolazione residua  $z(x)$  di età  $x$  se non vi fosse il vaiolo. Per ottenere gli elementi di questa colonna si parte da  $\xi(1) + M(1) = z(1)$  e si osserva che i morti per altre malattie si mantengono in proporzione al numero di individui che compongono la popolazione per cui, se su 1000 individui che nello stato naturale raggiungono il primo anno di vita, ne muoiono 133, partendo da  $z(1) = 1017,1$  individui, ne moriranno un numero  $y$  tale che

$$\xi(1) : z(1) = A(2) : y$$

cioè, essendo  $A(2) = 133$ ,  $y = 135,2$ . Questo numero rappresenta gli individui della popolazione morti nel passaggio tra il primo ed il secondo anno di vita nello stato ideale, senza vaiolo. Pertanto,  $z(2) = z(1) - y = 881,8$  ed in modo

analogo si costruiscono i valori successivi di  $z(x)$ . Infine la quarta colonna rappresenta il guadagno  $g(x)$  dello stato ideale, senza vaiolo, rispetto allo stato naturale (*gain absolu*) definito semplicemente come  $g(x) = z(x) - \xi(x)$  anche se in verità, come Bernoulli discute nel corso della memoria, è il guadagno relativo (*gain relatif*)  $\frac{g(x)}{\xi(x)}$  a dare una misura precisa dell'impatto dell'inoculazione sulla mortalità. Ora, mentre il guadagno assoluto può anche diminuire al crescere di  $x$ , quello relativo si mantiene crescente, tendendo ad approssimare il valore  $\frac{1}{7}$ , quando  $x$  è sufficientemente grande. Per trovare analiticamente il valore del rapporto  $\frac{z(x)}{\xi(x)}$  nel limite di  $x$  sufficientemente grandi, Bernoulli osservò che la diminuzione  $-d\xi$  nel tempuscolo  $dx$  della popolazione nello stato naturale ha una componente  $\frac{sdx}{mn}$  dovuta al vaiolo per cui la mortalità complessiva in assenza di vaiolo sarà  $-(d\xi + \frac{sdx}{mn})$ . Questa mortalità è però riferita ad una popolazione di  $\xi(x)$  individui; per riferirla alla popolazione dello stato senza vaiolo, occorre moltiplicarla per  $\frac{z}{\xi}$  ottenendo dunque

$$-\frac{z}{\xi} \left( d\xi + \frac{sdx}{mn} \right) = -dz$$

da cui segue

$$\frac{dz}{z} - \frac{d\xi}{\xi} = \frac{sdx}{\xi mn}$$

e, sostituendo il valore di  $s(x)$  trovato nell'equazione (6.17),

$$\frac{dz}{z} - \frac{d\xi}{\xi} = \frac{\frac{1}{n} dx}{(m-1)e^{\frac{x}{n}} + 1}$$

che si riscrive come

$$d \ln z - d \ln \xi = d \ln \frac{z}{\xi} = d \ln e^{\frac{x}{n}} - d \ln [(m-1)e^{\frac{x}{n}} + 1] = d \ln \frac{e^{\frac{x}{n}}}{(m-1)e^{\frac{x}{n}} + 1}$$

da cui si ottiene, osservando che  $z(0) = \xi(0)$ ,

$$\frac{z}{\xi} = \frac{me^{\frac{x}{n}}}{(m-1)e^{\frac{x}{n}} + 1}.$$

Bernoulli conclude che, se si prende un valore sufficientemente grande (*un peu grand*) per  $x$ , si ha  $\frac{z}{\xi} \simeq \frac{m-1}{m}$  che misura il vantaggio dell'inoculazione. Prendendo  $m = 8$  l'assenza di vaiolo offrirebbe ogni anno allo stato ed alla società francese 25 mila persone in più in età utile, cioè superiore ai 16 anni. Tutto questo ci riporta al nucleo del problema sollevato da Jurin: è chiaro che se non ci fosse rischio alcuno nell'inoculazione, non ci sarebbe alcun motivo per non intraprendere una campagna a favore di questa pratica. Il rischio di morte dovuto all'inoculazione, per quanto piccolo, gettava un'ombra di dubbio sulla validità della proposta. La questione viene formalizzata in questi termini da Bernoulli:

Cosa sarebbe dello stato dell'umanità se, con un certo numero di vittime, gli si potesse procurare l'esenzione dal vaiolo presente in natura?<sup>48</sup> ([3], p. 31)

Se una persona su  $N$  muore a seguito dell'inoculazione, occorre modulare i valori di  $z(x)$ , moltiplicandoli per  $\frac{N-1}{N}$ : nel caso numerico esaminato in cui  $N = 200$ , Bernoulli mostra come, per effetto di questi rischi, la vita media si abbassi di un mese e 20 giorni rispetto a quella in assenza di vaiolo ma rimanga ancora molto al di sopra del valore nello stato naturale. Per rincarare la dose, Bernoulli mostrò come  $N$  dovesse scendere al valore 9.43 perché non vi fosse differenza di vita media con lo stato naturale, in cui il vaiolo non viene contrastato con azioni specifiche. Anche con un tasso di mortalità per inoculazione così alto, resterebbe un vantaggio a procedere comunque con l'inoculazione perché la mortalità colpirebbe soprattutto i bambini non ancora utili alla società, mentre sopra i 16 anni vi resterebbe un vantaggio rispetto allo stato naturale.

Bernoulli iniziò a leggere la propria memoria sull'inoculazione il 30 aprile del 1760; il 12 novembre di quello stesso anno, D'Alembert lesse la propria risposta in cui metteva a fuoco due problemi: a suo dire il punto di vista corretto da cui partire nei calcoli sul vantaggio dell'inoculazione non era ancora stato toccato in letteratura; inoltre, le difficoltà, forse insormontabili, di ridurre al calcolo il vantaggio dell'inoculazione non costituiscono una buona ragione per non attuarla. D'Alembert svolse il ruolo di avvocato del diavolo a favore dell'inoculazione, cercando i punti deboli nelle motivazioni addotte in suo favore, per fornire argomenti migliori. D'Alembert contestava l'argomento della disparità del rischio corso da chi non si vaccina, a fronte di una bassa mortalità dovuta all'inoculazione, perché si confrontano rischi che agiscono su scale temporali molto diverse: poiché la morte dovuta all'inoculazione si verificava entro qualche giorno, i rischi connessi andrebbero confrontati con quelli di contrarre il vaiolo entro lo stesso intervallo temporale. La differenza più significativa è che, superata positivamente l'inoculazione, non si temerebbe più la malattia, che invece è continuamente temuta da chi non si premunisce [13]. Nella versione del lavoro [14] letta all'Accademia delle Scienze D'Alembert non fece uso di modelli matematici che sono invece descritti negli *Opuscules Mathématiques* ([15], [17]).

Tra i diversi problemi di indole statistica di cui è intessuto il Capitolo VIII della *Théorie analytique* di Laplace, troverà spazio anche lo studio della mortalità dovuta al vaiolo. Laplace si chiese quale sarebbe stata la durata media della vita se fosse stato possibile eliminare del tutto una causa di mortalità. I suoi calcoli non sono molto diversi da quelli di Bernoulli e non saranno discussi in questa sede.

---

<sup>48</sup>Quel seroit l'état de l'humanité, si moyennant un certain nombre de victimes on pouvoit lui procurer une exemption de la petite vérole naturelle?

# Bibliografia

- [1] L. Bachelier: *Calcul des probabilités*. Tome I. Gauthier-Villars, Paris, (1912).
- [2] D. Bernoulli: Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Petropolitanae*, **4**, 175-193, (1738).
- [3] D. Bernoulli: Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Mémoires de l'Académie Royale des Sciences*, 1-45, (1760).
- [4] J. Bernoulli: *Ars Conjectandi, opus posthumum*, Basel, (1713).
- [5] J. Bertrand: *Calcul des probabilités*. Gauthier-Villars, Paris, (1889).
- [6] É. Borel: *Le Hasard*. Alcan, Paris, (1920).
- [7] É. Borel: *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, Paris, (1952).
- [8] H. Bruns: *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*. Teubner, Leipzig, (1906).
- [9] G.-L. Leclerc de Buffon. *Essai d'arithmétique morale*. In *Œuvres Complètes de Buffon*, Vol. XV, Didot, Paris, (1829), pp. 338-447.
- [10] J.-A.-N. De Caritat, marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris, (1785).
- [11] J.-A.-N. De Caritat, marquis de Condorcet. *Éléments de probabilité, et son application aux jeux d'hazard, a la loterie et aux jugemens des hommes*. Royez, Paris, (1805).
- [12] E. Czuber: Das Petersburger Problem. *Archiv der Mathematik und Physik*, **67**, (1882), 1-28.
- [13] J.B. Le Rond D'Alembert: Reflexions sur le calcul des probabilités. In *Opuscules mathématiques*. Tome II. David, Paris, (1761), pp.1-25.

- [14] J.B. Le Rond D'Alembert: Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. In *Opuscules mathématiques*. Tome II. David, Paris, (1761), pp. 26-46.
- [15] J.B. Le Rond D'Alembert: Notes sur la mémoire précédent. In *Opuscules mathématiques*. Tome II. David, Paris, (1761), pp. 47-95.
- [16] J.B. Le Rond D'Alembert: *Opuscules mathématiques*. Tome IV. Brisson, Paris, 1768.
- [17] J.B. Le Rond D'Alembert: Sur la durée de la vie. In: *Opuscules mathématiques*. Tome IV. Brisson, Paris, 1768, pp. 92-105.
- [18] J.B. Le Rond D'Alembert: Sur le calcul de probabilités. In: *Opuscules mathématiques*. Tome VII. Jombert, Paris, (1780), pp. 39-59.
- [19] L. Daston: *Classical probability in the Enlightenment*. Princeton University Press, Princeton, New Jersey, U.S.A., (1988).
- [20] A. De Moivre: *The Doctrine of Chances: or, a method of calculating the probabilities in events of play*. III Edition. Millar, London, (1766).
- [21] P. Dupont: Un joyau dans l'histoire des sciences: le mémoire de Thomas Bayes de 1763. *Rendiconti del Seminario Matematico dell'Università Politecnica di Torino*, **37**, 105-138, (1979).
- [22] J. Dutka: On the St. Petersburg Paradox. *Archive for History of Exact Sciences*, **39**, 13-39, (1988).
- [23] D.A. Gillies: Was Bayes a Bayesian?. *Historia Mathematica*, **14**, 325-346, (1987).
- [24] N. Giocoli: The 'true' hypothesis of Daniel Bernoulli: what did the marginalists really know? *History of Economic Ideas*, **6**, 7-43, (1998).
- [25] G. Jorland: The Saint Petersburg Paradox 1713-1937. In *The probabilistic revolution*, Volume I. L. Krüger, L.J. Daston, M. Heidelberger, Eds. MIT Press, Cambridge, Massachusetts, U.S.A., (1987), pp. 157-190.
- [26] J. Jurin: A Letter to the Learned Dr. Caleb Cotes-Worth, F. R. S. of the College of Physicians, London, and Physician to St. Thomas's Hospital; Containing a Comparison between the Danger of the Natural Small Pox, and of That Given by Inoculation. *Philosophical Transactions of the Royal Society of London*, **32**, (1722-1723), 213-227.
- [27] E. Kauder: *The History of Marginal Utility Theory*. Princeton University Press, Princeton, New Jersey, U.S.A., (1965).
- [28] J.M. Keynes: *A treatise on probability*. Mac Millan & co., London, (1921).

- [29] P. S. de Laplace: Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie royale des Sciences de Paris (Savants étrangers)*, **6**, 621-656. In: *Œuvres de Laplace*. Vol. 8, Gauthier-Villars, Paris, (1891), 27-65.
- [30] K. Menger: Das Unsicherheitsmoment in der Wertlehre: Betrachtungen im Anschluss an das sogenannte Petersburger Spiel. *Zeitschrift für Nationalökonomie*, **5**, (1934), 459-485.
- [31] P. Rémond de Montmort: *Essay d'Analyse sur les Jeux de Hazard*. II Édition. Quillau, Paris, (1713).
- [32] J.-M. Rohrbasser: Les hasards de la variole. *Astérion. Philosophie, Histoire des idées, pensée politique*, **9**, (2011).
- [33] O. Spiess: Die Vorgeschichte des Petersburger Problems. In: *Die Werke von Jakob Bernoulli*, Band 3. Springer, Basel, (1975), 557-567.



## Capitolo 7

# Le origini della probabilità inversa

### 7.1 Il *Saggio* di Thomas Bayes

Nel 1763, sulle pagine delle *Philosophical Transactions* della Royal Society di Londra, apparve un saggio postumo di Thomas Bayes, morto il 17 aprile 1761 all'età di 59 anni: saggio destinato a lasciare un'impronta indelebile sull'evoluzione del calcolo delle probabilità e della statistica. Il lavoro [2] fu pubblicato grazie all'intervento di Richard Price (1723-1791) che gestì i manoscritti lasciati da Bayes ed aggiungendo diversi commenti all'originale. Il saggio—dal titolo: *An essay towards solving a problem in the Doctrine of Chance*—è formato da due sezioni principali: la prima consta di sette definizioni che servono a dimostrare altrettante proposizioni mentre la seconda sezione è occupata dalla soluzione del problema proposto da Bayes all'inizio del lavoro:

*Assegnato il numero di volte in cui un evento sconosciuto si è verificato o meno: Si richiede la possibilità (chance) che la probabilità (probability) che esso si realizzi in una prova singola giaccia tra due gradi di probabilità che si possono considerare.<sup>1</sup> ([2], p. 376)*

Nella lettera del 23 Dicembre 1763 in cui Price presentava il lavoro ad un altro socio della Royal Society, il fisico John Canton,<sup>2</sup> raccomandandone la pubblicazione sulle *Philosophical Transactions*, viene bene evidenziata la novità del punto di vista di Bayes, rispetto a quello adottato da Bernoulli e De Moivre:

Nelle sue *Laws of chance*, il sig. De Moivre, che ha molto migliorato questo ramo della matematica, ha fornito, sulla scorta di Bernoulli e con

---

<sup>1</sup> *Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.*

<sup>2</sup> Canton fu noto per i suoi studi sul magnetismo e per aver dimostrato sperimentalmente che l'acqua non è incomprimibile.

maggior grado di precisione, delle regole per trovare la probabilità che, se vengono ripetute un gran numero di prove riguardanti un qualche evento, la proporzione tra il numero di volte in cui esso si realizza ed il numero di volte in cui esso non si realizza in queste prove, differisce meno di limitazioni stringenti assegnate dal rapporto tra la probabilità di verificarsi e di non verificarsi in una singola prova. Non sono però a conoscenza di alcuno che abbia mostrato come dedurre la soluzione del problema inverso di questo; cioè, “assegnato il numero di volte in cui un evento sconosciuto si è verificato e quello in cui non si è verificato, trovare la possibilità che la probabilità che esso si realizzi si venga a trovare da qualche parte all’interno di due gradi di probabilità che si possono considerare.”<sup>3</sup> ([2], pp. 372-373)

Secondo Price, un elemento di merito del lavoro di Bayes è che viene meno la necessità di ritenere che il numero di prove tenda all’infinito, come richiesto per dare efficacia al teorema di Bernoulli-De Moivre:

Non si può ritenere che quanto fatto dal sig. De Moivre sia sufficiente a rendere l’esame di questo punto superfluo, soprattutto perché le regole da lui date non hanno la pretesa di essere esatte se non nell’ipotesi che il numero di prove sia infinito; da ciò risulta come non sia ovvio quanto debba essere grande il numero delle prove per far sì che le considerazioni siano sufficientemente esatte da farvi affidamento in pratica.<sup>4</sup> ([2], pp. 373)

Ancora, Price considerava un ulteriore elemento di pregio dell’approccio di Bayes la capacità di adattarsi a tentare di dirimere, se mai fosse possibile, un problema teologico formidabile:

dimostrare che motivo abbiamo per credere che nella costituzione delle cose siano iscritte delle leggi in base alle quali gli eventi si verificano e che dunque l’universo debba essere l’effetto della saggezza e della potenza di una causa intelligente; e confermare in questo modo l’argomento a

---

<sup>3</sup>Mr. De Moivre, indeed, the great improver of this part of mathematics, has in his *Laws of chance*, after Bernoulli, and to a greater degree of exactness, given rules to find the probability there is, that if a very great number of trials be made concerning any event, the proportion of the number of times it will happen, to the number of times it will fail in those trials, should differ less than by small assigned limits from the proportion of the probability of its happening to the probability of its failing in one single trial. But I know of no person who has shewn how to deduce the solution of the converse problem to this; namely, “the number of times an unknown event has happened and failed being given, to find the chance that the probability of its happening should lie somewhere between any two named degrees of probability”.

<sup>4</sup>What Mr. De Moivre has done therefore cannot be thought sufficient to make the consideration of this point unnecessary: especially, as the rules he has given are not pretended to be rigorously exact, except on supposition that the number of trials made are infinite; from whence it is not obvious how large the number of trials must be in order to make them exact enough to be depended on in practice.

favore dell'esistenza della Divinità a partire dalle cause finali.<sup>5</sup> ([2], pp. 373-374)

Vi sono evidenze piuttosto marcate [13] che suggeriscono come sia il saggio di Bayes che le aggiunte di Price fossero finalizzate a rispondere agli argomenti proposti nel 1748 dal filosofo scozzese David Hume (1711-1776), nel saggio *Enquiry concerning human understanding* dove figurava una limpida formulazione del problema dell'induzione, cioè del procedimento logico grazie al quale si passa dall'esame di casi particolari alla formulazione di un principio generale. In particolare, il Capitolo X dell'*Enquiry—Of Miracles*—attirò l'attenzione e fu bersaglio delle critiche di molti intellettuali per lo scetticismo nei confronti della religione che lo permeava. Il problema che Hume vi affrontava era, in generale, quello del valore da dare alla testimonianza degli uomini, che non può mai avere la stessa forza persuasiva dell'evidenza dei sensi. Hume era portato a non attribuire valore alle testimonianze di eventi miracolosi, perché distorte da diversi fattori quali: l'interesse di chi li narrava ad indurre l'ascoltatore od il lettore a credere nel miracolo narrato; il fascino esercitato dalla narrazione di eventi meravigliosi, in grado di catalizzare l'attenzione di una porzione non trascurabile della gente comune che diveniva preda di sfruttatori abili ad abusare della credulità popolare. Inoltre le testimonianze erano talora contraddittorie a seconda della fonte, segno di una manipolazione intenzionale mirata a far credere o a togliere credito ad un miracolo, a seconda dei sentimenti religiosi di chi redigeva il racconto. Ora, sia Bayes che Price erano esponenti del clero riformato non anglicano, uno dei bersagli preferiti dal sarcasmo di Hume nei confronti della religione. Come vedremo, soprattutto nell'appendice di Price all'*Essay* di Bayes, vi sono argomentazioni che puntano con decisione a rispondere alle tesi humiane. Nel breve Capitolo VI dell'*Enquiry—Probability*—dopo aver negato l'esistenza del *caso*, che è inteso come sinonimo della nostra ignoranza sulle vere cause alla base di una certa osservazione, Hume parlava di probabilità delle cause (*probability of causes*) in questi termini

Con la probabilità delle *cause* la situazione è la stessa che con la probabilità delle *eventualità*. Vi sono cause che sono del tutto uniformi e costanti nel produrre un certo tipo di effetto, senza che si sia mai verificata alcuna irregolarità o mancanza in questa operazione. Il fuoco ha sempre ustionato e l'acqua sempre soffocato tutte le creature umane. Che un urto o la gravità producano moto è una legge universale che sinora non ammette eccezioni. Succede però che vi siano altre cause che agiscono in modo più irregolare e dall'esito incerto: il rabarbaro non funziona sempre come purgante, né l'oppio come sedativo su tutti i pazienti che li assumono. È vero che, quando una causa non riesce a produrre il suo effetto consueto, gli scienziati non attribuiscono ciò a qualche irregolarità nella natura ma piuttosto suppongono che delle cause nascoste nella particolare struttura

---

<sup>5</sup>to shew what reason we have for believing that there are in the constitution of things fixt laws according to which events happen, and that, therefore, the frame of the world must be the effect of the wisdom and power of an intelligent cause; and thus to confirm the argument taken from final causes for the existence of the Deity.

delle parti abbia impedito l'attività. Tuttavia i nostri ragionamenti circa l'esito sono gli stessi che se questo principio delle "cause segrete" non si applicasse. L'abitudine ci ha convinti a trasferire il passato nel futuro in tutte le nostre inferenze e così, laddove il passato è stato del tutto regolare ed uniforme, noi ci aspettiamo l'esito consueto con la massima fiducia e non lasciamo alcuno spazio all'ipotesi contraria. Dove però si è trovato che effetti diversi sono seguiti a cause che appaiono del tutto simili, tutti questi effetti debbono essere presenti alla mente quando si sposta dal passato al futuro e debbono entrare nei nostri pensieri quando stimiamo la probabilità di un esito. Benché noi accordiamo il nostro favore a quello che si è visto essere il più comune e crediamo che questo effetto si verificherà anche la prossima volta, dobbiamo considerare anche gli altri effetti, assegnando loro un particolare peso ed autorità proporzionata alla loro maggiore o minore frequenza. In quasi tutti i paesi d'Europa è più probabile che ci sia gelo qualche volta a Gennaio piuttosto che il tempo si mantenga lontano dal gelo per tutto il mese, anche se questa probabilità varia a seconda dei vari climi e si avvicina ad essere certezza nelle terre più a nord. Sembra qui allora evidente che, quando trasferiamo il passato nel futuro per predire l'effetto che risulterà da una causa qualunque, noi trasferiamo tutti i diversi effetti nella stessa proporzione con cui si sono verificati nel passato e immaginiamo (ad esempio) che uno si sia verificato cento volte, un altro dieci ed un altro una sola volta. Quando un gran numero di opinioni si concentrano su un effetto lo rendono più forte e lo confermano all'immaginazione, generano quel sentimento che chiamiamo credenza (*belief*) e ci inducono a preferire quell'effetto all'effetto opposto che non è sostenuto da altrettante prove e non si palesa così spesso al nostro pensiero quando trasferiamo il passato nel futuro. Si provi a render conto di questa operazione della mente sulla base di uno qualsiasi dei sistemi filosofici esistenti e ci si renderà conto della difficoltà.<sup>6</sup>

---

<sup>6</sup>With the probability of *causes* the situation is the same as it is with the probability of *chance*. Some causes are entirely uniform and constant in producing a particular kind of effect, with no instance having ever been found of any failure or irregularity in their operation. Fire has always burned, and water has always suffocated, every human creature. The production of motion by impact and gravity is a universal law which up to now has had no exceptions. But other causes have been found to be more irregular and uncertain: rhubarb hasn't always worked as a purge, or opium as a soporific, on everyone who has taken these medicines. It is true that when any cause fails to produce its usual effect, scientists don't ascribe this to any irregularity in nature; but rather suppose that some secret causes in the particular structure of parts have prevented the operation. But our reasonings about the outcome are the same as if this principle concerning "secret causes" didn't apply. Custom has determined us to transfer the past to the future in all our inferences; so where the past has been entirely regular and uniform, we expect the usual outcome with the greatest confidence, and leave no room for any contrary supposition. But where different effects have been found to follow from causes that appear exactly alike, all these various effects must occur to the mind when it moves from the past to the future, and must enter into our thoughts when we estimate the probability of an outcome. Though we give preference to the one that has been found to be the most usual, and believe that this effect will occur this time too, we have to take into account the other effects, assigning to each a particular weight and authority in proportion as we have found it to be more or less frequent. In almost every country of Europe it is more probable that there will

Con la sua duttilità, l'approccio di Bayes permette di raccogliere la sfida lanciata da Hume in quanto

ci mostra, in modo distinto e preciso, in ogni caso nel quale si presenti un ordine particolare o una ripetizione di eventi, quale ragione vi sia di pensare che questa ripetizione o quest'ordine siano la conseguenza di cause stabili o regole della natura e non dell'irregolarità del caso.<sup>7</sup> ([2], p. 374)

Come accennavamo, la prima sezione dell'*Essay* di Bayes contiene sette definizioni da cui vengono dedotte altrettante proposizioni di calcolo delle probabilità. Le prime definizioni sono quelle di eventi incompatibili (*inconsistent*) e di evento complementare ad un certo evento. Soprattutto degna di nota è la definizione di probabilità

#### Definizione 5

La *probabilità di un qualsiasi evento* è il rapporto tra il valore a cui un'aspettazione subordinata al realizzarsi di un evento andrebbe calcolata ed il valore della cosa attesa se l'evento si realizza.<sup>8</sup> ([2], p. 376)

Essa è una piccola modifica della seconda definizione adottata da De Moivre e già illustrata nel Cap. IV. Scrivendo a John Canton, Price osservava che

Egli [Bayes] si è anche scusato per la singolare definizione data della parola caso o probabilità. La sua idea nel far questo era di stroncare ogni disputa circa il significato di una parola che, nel linguaggio comune è usata con significati differenti da persone di opinioni diverse, a seconda che si riferisca a fatti *passati* o *futuri*. Quali che siano i diversi significati che essa può avere, tutti (egli nota) concordano nel dire che un'aspettazione che dipenda dalla verità di un fatto *passato* o dal realizzarsi di un evento *futuro*, dovrebbe essere stimata di valore tanto maggiore quanto più è verosimile che il fatto sia vero o che l'evento accada. Pertanto, al posto del senso proprio della parola *probabilità*, egli [Bayes] gli ha confe-

---

be frost some time in January than that the weather will continue frost-free throughout that whole month; though this probability varies according to the different climates, and comes near to certainty in the more northern kingdoms. Here then it seems evident that when we transfer the past to the future in order to predict the effect that will result from any cause, we transfer all the different outcomes in the same proportion as they have appeared in the past, and conceive (for instance) one to have existed a hundred times, another ten times, and another once. As a great number of views here point to one outcome, they fortify and confirm it to the imagination, generate the sentiment that we call belief, and make us prefer that outcome to the contrary one that isn't supported by as many experiences and doesn't show up so frequently in our thought in transferring the past to the future. Try to account for this operation of the mind on the basis of any of the received systems of philosophy and you will become aware of the difficulty.

<sup>7</sup>shews us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any of the irregularities of chance.

<sup>8</sup>The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon it's happening.

rito quello che gli consentirà di esserne la misura propria in ogni caso in cui la parola venga usata.<sup>9</sup> ([2], p. 375)

Dunque ancora l'ambiguità del vocabolo di probabilità avrebbe spinto Bayes ad esprimerne l'idea ricorrendo al concetto di aspettazione di un evento. Si tratta ancora di un rapporto in cui il denominatore è il valore di ciò che mi aspetto di ricevere se si verifica l'evento, mentre il numeratore è il valore dell'aspettazione: se l'evento è certo, il valore dell'aspettazione coincide con quello di ciò che riceverò; se l'evento è impossibile, l'aspettazione ha valore nullo. Negli altri casi, il valore dell'aspettazione varia tra questi estremi. Non vi è motivo di dubitare delle intenzioni di Bayes, come riportate da Price, ma la definizione di probabilità proposta da Bayes—che ha una somiglianza con le considerazioni di de Witt cui si fece allusione nel Cap. IV—ha una particolare importanza perché non vi è nulla che le imponga di essere un numero *razionale*, potendo benissimo essere il rapporto tra numeri reali *qualsiasi*, vincolati a fornire per la probabilità un valore nell'intervallo  $[0, 1]$ . La necessità di avere alle spalle una nozione *continua* di probabilità è essenziale nel seguito del saggio di Bayes. Credo inoltre corretto il punto di vista di chi vede in questa definizione, una forma embrionale di approccio soggettivo alla probabilità [10]. Osserviamo infine che Bayes considerò i vocaboli *chance* e *probability* come sinonimi.

Egli introdusse la nozione di indipendenza tra due o più eventi in questi termini:

Eventi si dicono indipendenti quando il verificarsi di uno di questi non aumenti né riduca la probabilità di quel che rimane.<sup>10</sup> ([2], p. 376)

La definizione è sufficientemente vaga da non permettere di concludere se Bayes avesse consapevolezza della distinzione tra indipendenza a coppie o indipendenza tra eventi *tout-court*. Dal tenore della dimostrazione della successiva Prop. 6 sembrerebbe che tale distinzione sia assente, il che però non diminuisce il valore del saggio di Bayes.

OSSERVAZIONE. Ricordiamo in questo contesto un semplice esempio proposto da Sergej Bernštein (1880-1968) per chiarire la differenza tra indipendenza tra eventi a coppie o *tout-court*. Bernštein considerò un tetraedro equilibrato avente una faccia completamente colorata di rosso, una completamente gialla ed una completamente verde; la quarta ed ultima faccia era colorata a strisce rosse, gialle e verdi. Si considerino allora i tre eventi, legati al lancio del tetraedro

<sup>9</sup>He has also made an apology for the peculiar definition he has given of the word *chance* or *probability*. His design herein was to cut off all dispute about the meaning of the word, which in common language is used in different senses by persons of different opinions, and according as it is applied to *past* or *future* facts. But whatever different senses it may have, all (he observes) will allow that an expectation depending on the truth of any *past* fact, or the happening of any *future* event, ought to be estimated so much the more valuable as the fact is more likely to be true, or the event more likely to happen. Instead therefore, of the proper sense of the word *probability*, he has given that which all will allow to be its proper measure in every case where the word is used.

<sup>10</sup>Events are independent when the happening of any one of them does neither increase nor abate the probability of the rest.

- Evento  $A$ : esce<sup>11</sup> una faccia che ha del colore rosso;
- Evento  $B$ : esce una faccia che ha del colore giallo;
- Evento  $C$ : esce una faccia che ha del colore verde.

A ciascuno di questi eventi spetta una probabilità  $P(A) = P(B) = P(C) = \frac{1}{2}$ . Inoltre

$$P(A|B) = P(B|A) = P(A|C) = P(C|A) = P(B|C) = P(C|B) = \frac{1}{2}$$

per cui i tre eventi sono indipendenti a coppie. Tuttavia

$$P(A|B \cap C) = 1$$

e quindi non c'è indipendenza in generale (*cf.* [14], p. 59).

Vediamo ora all'opera la definizione di probabilità nella dimostrazione della prima proposizione dell'*Essay*:

PROP. 1

la probabilità che si verifichi uno o l'altro di più eventi incompatibili è la somma delle probabilità di ciascun evento.<sup>12</sup> ([2], p. 376)

Bayes considerò come illustrazione un caso di tre eventi incompatibili e chiamò  $N$  la cifra che si otterrebbe al verificarsi di uno qualunque degli eventi (*whichever of them happens*). Se  $\frac{a}{N}$ ,  $\frac{b}{N}$  e  $\frac{c}{N}$  sono le probabilità—nel senso della Def. 5—dei tre eventi, allora  $a$ ,  $b$  e  $c$  sono i valori delle aspettative associate al realizzarsi solo del primo, del secondo o del terzo evento. Il valore complessivo dell'aspettazione dovuta al realizzarsi di questi eventi è  $a + b + c$  e quindi la probabilità di ottenere uno o l'altro di essi è

$$\frac{a + b + c}{N} = \frac{a}{N} + \frac{b}{N} + \frac{c}{N}$$

cioè la somma delle probabilità relative a ciascuno di essi. Il corollario successivo specifica il fatto che, se si è *certi* del verificarsi di uno dei tre eventi, allora  $N = a + b + c$  perché

in questo caso, messe insieme tutte le aspettative, ammontano alla certezza di ricevere  $N$ , perciò la somma dei loro valori è uguale ad  $N$ .<sup>13</sup> ([2], p. 377)

Come immediata conseguenza di questo corollario, Bayes ottenne che, se un evento ha probabilità  $p$ , il suo complementare (*contrary*) ha probabilità  $1 - p$ . La seconda proposizione discende ancora direttamente dalla Def. 5 e dal corollario appena discusso:

<sup>11</sup>Trattandosi di un tetraedro, la faccia che "esce" è quella che, in seguito al lancio, è appoggiata al piano su cui è stato effettuato il lancio.

<sup>12</sup>When several events are inconsistent the probability of the happening of one or other of them is the sum of the probabilities of each of them.

<sup>13</sup>in this case all the expectations together amounting to a certain expectation of receiving  $N$ , their values together must be equal to  $N$ .

## PROP. 2

Se una persona ha un'aspettazione subordinata al realizzarsi di un evento, la probabilità di quell'evento sta a quella che l'evento non si realizzi come la perdita subita dalla persona se l'evento non si realizza sta al guadagno ottenuto se si realizza.<sup>14</sup> ([2], p. 377)

Infatti, dice Bayes, se un individuo riceverà  $N$  al realizzarsi di un evento di probabilità  $\frac{p}{N}$ , allora il valore di ciò che si aspetta è  $p$ , per la Def. 5. Se l'evento non si realizza, si perde  $p$  mentre se si realizza si ricava  $N$  ma l'aspettazione cessa e quindi il guadagno è  $N - p$  ed allora, siccome

$$\frac{p}{N} : \frac{N - p}{N} = p : N - p,$$

l'asserto è dimostrato. Parlare di  $N - p$  come guadagno sottintende l'accettazione della tradizionale definizione di gioco equo per cui il valore dell'aspettazione deve essere bilanciato dal versamento di una quota che garantisca il privilegio di poter ricavare la cifra  $N$ . Questo risultato viene subito impiegato per dimostrare la Proposizione 3, che equivale ad affermare che

$$p(A \cap B) = p(A)p(B|A) \quad (7.1)$$

per una coppia di eventi successivi  $A$  e  $B$ , di cui  $A$  è il primo:

## PROP. 3

la probabilità che due eventi successivi si verifichino entrambi è un rapporto composto dalla probabilità del I° evento e dalla probabilità del II° nell'ipotesi che si realizzi il I°.<sup>15</sup> ([2], p. 378)

Per la dimostrazione, Bayes suppose che, al realizzarsi di entrambi gli eventi, si riceverà  $N$  e che la probabilità che si realizzino entrambi sia  $\frac{p}{N}$ , quella che si realizzi solo il primo ( $A$ ) sia  $\frac{a}{N}$ , quella che si realizzi il secondo evento ( $B$ ), dato che si è realizzato  $A$  sia  $\frac{b}{N}$ . In virtù della Def. 5, il valore della aspettazione  $p$  diventa  $b$ , se  $A$  si è realizzato. Se dunque si realizza  $A$ , il guadagno è  $b - p$ , mentre se esso non si realizza, perderò  $p$ . Se si guarda al solo evento  $A$ , allora, per la Proposizione 2,

$$\frac{a}{N} : \frac{N - a}{N} = a : N - a$$

mentre, se si considera l'effetto del realizzarsi o meno di  $A$  sull'aspettazione legata ad  $A \cap B$ , la Proposizione 2 consente di concludere che

$$\frac{a}{N} : \frac{N - a}{N} = p : b - p$$

<sup>14</sup>If a person has an expectation depending on the happening of an event, the probability of the event is to the probability of its failure as his loss if it fails to his gain if it happens.

<sup>15</sup>The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2d on supposition the 1st happens.

da cui segue che

$$a : N - a = p : b - p \quad \text{e dunque} \quad a : N = p : b.$$

Siccome

$$\frac{p}{N} = \frac{p}{b} \frac{b}{N}$$

si conclude che

$$\frac{p}{N} = \frac{a}{N} \frac{b}{N}$$

che è quanto si voleva dimostrare. Il corollario che segue equivale a rileggere (7.1) nella forma

$$p(B|A) = \frac{p(A \cap B)}{p(A)}.$$

La Proposizione 4, in cui Bayes introduce  $p(A|B)$ , equivale a dire che

$$\frac{p(A \cap B)}{P(B)} = P(A|B) :$$

PROP. 4

Se ci sono due eventi successivi da determinare ogni giorno ed ogni giorno la probabilità del secondo è  $\frac{b}{N}$  e la probabilità che si realizzino entrambi è  $\frac{p}{N}$  e io riceverò  $N$  se entrambi gli eventi hanno luogo il primo giorno in cui ha luogo il secondo evento; affermo che, a queste condizioni, la probabilità che ho di ottenere  $N$  è  $\frac{p}{b}$ .<sup>16</sup> ([2], p. 380)

Una prima lettura di questa proposizione lascia perplessi per la scansione *temporale* dei due eventi, che chiameremo ancora  $A$ —il primo—e  $B$ , il secondo. Sembra corretto [10] intendere che per Bayes, l'evento  $C := A|B$ , cioè il realizzarsi di  $A$  sapendo che  $B$  si è verificato, può essere considerato in due casi: se  $B$  si realizza ad un certo giorno ed  $A$  si era realizzato lo stesso giorno, allora  $C$  si realizza; se  $B$  si verifica in un certo giorno, ma  $A$  non si era realizzato lo stesso giorno,  $C$  non si realizza; se  $B$  invece *non* si realizza ad un certo giorno, l'evento  $C$  non viene considerato e si passa al giorno successivo. Supponiamo che  $p(A|B) = \frac{x}{N}$ : allora, per la Def. 5,  $x$  è il valore dell'aspettazione di ottenere  $N$  se, sapendo che  $B$  si è verificato, si verifica anche  $A$  lo stesso giorno. Otterrò  $N$  in due casi tra loro incompatibili: se  $A$  e  $B$  si verificano in uno stesso giorno oppure se  $B$  non si verifica e la coincidenza degli eventi accade in un giorno successivo. Nel primo caso l'aspettazione è  $p$ , nel secondo caso, il gioco è, per così dire, congelato e l'aspettazione  $x$  resta sospesa. Bayes introduce una nuova quantità  $y$  tale che

$$y : x = N - b : N : \tag{7.2}$$

<sup>16</sup>If there be two subsequent events to be determined every day, and each day the probability of the 2d is  $\frac{b}{N}$  and the probability of both  $\frac{p}{N}$ , and I am to receive  $N$  if both the events happen the 1st day on which the 2d does; I say, according to these conditions, the probability of my obtaining  $N$  is  $\frac{p}{b}$ .

per la Prop. 2,  $y$  è il valore dell'aspettazione di ottenere  $x$  se  $B$  non si verifica ad un certo giorno. Se  $B$  non si realizza, mi posso aspettare di riottenere ad un giorno successivo  $x$  il cui valore, essendo legato alla probabilità che non si realizzi  $B$  è, per la (7.2), pari ad  $y$ . L'aspettazione  $x$  è così formata da due addendi legati al realizzarsi di eventi *incompatibili*:

- $p$ , il valore dell'aspettazione legata al verificarsi sia di  $A$  che di  $B$  il primo giorno in cui  $B$  si realizza
- $y$ , il valore dell'aspettazione legata al *non* verificarsi di  $B$  il primo giorno

e pertanto, conclude Bayes

$$p + y = x. \quad (7.3)$$

Ora, invertendo (7.2) ed applicando la proprietà dello scomporre, si ottiene

$$x : (x - y) = N : b, \quad \text{ovvero } x : p = N : b \quad (7.4)$$

da cui si ha

$$\frac{x}{N} = \frac{p}{b},$$

concludendo la dimostrazione. La discussione del corollario a questa proposizione conduce Bayes direttamente alla Prop. 5, il cui contenuto è una diversa formulazione di

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

COROLLARIO PROP. 5

Se ci sono due eventi successivi,  $\frac{b}{N}$  è la probabilità del secondo e  $\frac{p}{N}$  la probabilità che si realizzino entrambi, e si è scoperto che il secondo evento si è realizzato, e da ciò suppongo che anche il primo si sia realizzato, la probabilità che ho di essere nel giusto è  $\frac{p}{b}$ .<sup>17</sup> ([2], p. 381)

Come conseguenza diretta della definizione di eventi indipendenti e della Prop. 3 Bayes ricava la seguente

PROP. 6

la probabilità che più eventi indipendenti abbiano luogo è il rapporto composto dalle probabilità che si realizzino ciascun evento.<sup>18</sup> ([2], p. 382)

Questa proposizione viene applicata al caso in cui si consideri la probabilità che un evento si presenti al primo tentativo ma non al secondo ed al terzo e si ripresenti al quarto, ecc., le varie prove essendo tra loro indipendenti. La sezione 1 si conclude con la Proposizione 7 in cui si esplicita il significato combinatorio dei coefficienti dello sviluppo di  $(a + b)^n$  che abbiamo già visto più volte.

Inizia a questo punto la Sezione II dell'*Essay*, che si apre con la descrizione del celebre esempio della tavola:

<sup>17</sup>If there be two subsequent events, the probability of the 2d  $\frac{b}{N}$  and the probability of both together  $\frac{p}{N}$ , and it being 1st discovered that the 2d event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is  $\frac{p}{b}$ .

<sup>18</sup>The probability that several independent events shall all happen is a ratio compounded of the probabilities of each.

## POSTULATO I

Suppongo che una tavola quadrata, cioè un piano  $ABCD$ , sia costruita e levigata in modo tale che, se una delle sfere  $O$  o  $W$  viene lanciata su di essa, vi sarà la stessa probabilità che si fermi in una parte del piano come in un'altra e la sfera si fermerà necessariamente sulla tavola.<sup>19</sup> ([2], p. 385)

L'evento  $M$  che interessa a Bayes è descritto in questi termini:

## POSTULATO II

Suppongo che la sfera  $W$  sia lanciata per prima e che, passando per il punto in cui si arresta, sia tracciato un segmento  $os$  parallelo ad  $AD$ , che taglia  $CD$  in  $s$  e  $AB$  in  $o$ ; dopodiché, la pallina  $O$  viene lanciata per  $p+q$ , cioè  $n$ , volte e, quando in un lancio si ferma tra  $AD$  ed  $os$  diremo che l'evento  $M$  si è verificato in una prova.<sup>20</sup> ([2], p. 385)

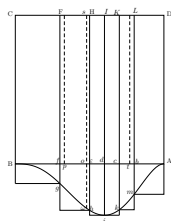


Figura 7.1: La *tavola* di Bayes

<sup>19</sup>I Suppose the square table or plane  $ABCD$  to be so made and levelled, that if either of the balls  $O$  or  $W$  be thrown upon it, there shall be the same probability that it rests upon any one equal part of the plane as another, and that it must necessarily rest somewhere upon it.

<sup>20</sup>I suppose that the ball  $W$  shall be 1st thrown, and through the point where it rests a line  $os$  shall be drawn parallel to  $AD$ , and meeting  $CD$  and  $AB$  in  $s$  and  $o$ ; and that afterwards the ball  $O$  shall be thrown  $p+q$  or  $n$  times, and that its resting between  $AD$  and  $os$  after a single throw be called the happening of the event  $M$  in a single trial.

Bayes dimostra un primo lemma in cui la probabilità che  $o$  appartenga ad una certa regione è caratterizzata *geometricamente*, come rapporto tra lunghezze di segmenti

## LEMMA I

La probabilità che il punto  $o$  cada tra due punti qualsiasi del segmento  $AB$  è il rapporto tra la distanza di questi punti e la lunghezza dell'intero segmento  $AB$ .<sup>21</sup> ([2], p. 385)

A partire da due punti  $f$  e  $b$  interni al segmento  $AB$ , si traccino i segmenti  $fF$  e  $bL$ , paralleli ad  $AD$ . Bayes distingue due casi:

- I rettangoli<sup>22</sup>  $Cf$ ,  $Fb$  ed  $LA$  sono commensurabili tra loro. Grazie a ciò è possibile suddividere tutte queste regioni in un numero intero di strisce di ugual larghezza, col che si può concludere che la probabilità che la prima palla  $W$  si fermi all'interno di una regione delimitata da un certo numero di queste strisce è la somma delle probabilità che essa si fermi su ciascuna di esse, trattandosi di eventi incompatibili. Poiché poi la probabilità che  $W$  si arresti entro una certa striscia non dipende dalla striscia scelta, per il I Postulato, la probabilità  $p_1$  di fermarsi all'interno di  $Fb$  è pari alla probabilità di arrestarsi in una delle strisce che la compongono per il numero di strisce che formano  $Fb$ . La probabilità  $q_1$  che  $W$  si fermi in una regione di  $AB \setminus Fb$  è data dal numero di parti in  $Cf$  ed  $LA$  e quindi

$$p_1 : q_1 = \# \text{ parti di } Fb : \# \text{ parti di } Cf \text{ e } LA$$

o, più semplicemente,

$$p_1 : q_1 = Fb : Cf \cup LA \quad \text{o ancora} \quad p_1 : q_1 = fb : Bf \cup Ab,$$

dove usiamo la stessa notazione per i segmenti e le loro lunghezze. Usando la proprietà del comporre ed osservando che  $p_1 + q_1 = 1$ , conseguenza diretta della Prop. 2, si ha

$$p_1 : 1 = fb : AB$$

e dunque la tesi.

- I rettangoli  $Cf$ ,  $Fb$  ed  $LA$  non sono commensurabili tra loro. In questo caso Bayes adotta un metodo di esaustione mostrando che  $p_1$  non può essere né minore né maggiore del rapporto  $\frac{fb}{AB}$ . Infatti, se fosse  $p_1 < \frac{fb}{AB}$ , potremmo porre<sup>23</sup>  $p_1 = \frac{fc}{AB}$  e considerare su  $fb$  due punti  $p$  e  $t$  tali che

$$pt > fc \tag{7.5}$$

<sup>21</sup>The probability that the point  $o$  will fall between any two points in the line  $AB$  is the ratio of the distance between the two points to the whole line  $AB$ .

<sup>22</sup>Secondo l'uso del tempo, Bayes indica un rettangolo o la sua area allo stesso modo, usando le lettere di due vertici *opposti*. Mi atterrò, nel seguito, a questo uso.

<sup>23</sup>Osserviamo che la continuità della probabilità è cruciale anche in questo punto.

e che i tre segmenti  $Bp$ ,  $pt$  e  $tA$  siano commensurabili tra loro: concretamente, basta bisecare  $AB$  più volte fino a trovare tra i punti di suddivisione due punti  $p$  e  $t$ , interni ad  $fb$  e tali che  $fp < ct$ . Dalla commensurabilità dei segmenti  $Bp$ ,  $pt$  e  $tA$  segue quella dei rettangoli  $Cp$ ,  $Dt$  e del rettangolo di base  $pt$  e altezza  $AD$ . Detta  $p_2$  la probabilità che  $o$  cada nella regione rettangolare della tavola individuata dai punti  $p$  e  $t$  si ha, per quanto visto al passo precedente

$$p_2 = \frac{pt}{AB}$$

e, per (7.5),

$$p_2 = \frac{pt}{AB} > \frac{fc}{AB} = p_1,$$

che è assurda, dal momento che  $pt$  è interno ad  $fb$ . In questo punto Bayes si serve di una proprietà di monotonìa della probabilità. Similmente si procede ad invalidare l'altra disuguaglianza, dimostrando il teorema.

Come conseguenza, la probabilità dell'evento  $M$  è

$$p(M) = \frac{Ao}{AB}. \quad (7.6)$$

La parte più tecnica del saggio di Bayes inizia dalla Proposizione 8.

PROP. 8

Se si costruisce su  $AB$  la curva  $BghikmA$  tale che, divisa la base  $AB$  in due parti  $Ab$  e  $Bb$  e condotta dal punto di divisione  $b$  la perpendicolare che taglia la curva in  $m$ ; detti  $y$ ,  $x$ ,  $r$ , rispettivamente, i rapporti di  $bm$ ,  $Ab$  e  $Bb$  con  $AB$ , e detto  $E$  il coefficiente del termine in cui compare  $a^p b^q$  quando si sviluppa il binomio  $(a + b)^{p+q}$ ,  $y = Ex^p r^q$ . Dico che, prima di lanciare la sfera  $W$ , la probabilità che  $o$  cada tra  $f$  e  $b$ , due punti qualsiasi del segmento  $AB$ , e che inoltre  $M$  si realizzi  $p$  volte e non si realizzi  $q$  volte in  $p + q$  prove, è il rapporto tra  $fghikmb$ , la parte di  $BghikmA$  delimitata dalle perpendicolari  $fg$  e  $bm$  al segmento  $AB$ , e il quadrato  $CA$  costruito su  $AB$ .<sup>24</sup> ([2], p. 388)

Assunta  $AB$  come asse delle ascisse, con origine in  $A$  e verso positivo da  $A$  a  $B$  e dette  $(x, y)$  le coordinate del punto  $m$ , normalizzate ad  $AB$ :  $x = \frac{Ab}{AB}$ ,  $y = \frac{bm}{AB}$ , mentre  $r = \frac{bB}{AB} = 1 - x$ , Bayes costruisce sotto la base  $AB$  il grafico della funzione

$$y(x) = E_{p,q} x^p r^q = E_{p,q} x^p (1 - x)^q,$$

<sup>24</sup>If upon  $BA$  you erect the figure  $BghikmA$  whose property is this, that (the base  $BA$  being divided into any two parts, as  $Ab$ , and  $Bb$  and at the point of division  $b$  a perpendicular being erected and terminated by the figure in  $m$ ; and  $y$ ,  $x$ ,  $r$  representing respectively the ratio of  $bm$ ,  $Ab$ , and  $Bb$  to  $AB$ , and  $E$  being the the coefficient of the term in which occurs  $a^p b^q$  when the binomial  $(a + b)^{p+q}$  is expanded)  $y = Ex^p r^q$ . I say that before the ball  $W$  is thrown, the probability the point  $o$  should fall between  $f$  and  $b$ , any two points named in the line  $AB$ , and withall that the event  $M$  should happen  $p$  times and fail  $q$  in  $p + q$  trials, is the ratio of  $fghikmb$ , the part of the figure  $BghikmA$  intercepted between the perpendiculars  $fg$ ,  $bm$  raised upon the line  $AB$ , to  $CA$  the square upon  $AB$ .

con

$$E = E_{p,q} := \binom{n}{p} = \binom{n}{q},$$

dove  $n := p + q$ . Per quanto visto prima,  $x$  ed  $r$  rappresentano le probabilità che  $W$  si fermi nel rettangolo  $LA$  ovvero nel rettangolo  $CL$  ed  $y(x)$  è la probabilità che  $O$  si fermi  $p$  volte a destra di  $bL$  e  $q$  volte si fermi a sinistra, cioè che l'evento  $M$  si realizzi  $p$  volte su  $n$  prove. Questo per quanto riguarda l'enunciato ma, nel corso della dimostrazione, occorre pensare ad  $x$  come all'ascissa normalizzata del punto mobile  $o$ :  $x = \frac{Ao}{AB}$ , mentre  $r = \frac{oB}{AB} = 1 - x$ . Indicheremo con  $E_1$  ed  $E_2$  i seguenti eventi

- $E_1$ : il punto  $o$  appartiene al segmento  $bf$ ;
- $E_2$ : in  $n = p + q$  prove, l'evento  $M$  si è verificato  $p$  volte e non si è verificato  $q$  volte:

Bayes intende trovare la probabilità di  $E_1 \cap E_2$ . La dimostrazione segue ancora il metodo di esaustione, supponendo dapprima che la probabilità cercata sia il rapporto  $\frac{D}{\text{area}(AC)}$  dove  $D$  rappresenta l'area di una certa figura ed è *maggiore* dell'area sottesa dal tratto di curva  $fghikmb$ . Presi i punti  $\varepsilon, d, c, \dots$  su  $AB$  e detti  $h, i, k, \dots$  i punti corrispondenti sulla curva  $y(x)$ , con  $d$  punto per il quale  $y(x)$  raggiunge il massimo assoluto  $i$ , è sempre possibile operare la scelta in modo che la somma delle aree dei rettangoli  $bk, ci, i\varepsilon, hf$  circoscritti alla porzione di piano  $fghikmb$ , sia *minore* dell'area  $D$ . Se il punto  $o$  coincidesse con  $\varepsilon$ , la probabilità che si realizzi l'evento  $M$  in un unico lancio sarebbe  $\frac{A\varepsilon}{AB}$ , per l'equazione (7.6), mentre  $\frac{B\varepsilon}{AB}$  sarebbe la probabilità dell'evento complementare. In questo caso, la probabilità che l'evento  $M$  si realizzi  $p$  volte e non si realizzi  $q$  volte è pari a  $E_{p,q}x^p r^q$  e coincide, per costruzione, con il rapporto  $\frac{\varepsilon h}{AB}$ . Similmente, se  $o$  fosse sovrapposto ad  $f$ , la probabilità che  $M$  si realizzi  $p$  volte e non si realizzi  $q$  volte, sarebbe pari al rapporto  $\frac{fg}{AB}$ . Poiché  $y(x)$  è monotona nel tratto  $fd$ , cui si suppone appartenga  $\varepsilon$ , la probabilità che  $M$  si realizzi  $p$  volte su  $n = p + q$  tentativi, sapendo che  $o \in \varepsilon f$ , è

$$P(E_2|E'_1) \leq \frac{\varepsilon h}{AB},$$

dove abbiamo indicato con  $E'_1$  l'evento: il punto  $o$  appartiene ad  $\varepsilon f$ . Poiché la probabilità che  $o$  stia tra  $\varepsilon$  ed  $f$  è  $\frac{\varepsilon f}{AB}$ , per la proposizione 3 la probabilità che  $o$  cada nel segmento  $\varepsilon f$  e che  $M$  abbia luogo  $p$  volte su  $n = p + q$  tentativi è

$$P(E'_1 \cap E_2) = P(E_2|E'_1)p(E'_1) \leq \frac{\varepsilon f}{AB} \frac{\varepsilon h}{AB} = \frac{\text{area}(fh)}{\text{area}(AC)}.$$

Similmente, indicando con  $E''_1$  l'evento: il punto  $o$  appartiene ad  $\varepsilon d$ , la probabilità che  $o$  cada nel segmento  $\varepsilon d$  e che  $M$  abbia luogo  $p$  volte su  $n = p + q$  è

$$P(E''_1 \cap E_2) \leq \frac{\text{area}(i\varepsilon)}{\text{area}(AC)}.$$

Ragionando in modo analogo se si ipotizza che  $o$  cada entro  $db$ , ed osservando che  $E_1 = E'_1 \cup E''_1 \cup \dots$  dove gli eventi  $E'_1, E''_1$  sono incompatibili, sommando  $P(E'_1 \cap E_2), P(E''_1 \cap E_2)$  si otterrà che  $P(E_1 \cap E_2)$  è inferiore al rapporto tra l'area del plurirettangolo circoscritto formato da  $fh, \varepsilon i, ci, \dots, bk$  e del quadrato  $AC$  che, a sua volta, non supera  $\frac{D}{\text{area}(AC)}$ , violando l'ipotesi da cui si era partiti. Se invece si considerano dei rettangoli *inscritti* nel grafico di  $y(x)$ , si dimostra che  $P(E_1 \cap E_2)$  non può essere il rapporto tra un'area minore di quella sottesa da  $fghikmb$  e quella del quadrato  $AC$ , concludendo la validità della tesi. A questa proposizione Bayes fa seguire un corollario, che equivale a prendere  $b = A$  ed  $f = B$  nel risultato precedente.

Corollario alla Prop. 8

Prima di lanciare la pallina  $W$ , la probabilità che il punto  $o$  giaccia da qualche parte tra  $A$  e  $B$ , ovvero da qualche parte lungo il segmento  $AB$ , e inoltre che l'evento  $M$  accada  $p$  volte e non accada  $q$  volte in  $p + q$  prove è il rapporto tra l'intera figura  $AiB$  e  $CA$ <sup>25</sup>

Rispetto alla proposizione, l'evento  $E_1$  è modificato in

$$\tilde{E}_1 : \quad o \in AB$$

e, siccome per la seconda parte del Postulato I,  $o$  appartiene sicuramente al segmento  $AB$ , dalla proposizione 8 si ha

$$P(\tilde{E}_1 \cap E_2) = P(\tilde{E}_1)P(E_2|\tilde{E}_1) = P(E_2|\tilde{E}_1) = P(E_2) = \frac{\text{area}(AiB)}{\text{area}(AC)}.$$

Analiticamente, la proposizione 8 ed il suo corollario si traducono nelle formule

$$\begin{aligned} P(E_1 \cap E_2) &= \frac{1}{\text{area}(AC)} \int_{x_b}^{x_f} E_{p,q} x^p (1-x)^q dx \\ P(E_2) &= \frac{1}{\text{area}(AC)} \int_0^1 E_{p,q} x^p (1-x)^q dx. \end{aligned} \tag{7.7}$$

Se la Prop. 8 era rivolta a determinare la probabilità del verificarsi di due eventi, la Prop. 9 mira a determinare una probabilità condizionata:

PROP. 9

Se, prima di scoprire qualunque cosa sulla posizione del punto  $o$ , si dovesse manifestare che l'evento  $M$  si è verificato  $p$  volte e non verificato  $q$  volte su  $p + q$  tentativi, e da ciò io traessi la conclusione (*I guess*) che il punto  $o$  giaccia tra due punti qualsiasi del segmento  $AB$ , come  $f$  e  $b$ , e di conseguenza che la probabilità dell'evento  $M$  in una singola prova sia stato compreso tra il rapporto di  $Ab$  con  $AB$  e quello di  $Af$  con  $AB$ : la probabilità che ho di essere nel vero è il rapporto della porzione della

<sup>25</sup>Before the ball  $W$  is thrown the probability that the point  $o$  will lie somewhere between  $A$  and  $B$ , or somewhere upon the line  $AB$ , and withal that the event  $M$  will happen  $p$  times, and fail  $q$  in  $p + q$  trials is the ratio of the whole figure  $AiB$  to  $CA$ .

figura  $AiB$  descritta in precedenza e che è racchiusa tra le perpendicolari condotte da  $AB$  nei punti  $f$  e  $b$  all'intera figura  $AiB$ .<sup>26</sup> ([2], pp. 391-392)

Infatti, grazie alle equazioni che figurano nella (7.7) ed alla proposizione 5 si conclude che

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{\text{area}(fghimb)}{\text{area}(AiB)} = \frac{\int_{x_f}^{x_b} E_{p,q} x^p (1-x)^q dx}{\int_0^1 E_{p,q} x^p (1-x)^q dx} = \frac{\int_{x_f}^{x_b} x^p (1-x)^q dx}{\int_0^1 x^p (1-x)^q dx} \quad (7.8)$$

visto che il fattore numerico  $E_{p,q}$  compare sia a numeratore che a denominatore. Come corollario, Bayes osserva che, ipotizzando che l'evento  $M$  abbia probabilità non superiore ad  $\frac{Ab}{AB}$ , si ha una probabilità di essere nel vero che è data da

$$\frac{\int_0^{x_b} x^p (1-x)^q dx}{\int_0^1 x^p (1-x)^q dx}.$$

Più che per la sua dimostrazione, la Prop. 9 è interessante per l'applicazione che Bayes ne fa al problema dell'induzione posto da Hume. Il teorema ottenuto misura il grado di attendibilità di una congettura elevata sull'evento  $E_1$  a partire dalla frequenza con cui l'evento  $M$  si è verificato. Notiamo che cruciale al riguardo è l'interpretazione dell'evento  $E_1$  come equivalente al realizzarsi di  $M$  in una prova *successiva*. È questa interpretazione che viene ancora utilizzata nello *scholium* seguente il cui obiettivo è di superare lo schema della tavola usato sin qui, esportandolo a situazioni più generali.

#### SCHOLIUM

Dalla Proposizione precedente è evidente che, nel caso in cui un evento come quello che ho chiamato  $M$ , di cui non conosco altro se non il numero di volte in cui esso si realizza o meno in un certo numero di prove, è possibile congetturare (*guess*) dove si trovi la sua probabilità e, con il metodo tradizionale di calcolare le aree cui ho fatto riferimento in precedenza, stimare la probabilità (*chance*) che la congettura sia corretta. E che questa regola sia quella appropriata da usare per un evento sulla cui probabilità non sappiamo assolutamente nulla prima che si facciano delle prove che lo riguardano, sembra risultare da questa considerazione; cioè che, circa un evento di quel tipo, non ho motivo di pensare che, in un dato numero di prove, esso debba verificarsi un certo numero di volte piuttosto che un altro. Perciò, sulla scorta di questo, posso correttamente ragionare a suo riguardo come se la sua probabilità non sia stata inizialmente fissata e sia stata determinata in seguito in modo tale da non darmi alcun

<sup>26</sup>If before any thing is discovered concerning the place of the point  $o$ , it should appear that the event  $M$  had happened  $p$  times and failed  $q$  in  $p + q$  trials, and from hence I guess that the point  $o$  lies between any two points in the line  $AB$ , as  $f$  and  $b$  and consequently that the probability of the event  $M$  in a single trial was somewhere between the ratio of  $Ab$  to  $AB$  and that of  $Af$  to  $AB$ : the probability I am in the right is the ratio of that part of the figure  $AiB$  described as before which is intercepted between perpendiculars erected upon  $AB$  at the points  $f$  and  $b$  to the whole figure  $AiB$ .

motivo di ritenere che, in un certo numero di prove, debba presentarsi un certo numero di volte piuttosto che un altro. Questo è proprio il caso dell'evento  $M$  perché, prima che la sferetta  $W$  sia lanciata e quindi risulti determinata la sua probabilità in una prova singola (per il corollario della proposizione 8), la probabilità che [l'evento  $M$ ] ha di realizzarsi  $p$  volte e di non realizzarsi  $q$  volte in  $p + q$  od  $n$  prove è il rapporto tra  $AiB$  e  $CA$ , rapporto che è lo stesso quando  $p + q$  od  $n$  è assegnato, qualunque sia il valore di  $p$ ; (...) e di conseguenza, prima che la posizione di  $o$  sia scoperta o [sia scoperto] il numero di prove in cui  $M$  si è verificato in  $n$  prove, non vi sono ragioni per pensare che debba verificarsi un numero di volte piuttosto che un altro.<sup>27</sup> ([2], pp. 392-393)

Dopo aver dimostrato, nella proposizione 10, un'altra conseguenza della proposizione 9, Bayes fornì un metodo per il calcolo dell'integrale definito

$$\int_{x_0}^{x_1} x^p(1-x)^q dx.$$

L'integrando viene sviluppato grazie al teorema del binomio di Newton e l'integrale è eseguito termine a termine. Il risultato di Bayes viene espresso mediante una regola che fornisce una buona approssimazione quando  $p$  o  $q$  hanno valori piccoli.

Nell'appendice al lavoro di Bayes, Price offrì diversi esempi di applicazione della formula (7.8), affrontando problemi generali e spinosi.

Supponiamo anzitutto che di un evento, come quello chiamato  $M$  nel saggio, sulla cui probabilità, prima di effettuare delle prove, non sappiamo nulla, se non che si è verificato una volta, e che si cerchi quali conclusioni possiamo trarre da ciò, in merito alla possibilità che esso si verifichi in una seconda prova. La risposta è che si può scommettere tre contro uno per

---

<sup>27</sup>From the preceding proposition it is plain, that in the case of such an event as I there call  $M$ , from the number of times it happens and fails in a certain number of trials, without knowing anything more concerning it, one may give a guess whereabouts it's probability is, and, by the usual methods computing the magnitudes of the areas there mentioned, see the chance that the guess is right. And that the same rule is the proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. For, on this account, I may justly reason concerning it as if its probability had been at first unfixed, and then determined in such a manner as to give me no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. But this is exactly the case of the event  $M$ . For before the ball  $W$  is thrown, which determines it's probability in a single trial (by cor. prop. 8), the probability it has to happen  $p$  times and fail  $q$  in  $p + q$  or  $n$  trials is the ratio of  $AiB$  to  $CA$ , which ratio is the same when  $p + q$  or  $n$  is given, whatever number  $p$  is; (...) And consequently before the place of the point  $o$  is discovered or the number of times the event  $M$  has happened in  $n$  trials, I can have no reason to think it should rather happen one possible number of times than another.

avere un po' più del 50% di probabilità che si verificherà in una seconda prova<sup>28</sup> ([2], p. 405)

In altri termini, sapendo che  $p = 1$ ,  $q = 0$ ,  $n = 1$  si ha una probabilità  $\frac{3}{4}$  di essere corretti nell'affermare che l'evento ha probabilità compresa nell'intervallo  $[\frac{1}{2}, 1]$ . Ponendo infatti  $x_b = \frac{1}{2}$  e  $x_f = 1$  nella formula (7.8) si ottiene

$$P(E_1|E_2) = P\left(\frac{1}{2} < x < 1|M\right) = 2 \left[\frac{x^2}{2}\right]_{\frac{1}{2}}^1 = \frac{3}{4}.$$

Un procedimento analogo permette di vedere che, se l'evento  $M$  si verifica per un certo numero  $n = p$  di volte senza che si verifichi in alcun caso l'evento complementare ( $q = 0$ ), la probabilità di essere corretti nel ritenere che la probabilità di  $M$  sia nell'intervallo  $[\frac{1}{2}, 1]$  è

$$P\left(\frac{1}{2} < x < 1|M \text{ avvenuto } n \text{ volte su } n\right) = \frac{2^{n+1} - 1}{2^{n+1}}$$

e dunque tende ad 1 piuttosto rapidamente. Price considera un altro caso

supponiamo che tutto ciò che si conosce di un evento sia che si è verificato dieci volte senza eccezione alcuna e che si ricerchi quale sia la ragione che abbiamo di ritenere di essere nel giusto se congetturiamo che la probabilità che esso ha di verificarsi in una singola prova giaccia nell'intervallo tra  $\frac{16}{17}$  e  $\frac{2}{3}$ , o che il rapporto delle cause del suo verificarsi stia a quello del suo non verificarsi sia compreso tra sedici contro uno o due contro uno.<sup>29</sup> ([2], pp. 405-406)

Si tratta di una lunga perifrasi per chiedere di calcolare

$$P\left(\frac{2}{3} < x < \frac{16}{17}|M \text{ avvenuto } 10 \text{ volte su } 10\right).$$

Presi  $q = 0$ ,  $p = n = 10$ ,  $x_b = \frac{2}{3}$  e  $x_f = \frac{16}{17}$  nella (7.8), si ottiene il valore 0.5013 e Price annotò, quasi volesse rispondere a Hume:

In questo modo, noi determiniamo in ogni caso quali conclusioni possiamo trarre da un certo numero di esperimenti le cui conclusioni non sono mai state smentite. Non c'è chi non veda in generale che si ha motivo di

<sup>28</sup>Let us then first suppose, of such an event as that called  $M$  in the essay, or an event about the probability of which, antecedently to trials, we know nothing, that it has happened once, and that it is enquired what conclusion we may draw from hence with respect to the probability of it's happening on a second trial. The answer is that there would be an odds of three to one for somewhat more than an even chance that it would happen on a second trial.

<sup>29</sup>suppose all I know of an event to be that it has happened ten times without failing, and the enquiry to be what reason we shall have to think we are right if we guess that the probability of it's happening in a single trial lies somewhere between  $\frac{16}{17}$  and  $\frac{2}{3}$ , or that the ratio of the causes of it's happening to those of it's failure is some ratio between that of sixteen to one and two to one.

attendarsi un evento con maggiore o minor fiducia a seconda del maggior o minor numero di volte in cui, in certe circostanze, esso si è verificato senza eccezioni; ora però noi comprendiamo esattamente quale sia questa ragione, su quali principii si basi e come dobbiamo regolare le nostre aspettative.<sup>30</sup> ([2], p. 406)

Esaurito l'esempio originale di Bayes, Price considera un nuovo problema:

Si supponga di non saper nulla sul numero di facce e sulla natura di un solido o di un dado; e che dobbiamo giudicare su ciò da esperimenti fatti lanciandolo.<sup>31</sup> ([2], p. 406)

Il primo lancio del dado non fornisce alcuna informazione perché:

In questo caso, bisogna osservare che sarebbe estremamente improbabile che il solido dovesse, al primo lancio, mostrare una faccia scelta in precedenza perché sarebbe noto che qualche faccia deve presentarsi e che c'era un'infinità di altre facce, o di facce segnate in qualche modo, che avevano le stesse possibilità di presentarsi. Il primo lancio mostra solo che il dado *ha* la faccia che si è mostrata, senza fornire alcun motivo di ritenere che *ne abbia* un certo numero piuttosto che un altro. Risulterà quindi che, *dopo* e non prima del primo lancio, noi ci troviamo nelle condizioni richieste dal nostro problema e che l'effetto complessivo di questo lancio sarà di portarci in questa situazione.<sup>32</sup> ([2], pp. 406-407)

A partire dal secondo lancio, iniziano i calcoli simili a quelli visti in precedenza e, se la stessa faccia compare sempre in una serie di lanci, cresce la probabilità che il dado abbia più facce del tipo di quella emersa nei lanci che di altro tipo, senza che però si raggiunga la certezza.

Non bisogna però immaginare che un numero qualsivoglia di esperimenti di questo tipo dia motivi sufficienti per pensare che non si presenterà *mai* un'altra faccia.<sup>33</sup> ([2], p. 407)

---

<sup>30</sup>In this manner we may determine in any case what conclusion we ought to draw from a given number of experiments which are unopposed by contrary experiments. Every one sees in general that there is reason to expect an event with more or less confidence according to the greater or less number of times in which, under given circumstances, it has happened without failing; but we here see exactly what this reason is, on what principles it is founded, and how we ought to regulate our expectations.

<sup>31</sup>Suppose a solid or die of whose number of sides and constitution we know nothing; and that we are to judge of these from experiments made in throwing it.

<sup>32</sup>In this case, it should be observed, that it would be in the highest degree improbable that the solid should, in the first trial, turn any one side which could be assigned beforehand; because it would be known that some side it must turn, and that there was an infinity of other sides, or sides otherwise marked, which it was equally likely that it should turn. The first throw only shews that *it has* the side then thrown, without giving any reason to think that *it has* it any one number of times rather than any other. It will appear, therefore, that *after* the first throw and not before, we should be in the circumstances required by the conditions of the present problem, and that the whole effect of this throw would be to bring us into these circumstances.

<sup>33</sup>It should not, however, be imagined that any number of such experiments can give sufficient reason for thinking that it would *never* turn any other side.

Price ripropone ora uno degli esempi di induzione proposti da Hume:

Ho fatto queste osservazioni soprattutto perché sono espressamente applicabili agli eventi che si presentano in natura. Prima di qualunque tipo di esperienza sarebbe improbabile come infinito contro uno che qualsiasi evento particolare, immaginato a priori, debba seguire dall'azione di un oggetto naturale su di un altro; ci sarebbero pari possibilità per qualsiasi altra infinità di eventi. Ma se per caso abbiamo visto una volta degli effetti particolari, come il fatto che il legno bruci quando messo sul fuoco, o che una pietra, staccatasi da una roccia, cada, allora le conclusioni che si possono trarre da un numero qualsiasi di eventi dello stesso tipo in successione, andrebbero determinate nello stesso modo delle conclusioni appena menzionate e che si riferiscono alla natura del solido che ho immaginato. In altre parole: il primo esperimento che si immagina essere fatto su un oggetto naturale ci informerebbe soltanto su un evento che può presentarsi in seguito ad una modifica particolare nelle circostanze di questi oggetti; non ci darebbe però alcuna idea di uniformità nella natura, né ci fornirebbe la minima motivazione per comprendere, in quel caso o in un altro qualsiasi, sulla regolarità o sull'irregolarità nelle sue operazioni. Se però lo stesso evento si è ripresentato senza eccezione in un numero qualsiasi di esperimenti successivi, allora si osserverà un certo grado di uniformità; ci saranno dei motivi per attendersi lo stesso esito in un ulteriore esperimento, ed i calcoli destinati alla soluzione di questo problema si possono svolgere.<sup>34</sup> ([2], pp. 408-409)

Price analizzò con i nuovi strumenti matematici a sua disposizione un esempio destinato ad essere riproposto molte volte in seguito e che era stato esposto [20] in una forma simile già nel 1754 nel *Traité des sensations* del filosofo francese Étienne Bonnot de Condillac (1714-1780): stimare la probabilità che il Sole sorga ancora un certo giorno, sapendo che è sorto in  $n$  giorni precedenti.

Immaginiamo il caso di una persona che, appena trasportata sulla Terra, sia lasciata a valutare dalle osservazioni sull'ordine ed il corso degli eventi,

---

<sup>34</sup>I have made these observations chiefly because they are all strictly applicable to the events and appearances of nature. Antecedently to all experience, it would be improbable as infinite to one, that any particular event, beforehand imagined, should follow the application of any one natural object to another; because there would be an equal chance for any one of an infinity of other events. But if we had once seen any particular effects, as the burning of wood on putting it into fire, or the falling of a stone on detaching it from all contiguous objects, then the conclusions to be drawn from any number of subsequent events of the same kind would be to be determined in the same manner with the conclusions just mentioned relating to the constitution of the solid I have supposed. In other words. The first experiment supposed to be ever made on any natural object would only inform us of one event that may follow a particular change in the circumstances of those objects; but it would not suggest to us any ideas of uniformity in nature, or give us the least reason to apprehend that it was, in that instance or in any other, regular rather than irregular in its operations. But if the same event has followed without interruption in any one or more subsequent experiments, then some degree of uniformity will be observed; reason will be given to expect the same success in further experiments, and the calculations directed by the solution of this problem may be made.

quali potenze e cause possano essere presenti. Il primo oggetto che attirerà la sua attenzione sarà forse il Sole; dopo averlo però perso di vista la prima notte, non saprà assolutamente se lo rivedrà o meno. Supponiamo che però egli veda il Sole una seconda volta, ovvero che il Sole sia ritornato una volta; si creerà in lui l'aspettazione di un secondo ritorno ed egli potrebbe sapere che si può scommettere tre contro uno per avere qualche probabilità che ciò accada. Come visto in precedenza, questa probabilità crescerà con il numero di ritorni di cui è stato testimone. Nessun numero finito di ritorni però sarà sufficiente per produrre la certezza assoluta o fisica del ritorno del Sole. Si supponga che lo abbia visto ritornare ad intervalli regolari e determinati per un milione di volte. Quanto gli è consentito di concludere da ciò sarebbe di questo tipo: si potrebbe scommettere la milionesima potenza di 2 contro 1 che il Sole tornerà nuovamente, al termine dell'intervallo consueto.<sup>35</sup> ([2], pp. 409-410)

Vi è un'osservazione molto importante aggiunta da Price, che invita alla cautela nell'uso di argomenti probabilistici svincolati da un sano spirito critico, una osservazione che non fu sempre tenuta presente nelle riproposizioni dell'esempio appena visto.

Bisogna ricordare con cura che queste deduzioni presuppongono una completa ignoranza pregressa della natura. Dopo aver osservato per qualche tempo il corso degli eventi, si troverebbe che le operazioni della natura sono in genere regolari e che le potenze e le leggi che vi prevalgono sono stabili e permanenti. La considerazione di ciò fa sì che uno o alcuni esperimenti spesso producano un'aspettazione più forte di successo negli esperimenti successivi di quanto non sarebbe altrimenti lecito attendersi; alla stessa stregua in cui, l'osservazione frequente che le cose di un certo tipo si dispongono insieme in un posto, ci porterebbe a concludere, avendovi scoperto un oggetto qualsiasi di quel tipo particolare, che ce ne sarebbero molti altri dello stesso tipo. È ovvio che questo, lungi dal contraddire le deduzioni precedenti, è solo un caso particolare in cui queste si possono applicare. Quanto detto ci pare sufficiente per mostrarci quali conseguenze trarre da un'esperienza *uniforme*. Ciò dimostra, in particolare, che invece di provare che gli eventi si realizzeranno *sempre* in accordo

---

<sup>35</sup>Let us imagine to ourselves the case of a person just brought forth into this world, and left to collect from his observation of the order and course of events what powers and causes take place in it. The Sun would, probably, be the first object that would engage his attention; but after losing it the first night he would be entirely ignorant whether he should ever see it again. He would therefore be in the condition of a person making a first experiment about an event entirely unknown to him. But let him see a second appearance or one return of the Sun, and an expectation would be raised in him of a second return, and he might know that there was an odds of 3 to 1 for some probability of this. This odds would increase, as before represented, with the number of returns to which he was witness. But no finite number of returns would be sufficient to produce absolute or physical certainty. For let it be supposed that he has seen it return at regular and stated intervals a million of times. The conclusions this would warrant would be such as follow. There would be the odds of the millionth power of 2, to one, that it was likely that it would return again at the end of the usual interval.

ad essa, c'è sempre motivo contro questa conclusione. In altre parole, laddove il corso della natura è stato costante in sommo grado, noi possiamo solo stimare il grado di questa costanza in proporzione alla ricorrenza di un evento; non avremo però alcun motivo di ritenere che non esistano cause in natura che non possano interferire *mai* con le cause da cui questa costanza è tratta o che non possano darsi circostanze in cui l'evento non si presenterà. E se ciò è vero, supponendo di aver tratto i *dati* solo dall'esperienza, troveremo motivi ulteriori per pensarla così se applichiamo altri principii o se ricorriamo a considerazioni che la ragione può suggerire, indipendentemente dall'esperienza.<sup>36</sup> ([2], pp. 410-411)

L'ultimo esempio riportato da Price riguarda la stima della probabilità che il rapporto tra i biglietti vincenti e quelli non vincenti in una lotteria possa essere desunto dalla proporzione osservata in un numero di estrazioni successive.

Immaginiamo una persona presente all'estrazione di una lotteria e che non conosca nulla della sua struttura o della proporzione dei [biglietti] Vuoti rispetto ai Premi. Si supponga ulteriormente di obbligare la persona ad inferire questa proporzione a partire dal numero di vuoti che ha sentito nell'estrazione, confrontato al numero dei premi; e supponiamo che sia richiesto quali conclusioni egli può ragionevolmente trarre in questa situazione. Supponiamo che abbia assistito all'estrazione di dieci biglietti vuoti e di un premio e che si richieda quale probabilità egli ha di essere nel vero se congettura che la proporzione dei [biglietti] vuoti rispetto ai pieni nella lotteria sia compresa tra 9 ad 1 ed 11 ad 1.<sup>37</sup> ([2], p. 411)

<sup>36</sup>It should be carefully remembered that these deductions suppose a previous total ignorance of nature. After having observed for some time the course of events it would be found that the operations of nature are in general regular, and that the powers and laws which prevail in it are stable and permanent. The consideration of this will cause one or a few experiments often to produce a much stronger expectation of success in further experiments than would otherwise have been reasonable; just as the frequent observation that things of a sort are disposed together in any place would lead us to conclude, upon discovering there any object of a particular sort, that there are laid up with it many others of the same sort. It is obvious that this, so far from contradicting the foregoing deductions, is only one particular case to which they are to be applied. What has been said seems sufficient to shew us what conclusions to draw from *uniform* experience. It demonstrates, particularly, that instead of proving that events will *always* happen agreeably to it, there will be always reason against this conclusion. In other words, where the course of nature has been the most constant, we can have only reason to reckon upon a recurrency of events proportioned to the degree of this constancy; but we can have no reason for thinking that there are no causes in nature which will *ever* interfere with the operations of the causes from which this constancy is derived, or no circumstances of the world in which it will fail. And if this is true, supposing our only *data* derived from experience, we shall find additional reason for thinking thus if we apply other principles, or have recourse to such considerations as reason, independently of experience, can suggest.

<sup>37</sup>Let us then imagine a person present at the drawing of a lottery, who knows nothing of its scheme or of the proportion of Blanks to Prizes in it. Let it further be supposed, that he is obliged to infer this from the number of blanks he hears drawn compared with the number of prizes; and that it is enquired what conclusions in these circumstances he may reasonably make. Let him first hear ten blanks drawn and one prize, and let it be enquired what chance he will have for being right if he guesses that the proportion of blanks to prizes in the lottery lies somewhere between the proportions of 9 to 1 and 11 to 1.

Si tratta di calcolare

$$P\left(\frac{9}{10} \leq x \leq \frac{11}{12} \mid p = 10, q = 1\right)$$

e quindi occorre applicare la formula (7.8) prendendo  $x_b = \frac{9}{10}$ ,  $x_f = \frac{11}{12}$ ,  $p = 10$ ,  $q = 1$ ,  $n = 11$ ,  $E_{10,1} = 11$ . Si ottiene come stima per la probabilità cercata il valore 0.07699:

Ci sarà dunque una probabilità di 923 contro 76, o circa di 12 contro 1, contro il fatto che sia corretto. Se avesse congetturato che vi fossero meno di 9 vuoti per ogni premio, avrebbe avuto una probabilità di essere nel giusto di .6589, ovvero di 65 contro 34.<sup>38</sup> ([2], p. 411)

Concludiamo questa esposizione, riassumendo i passi salienti dell'*Essay* di Bayes e delle aggiunte di Price.

- L'*Essay* sembra nascere più dal desiderio di fornire una risposta quantitativa al problema dell'induzione proposto da Hume che dall'invertire il teorema di Bernoulli-De Moivre;
- la definizione di probabilità adottata da Bayes appare piuttosto involuta ma ha il merito di renderne possibile l'*estensione* al continuo;
- la seconda parte dell'*Essay* che inizia con i Postulati I e II non è la ripetizione delle prime proposizioni. Il legame tra  $P(A \cap B)$ ,  $P(A|B)$ ,  $P(B|A)$ ,  $P(A)$  e  $P(B)$  era già espresso nelle proposizioni 3 e 4; con le proposizioni 7-9 Bayes dà una misura quantitativa delle probabilità coinvolte a partire dallo schema della tavola che, come si capisce procedendo nella lettura, nelle intenzioni di Bayes, è il prototipo dei problemi legati all'induzione;
- alcuni esempi inseriti da Price a conclusione del saggio di Bayes saranno ripresi in seguito senza le cautele necessarie e ciò creerà problemi alla credibilità del calcolo delle probabilità;
- la formula di Bayes, come la conosciamo oggi, *non* compare nell'*Essay* di Bayes.

## 7.2 Probabilità inversa nella *Mathématique sociale* di Condorcet

Abbiamo già fatto cenno, nel capitolo precedente, al progetto di *mathématique sociale* propugnato da Condorcet e sviluppato soprattutto nell'opera più originale e controversa, *l'Essai sur l'application de l'analyse à la probabilité des*

<sup>38</sup>There would therefore be an odds of about 923 to 76, or nearly 12 to 1 against his being right. Had he guessed only in general that there were less than 9 blanks to a prize, there would have been a probability of his being right equal to .6589, or the odds of 65 to 34. Notiamo che un'approssimazione corretta sarebbe stata di 66 contro 34.

*décisions rendues à la pluralité des voix* del 1785 [7]. Come suggerisce il titolo, in quest'opera Condorcet intendeva applicare il calcolo delle probabilità come guida per comprendere come comporre una giuria perché la probabilità di emettere un verdetto corretto in un processo fosse la più grande possibile e, più in generale, che un corpo di elettori prendesse la miglior decisione possibile o scegliesse il candidato migliore, pur con tutte le cautele con cui questo aggettivo può essere impiegato. Il significato più duraturo del lavoro di Condorcet è soprattutto *politico*, in quanto pose al centro dell'attenzione il problema fondamentale nella teoria della democrazia, quello di fornire una giustificazione alla prassi di attenersi alle decisioni prese dalla maggioranza. L'*Essai* è un'opera di ampio respiro che contiene un lungo *Discours Préliminaire* di quasi 200 pagine cui fa seguito il saggio vero e proprio, diviso in cinque parti. In particolare, l'obiettivo della terza parte viene descritto nel *Discours* in questi termini ([7], pp. lxxxix):

In questa terza parte ci proponiamo di fornire i mezzi 1° per stabilire, grazie all'osservazione, la probabilità della verità o falsità del voto di un uomo o della decisione di un tribunale; 2° per stabilire equamente, nei diversi tipi di questioni che si può essere chiamati a risolvere, la probabilità che si può considerare come idonea a conferire una sufficiente confidenza, cioè la più piccola probabilità della quale la giustizia o la prudenza possono permettere di accontentarsi.<sup>39</sup>

Il *Discours* contiene anche chiari riferimenti alla probabilità inversa ([7], p. lxxxiv):

Si vede infine che, per avere la probabilità di un evento futuro, sulla base della legge seguita dagli eventi passati, occorre prendere, 1° la probabilità di questo evento nell'ipotesi che la sua realizzazione sia soggetta a leggi costanti; 2° la probabilità dello stesso evento nel caso in cui la sua realizzazione non sia soggetta ad alcuna legge; moltiplicare ciascuna di queste probabilità per quella dell'ipotesi in virtù del quale è stata determinata, e dividere la somma dei prodotti per quello delle probabilità delle due ipotesi.<sup>40</sup>

Condorcet risolse una serie di problemi di probabilità inversa, alcuni dei quali molto articolati, che intendeva applicare per ([7], p. 213)

<sup>39</sup>Nous nous proposons dans cette troisième Partie de donner les moyens 1°. de déterminer par l'observation la probabilité de la vérité ou de la fausseté de la voix d'un homme ou de la décision d'un Tribunal; 2°. de déterminer également, pour les différentes espèces de questions qu'on peut avoir à résoudre, la probabilité que l'on peut regarder comme donnant une assurance suffisante, c'est-à-dire, la plus petite probabilité dont la justice ou la prudence puisse permettre de se contenter.

<sup>40</sup>On voit enfin que pour avoir la probabilité d'un événement futur, d'après la loi que suivent les événements passés, il faut prendre, 1° la probabilité de cet événement dans l'hypothèse que la production en est assujettie à des lois constantes; 2° la probabilité du même événement dans le cas où la production n'est assujettie à aucune loi; multiplier chacune de ces probabilités par celle de la supposition en vertu de laquelle on l'a déterminée, & diviser la somme des produits par celle des probabilités des deux hypothèses.

trovare degli strumenti per determinare, in seguito all'osservazione, il valore della probabilità del voto di un membro di un tribunale e della decisione di un certo tribunale.<sup>41</sup>

Un primo metodo di Condorcet richiede di conoscere un certo numero  $r$  di decisioni prese da un tribunale i cui componenti si suppone abbiano uguale capacità di discernimento (*égaux en lumières*) e di sapere con quale maggioranza siano state prese le decisioni; inoltre Condorcet suppone che le questioni da dirimere non siano tra quelle tali da suscitare pregiudizi o passioni. Infine egli ipotizza che i temi su cui il tribunale è stato chiamato a decidere siano di natura simile tra loro e tali che, esaminando dall'esterno la questione, sia possibile stabilire se un votante abbia preso la decisione corretta o meno. Condorcet propose diverse semplificazioni del metodo per poi passare ad illustrarne un secondo ([7], p. 220) in cui egli supponeva che la probabilità che un votante ha di pronunciarsi per il vero sia sempre ben al di sopra di  $\frac{1}{2}$ . Inoltre, Condorcet avanzò tre ipotesi:

1° che in ogni decisione il voto di ciascun votante abbia una probabilità costante; 2° che la probabilità sia variabile in ogni decisione e per ogni votante; 3° che si ammettano entrambe le ipotesi, moltiplicando la probabilità che ne consegue per quella che ciascuna ipotesi abbia luogo.<sup>42</sup>

Scartata la prima ipotesi come non realistica, Condorcet afferma che la seconda ipotesi conduce ai problemi IV e XII trattati nella parte terza dell'*Essai*:

#### Problema IV

Si supponga che un evento  $A$  si sia verificato  $m$  volte ed un evento  $N$ ,  $n$  volte; si domanda, sotto le tre ipotesi dei tre problemi precedenti,<sup>43</sup> 1° la probabilità che sia  $A$  o  $N$  l'evento cui compete probabilità da 1 ad  $\frac{1}{2}$  di accadere nella nuova prova; la probabilità che sia  $A$  o  $N$  a realizzarsi in una nuova prova; 3° la probabilità che si presenti un evento cui compete una probabilità tra 1 ed  $\frac{1}{2}$ .<sup>44</sup>

<sup>41</sup>trouver des moyens de déterminer, d'après l'observation, la valeur de la probabilité de la voix d'un des Votans d'un Tribunal & celle de la décision d'un Tribunal donné.

<sup>42</sup>1° que dans chaque décision la voix de tous les Votans ait une probabilité constante; 2° que dans chaque décision & chaque Votant, la probabilité varie; 3° qu'on admette ensemble les deux hypothèses, en multipliant la probabilité qui résulte de chacune par celle que cette hypothèse a lieu.

<sup>43</sup>Cioè che le probabilità che si realizzino  $A$  o  $N$  in una prova non dipendano dalla prova; che possano cambiare di prova in prova, prendendo un valore qualsiasi tra 0 ed 1, oppure di ignorare se queste probabilità siano o meno costanti

<sup>44</sup>On suppose ici un événement  $A$  arrivé  $m$  fois, & un événement  $N$  arrivé  $n$  fois; que l'on sache que la probabilité inconnue d'un des évènements soit depuis 1 jusqu'à  $\frac{1}{2}$ , & celle de l'autre depuis  $y$  jusqu'à zéro, & l'on demande, dans les trois hypothèses des trois problèmes précédens, 1° la probabilité que c'est  $A$  ou  $N$  dont la probabilité est depuis 1 jusqu'à  $\frac{1}{2}$ ; 2° la probabilité d'avoir  $A$  ou  $N$  dans le cas d'un nouvel événement; 3° la probabilité d'avoir un événement dont la probabilité soit depuis 1 jusqu'à  $\frac{1}{2}$ .

## Problema XII

Si supponga che la probabilità di uno degli eventi vari tra 1 ed  $\frac{1}{2}$  e quella dell'altro tra  $\frac{1}{2}$  e zero; si richiede, sotto questa ipotesi 1° La probabilità di ottenere  $A$   $q - q'$  volte in  $q$  prove ed  $N$   $q'$  volte; o che l'evento la cui probabilità è compresa tra 1 ed  $\frac{1}{2}$  si verificherà  $q - q'$  volte e quello la cui probabilità è tra  $\frac{1}{2}$  e zero,  $q'$  volte. 2° La probabilità che su  $2q + 1$  prove  $N$  non si realizzerà affatto  $2q' + 1$  volte più spesso di  $A$ ; 3° La probabilità che su  $2q + 1$  prove, l'evento  $A$  si verificherà  $2q' + 1$  volte più spesso di  $N$ , ovvero che la probabilità dell'evento la cui probabilità è compresa tra 1 ed  $\frac{1}{2}$ , arriverà  $2q' + 1$  volte più spesso di quello che ha probabilità compresa tra  $\frac{1}{2}$  e zero.<sup>45</sup>

La risoluzione di questi problemi si riconduce a valutare quozienti di integrali del tipo di quelli già calcolati da Bayes e Price. La novità è l'applicazione ed il punto delicato del lavoro di Condorcet è l'attendibilità, nel contesto giuridico proposto, delle ipotesi necessarie per rendere affrontabili matematicamente i problemi. Uno dei risultati più noti dell'opera di Condorcet è il cosiddetto *teorema della giuria* che apre la prima sezione dell'*Essai* ([7], p. 3):

Sia  $2q + 1$  il numero di votanti; si cerca la probabilità di avere una maggioranza di un sol voto.<sup>46</sup>

Condorcet suppone che ciascun votante, indipendentemente dagli altri, abbia probabilità  $v$  di votare “conformemente alla verità” ed  $e$  di votare per la decisione errata. Se si sviluppa  $(v + e)^{2q+1}$ , considerando i termini in cui l'esponente di  $v$  è *maggiore* di quello di  $e$ , la probabilità che il verdetto sia conforme al vero con una maggioranza qualsiasi è

$$\varphi(q) := v^{2q+1} + (2q+1)v^{2q}e + \binom{2q+1}{2}v^{2q-1}e^2 + \dots + \binom{2q+1}{q+1}v^{q+1}e^q.$$

In modo analogo, si trova  $\varphi(q+1)$ , la probabilità che il verdetto veritiero sia stato emesso quando i votanti sono  $2q + 3$ , con una maggioranza qualsiasi:

$$\varphi(q+1) := v^{2q+3} + (2q+3)v^{2q+2}e + \binom{2q+3}{2}v^{2q+1}e^2 + \dots + \binom{2q+3}{q+1}v^{q+2}e^{q+1}.$$

<sup>45</sup>On suppose que la probabilité d'un des événemens est depuis 1 jusqu'à  $\frac{1}{2}$ , & celle de l'autre depuis  $\frac{1}{2}$  jusqu'à zéro, & on demande dans cette hypothèse; 1° La probabilité que  $A$  arrivera  $q/q'$  fois dans  $q$  événemens,  $N$   $q'$  fois; ou que l'évènement dont la probabilité est depuis 1 jusqu'à  $\frac{1}{2}$ , arrivera  $q/q'$  fois, & celui dont la probabilité est depuis  $\frac{1}{2}$  jusqu'à zéro,  $q'$  fois. 2° La probabilité que sur  $2q + 1$  événemens,  $N$  n'arrivera point  $2q' + 1$  fois plus souvent que  $A$ ; ou que l'évènement dont la probabilité est depuis  $\frac{1}{2}$  jusqu'à zéro, n'arrivera pas  $2q' + 1$  fois plus souvent que l'évènement dont la probabilité est depuis 1 jusqu'à  $\frac{1}{2}$ . 3° La probabilité que sur  $2q + 1$  événemens, l'évènement  $A$  arrivera  $2q' + 1$  fois plus que  $N$ ; ou que l'évènement dont la probabilité est depuis 1 jusqu'à  $\frac{1}{2}$ , arrivera  $2q' + 1$  fois plus souvent que celui dont la probabilité est depuis  $\frac{1}{2}$  jusqu'à zéro.

<sup>46</sup>Le nombre des Votans est  $2q + 1$  & l'on cherche la probabilité de la pluralité d'une seule voix.

Ora, essendo  $v + e = 1$ , vale l'identità

$$\varphi(q+1) - \varphi(q) = \varphi(q+1) - (v+e)^2 \varphi(q)$$

che rende la differenza a sinistra uguale ad una differenza tra due polinomi in  $v$  di ugual grado. Servendosi di elementari identità sui coefficienti binomiali, si verifica che

$$\varphi(q+1) - \varphi(q) = \binom{2q+1}{q+1} (v-e)v^{q+1}e^{q+1}.$$

Se il procedimento viene ripetuto per valori inferiori di  $q$  ed i risultati ottenuti vengono sommati, si ottiene

$$\varphi(q) = v + (v-e) \sum_{k=1}^q \binom{2k-1}{k} v^k e^k$$

e la somma che figura in questa espressione è una ridotta nello sviluppo in serie di potenze di  $ve < 1$  della funzione

$$-\frac{1}{2} + \frac{1}{2} \frac{1}{\sqrt{1-4ve}}$$

per cui, supponendo  $q \gg 1$  possiamo scrivere

$$\varphi(q) \simeq v + (v-e) \left[ -\frac{1}{2} + \frac{1}{2} \frac{1}{\sqrt{1-4ve}} \right].$$

Ora, siccome

$$1 - 4ve = (v+e)^2 - 4ve = (v-e)^2,$$

se  $v > e$  ovvero se ciascun giurato ha più probabilità di prendere la decisione vera piuttosto che quella errata, avremo

$$v - e = \sqrt{1 - 4ve}$$

e quindi

$$\varphi(q) \simeq v + (v-e) \left[ -\frac{1}{2} + \frac{1}{2(v-e)} \right] = v + e = 1 :$$

si conclude dunque che, in queste condizioni, se la giuria è composta da un numero sufficientemente alto di membri, ciascuno dei quali ha probabilità  $v > \frac{1}{2}$  di pronunciarsi per la verità, la decisione del comitato avrà probabilità prossima ad 1 quanto si vuole di essere quella corretta. Al contrario, se  $v < e$ ,  $\varphi(q) \mapsto 0$  quando  $q \gg 1$ . Il teorema che, di per sé, è un bell'esercizio sulla distribuzione binomiale, richiede che i giudizi dei singoli votanti siano *indipendenti* uno dall'altro, un'ipotesi difficilmente sostenibile nella pratica e che attirò il giudizio negativo di molti studiosi, sia nell'accademia delle Scienze di Parigi, di cui Condorcet fu membro, che all'esterno. In particolare è rimasto celebre il giudizio del filosofo britannico John Stuart-Mill (1806-1873) che, nell'edizione del 1863 della sua *Logica* si riferiva a questo genere di risultati, come ad ([25], p. 356)

applicazioni fuorvianti del calcolo delle probabilità che lo hanno reso il vero obbrobrio della matematica.<sup>47</sup>

Sul finire dell'Ottocento, nella lunga introduzione al suo trattato sul calcolo delle probabilità ([3], p. xliii), Joseph Bertrand (1822-1900) scriverà a questo proposito

L'accusa è ingiusta. Si può pesare del rame e farlo passare per oro ma non si può incolpare la bilancia. Nei loro lavori sulla teoria dei giudizi, Condorcet, Laplace e Poisson non hanno fatto altro che pesare del rame.<sup>48</sup>

Il teorema della giuria di Condorcet fu relegato a lungo nell'oblio, per riemergere a metà del XX secolo quando, sotto opportune ipotesi restrittive, è divenuto un importante punto di riferimento della teoria delle funzioni di scelta sociale [4, 12].

Sempre nell'*Essai*, Condorcet si pose il problema di trovare il sistema elettorale più appropriato per fare in modo che le decisioni prese da un comitato o da un tribunale fossero conformi al vero. Ciascuno dei membri del tribunale è chiamato a pronunciarsi sulla verità o meno di una proposizione, che può essere semplice oppure spezzarsi in più proposizioni, su ciascuna delle quali occorre esprimere un giudizio. Quando è coinvolta una sola proposizione, se sia vera o falsa, conoscendo la probabilità che ciascun votante ha di scegliere correttamente, è possibile ottenere la probabilità che la decisione, presa a maggioranza, sia conforme al vero. Quando si hanno proposizioni composte, una decisione richiede che ci si pronunci sulla verità o meno di tutte le proposizioni semplici e le decisioni che si possono prendere ammontano a tutte le combinazioni che si possono formare prendendo, per ciascuna proposizione, una decisione a favore o contraria. In altre parole, se un giudizio richiede di pronunciarsi sulla verità o meno di  $n$  proposizioni, gli esiti possibili saranno  $2^n$ . Conoscendo la probabilità che un votante prenda, su ogni proposizione, una decisione conforme al vero, sarebbe possibile risalire alla probabilità che il giudizio finale sia veritiero o contenga un certo numero di errori. Condorcet osserva però come il metodo di prendere decisioni a maggioranza possa presentare delle difficoltà, se applicato a decisioni composte:

Occorre fare qui un'importante osservazione. È certamente possibile che l'opinione che ha la maggioranza dei voti non sia formata da proposizioni ciascuna delle quali abbia davvero la maggioranza e questa osservazione rende completamente difettoso il metodo di prendere la decisione in base alla maggioranza dei voti per ciascuna opinione e di determinarne la probabilità con il metodo precedente.<sup>49</sup>

<sup>47</sup>Misapplications of the calculus of probabilities which have made it the real opprobrium of mathematics.

<sup>48</sup>L'accusation est injuste. On peut peser du cuivre et le donner pour or, la balance reste sans reproche. Dans leurs travaux sur la théorie des jugements, Condorcet, Laplace et Poisson n'ont pesé que du cuivre.

<sup>49</sup>Mais il faut ici faire une observation importante. Il est très-possible que l'avis qui a la pluralité des voix, ne soit pas formé de propositions qui chacune aient réellement la pluralité,

Condorcet considera un'opinione (*avis*) che si forma combinando due proposizioni  $A$  ed  $a$  e le loro negazioni  $N$  ed  $n$ . Le opinioni possibili sono le coppie:  $(A, a)$ ,  $(A, n)$ ,  $(N, a)$  ed  $(N, n)$ . Supponiamo che vi siano 33 votanti che operano queste scelte:

1. 11 votano per  $(A, a)$ ;
2. 10 votano per  $(A, n)$ ;
3. 3 votano per  $(N, a)$ ;
4. 9 votano per  $(N, n)$ .

A maggioranza si adotta la mozione  $(A, a)$ . Condorcet osservò che i primi due tipi di elettori avrebbero adottato la proposizione  $A$  che avrebbe raggiunto 21 preferenze, risultando così preferita ad  $N$ . D'altra parte gli elettori di tipo 1 e 3 avrebbero adottato  $a$ , che avrebbe ricevuto 14 voti, risultando perdente contro  $n$ . L'opinione da adottare è allora  $(A, n)$ , diversamente da quanto deciso a maggioranza.

Poco dopo, Condorcet esaminò una elezione tra tre candidati, indicati con le lettere  $C_1$ ,  $C_2$  e  $C_3$ , osservando che, chi preferisce  $C_1$ , sottintende con la sua scelta che

1.  $C_1$  è migliore di  $C_2$  (in simboli moderni  $C_1 \succ C_2$ ): è la proposizione  $A$
2.  $C_1$  è migliore di  $C_3$  ( $C_1 \succ C_3$ ): è la proposizione  $a$ .

Similmente chi preferisce il candidato  $C_2$  sottintende che

3.  $C_2$  è migliore di  $C_1$  ( $C_2 \succ C_1$ ): è la proposizione  $N$
4.  $C_2$  è migliore di  $C_3$  ( $C_2 \succ C_3$ ): è la proposizione  $\alpha$ .

Infine, chi preferisce il candidato  $C_3$  ragiona dicendo che

5.  $C_3$  è migliore di  $C_1$  ( $C_3 \succ C_1$ ): è la proposizione  $n$
6.  $C_3$  è migliore di  $C_2$  ( $C_3 \succ C_2$ ): è la proposizione  $\nu$ .

Le opinioni da considerare sarebbero 8:

- I.  $(A, a, \alpha)$    II.  $(A, a, \nu)$    III.  $(A, n, \alpha)$    IV.  $(A, n, \nu)$   
 V.  $(N, a, \alpha)$    VI.  $(N, a, \nu)$    VII.  $(N, n, \alpha)$    VIII.  $(N, n, \nu)$ .

Condorcet osservò che le opinioni I e II comportano, rispettivamente, gli ordinamenti di preferenze  $C_1 \succ C_2 \succ C_3$  e  $C_1 \succ C_3 \succ C_2$  e sono associabili ad un voto per il candidato  $C_1$ ; le opinioni V e VII forniscono gli ordinamenti  $C_2 \succ C_1 \succ C_3$  e  $C_2 \succ C_3 \succ C_1$ , rispettivamente, e sono associabili ad un voto per il candidato  $C_2$ ; le opinioni IV ed VIII corrispondono ai sistemi di preferenze  $C_3 \succ C_1 \succ C_2$  e  $C_3 \succ C_2 \succ C_1$  per cui conducono a dare il voto a  $C_3$ . Le opinioni III e VI invece non conducono ad alcuna decisione perché sono tali che ([7], p. lvij)

---

& cette réflexion rend absolument défectueuse la manière de former la décision à la pluralité des voix pour chaque avis, & d'en déterminer la probabilité d'après la méthode précédente.

da due qualsiasi delle tre proposizioni che la compongono, scaturisce una conclusione contraria alla terza.<sup>50</sup>

Ad esempio, l'opinione III, contenendo voti per le proposizioni  $A$  ed  $n$ , porta a ritenere che, essendo sia  $C_1 \succ C_2$  e  $C_3 \succ C_1$ , per l'elettore dovrebbe valere  $C_3 \succ C_2$ , contraddetta dalla sua scelta in favore della proposizione  $\alpha$ .

Il motivo per cui l'elezione a maggioranza può tradire le reali preferenze del corpo elettorale è chiarito con un esempio in cui si considerano 60 votanti dei quali 23 hanno  $C_1$  come candidato preferito, 19 hanno  $C_2$  e 18 hanno  $C_3$ . Inoltre, Condorcet suppone che tutti i 23 votanti che preferiscono  $C_1$  siano concordi nel ritenere  $C_3 \succ C_2$ , che i 19 votanti per  $C_2$  concordino nel ritenere  $C_3 \succ C_1$  e infine che, tra i 18 votanti per  $C_3$ , 16 concordino nel ritenere  $C_2 \succ C_1$  e 2 nel ritenere  $C_1 \succ C_2$ . In definitiva, si ha la seguente ripartizione del corpo elettorale in  $6 = 3!$  sistemi transitivi di preferenze:

1.  $C_1 \succ C_2 \succ C_3$ , profilo condiviso da  $n_1 = 0$  elettori;
2.  $C_1 \succ C_3 \succ C_2$ , profilo condiviso da  $n_2 = 23$  elettori;
3.  $C_2 \succ C_1 \succ C_3$ , profilo condiviso da  $n_3 = 0$  elettori;
4.  $C_2 \succ C_3 \succ C_1$ , profilo condiviso da  $n_4 = 19$  elettori;
5.  $C_3 \succ C_1 \succ C_2$ , profilo condiviso da  $n_5 = 2$  elettori;
6.  $C_3 \succ C_2 \succ C_1$ , profilo condiviso da  $n_6 = 16$  elettori,

Pertanto  $n_4 + n_5 + n_6 = 37$  elettori optano per  $C_3 \succ C_1$ , mentre 23 per  $C_1 \succ C_3$ ;  $n_3 + n_4 + n_6 = 35$  elettori optano per  $C_2 \succ C_1$ , mentre 25 per  $C_1 \succ C_2$ ;  $n_1 + n_3 + n_4 = 19$  elettori optano per  $C_2 \succ C_3$  e 41 per  $C_3 \succ C_2$ . Dunque, mentre a maggioranza vincerebbe il candidato  $C_1$  con 23 voti,  $C_2$  sarebbe secondo con 19 voti e  $C_3$  terzo con 18 voti, nei confronti a coppie  $C_1$  perde sia con  $C_2$  che con  $C_3$  mentre  $C_3$ , che a maggioranza sarebbe giunto terzo, si aggiudica entrambi i confronti ristretti. La situazione peggiora se si suppone che i 23 votanti per  $C_1$  preferiscano  $C_2$  a  $C_3$ ; che tra i 19 votanti per  $C_2$  ve ne siano 17 per i quali  $C_3 \succ C_1$  e 2 per i quali  $C_1 \succ C_3$ ; che tra i 18 votanti per  $C_3$  ve ne siano 10 per i quali  $C_1 \succ C_2$  ed 8 per i quali  $C_2 \succ C_1$ . In questo caso, i profili elettorali sono ripartiti secondo lo schema:

1.  $C_1 \succ C_2 \succ C_3$ , profilo condiviso da  $n_1 = 23$  elettori;
2.  $C_1 \succ C_3 \succ C_2$ , profilo condiviso da  $n_2 = 0$  elettori;
3.  $C_2 \succ C_1 \succ C_3$ , profilo condiviso da  $n_3 = 2$  elettori;
4.  $C_2 \succ C_3 \succ C_1$ , profilo condiviso da  $n_4 = 17$  elettori;
5.  $C_3 \succ C_1 \succ C_2$ , profilo condiviso da  $n_5 = 10$  elettori;

<sup>50</sup>de deux quelconques des trois propositions qui la forment, résulte une conclusion contraire à la troisième.

6.  $C_3 \succ C_2 \succ C_1$ , profilo condiviso da  $n_6 = 8$  elettori.

In questo caso si verifica che la proposizione  $C_2 \succ C_3$  è preferita da 42 elettori su 60; la proposizione  $C_3 \succ C_1$  è preferita da 35 elettori su 60 e quella  $C_1 \succ C_2$  è preferita da 33 elettori mosrando così che, nel suo complesso, il corpo elettorale opta per un giudizio contraddittorio  $C_1 \succ C_2$ ,  $C_2 \succ C_3$  e  $C_3 \succ C_1$ .

Non è il caso in questa sede di soffermarsi sulle importanti conseguenze che questo paradosso ha avuto sulla teoria dei sistemi elettorali per le quali rimando alla letteratura specializzata [12]; vediamo invece come Condorcet intendeva servirsi del calcolo delle probabilità per stimare l'attinenza dell'esito di una votazione con la realtà delle cose. Consideriamo ancora un esempio numerico ([7], p. lxiii) con 60 votanti le cui preferenze sono così ripartite

1.  $C_1 \succ C_2 \succ C_3$ , profilo condiviso da  $n_1 = 10$  elettori;
2.  $C_1 \succ C_3 \succ C_2$ , profilo condiviso da  $n_2 = 13$  elettori;
3.  $C_2 \succ C_1 \succ C_3$ , profilo condiviso da  $n_3 = 6$  elettori;
4.  $C_2 \succ C_3 \succ C_1$ , profilo condiviso da  $n_4 = 13$  elettori;
5.  $C_3 \succ C_1 \succ C_2$ , profilo condiviso da  $n_5 = 0$  elettori;
6.  $C_3 \succ C_2 \succ C_1$ , profilo condiviso da  $n_6 = 18$  elettori.

Gli elettori per i quali  $C_2 \succ C_1$  sono 37 contro 23 per i quali  $C_1 \succ C_2$ ; gli elettori per cui  $C_3 \succ C_1$  sono 31 contro 29 per i quali  $C_1 \succ C_3$  e, similmente l'opzione  $C_3 \succ C_2$  prevale su  $C_2 \succ C_3$  per 31 contro 29 voti. Non vi sono cicli e  $C_3$  è, per Condorcet, la proposizione vincente e l'ordinamento emerso dai confronti a coppie è  $C_3 \succ C_2 \succ C_1$ . Tuttavia i successi di  $C_3$  avvengono con una maggioranza molto ristretta mentre nel successo su  $C_1$ ,  $C_2$  ha una maggioranza molto ampia. L'ordinamento ottenuto lascia ancora qualche dubbio e per essere più sicuri, Condorcet invoca il calcolo delle probabilità, cercando la verosimiglianza dei risultati ottenuti. Condorcet suppone che ogni elettore abbia la stessa probabilità  $v$  di essere nel vero ed  $e = 1 - v$  di errare in ogni proposizione. I ragionamenti di Condorcet sono piuttosto involuti e seguiamo la ricostruzione fornita da Young, due secoli dopo la pubblicazione dell'*Essai* [27]. Il candidato  $C_3$  è veramente il migliore? Se così fosse, la probabilità di ottenere un esito in cui  $C_3$  vince su  $C_1$  per 31 a 29 sarebbe

$$P(\text{voto} | C_3 \succ C_1) = \frac{60!}{31!29!} v^{31} e^{29}$$

perché i 31 elettori che lo hanno scelto, hanno optato per la decisione conforme al vero. Viceversa se  $C_3$  non fosse migliore di  $C_1$ , la probabilità di osservare lo stesso esito elettorale sarebbe

$$P(\text{voto} | C_1 \succ C_3) = \frac{60!}{31!29!} v^{29} e^{31}.$$

Supponendo *a priori* equiprobabile che  $C_1$  sia migliore di  $C_3$  o no, ovvero che  $P(C_1 \succ C_3) = P(C_3 \succ C_1) = \frac{1}{2}$ , la probabilità di avere l'esito ottenuto nel confronto tra  $C_1$  e  $C_3$  è

$$\begin{aligned} P(\text{voto}) &= P(\text{voto}|C_3 \succ C_1)P(C_3 \succ C_1) + P(\text{voto}|C_1 \succ C_3)P(C_1 \succ C_3) = \\ &= \frac{1}{2} \frac{60!}{31!29!} v^{29} e^{29} (v^2 + e^2). \end{aligned}$$

D'altra parte, per la formula di Bayes, abbiamo

$$P(C_3 \succ C_1 | \text{voto}) = \frac{P(\text{voto}|C_3 \succ C_1)P(C_3 \succ C_1)}{P(\text{voto})} = \frac{v^{31} e^{29}}{v^{29} e^{29} (v^2 + e^2)} = \frac{v^2}{v^2 + e^2},$$

che coincide con  $P(C_3 \succ C_2 | \text{voto})$ , dal momento che il confronto tra  $C_2$  e  $C_3$  conduce allo stesso esito di quello tra  $C_1$  e  $C_3$ . Dunque, la probabilità che  $C_3$  sia il miglior candidato, dato l'esito del voto è

$$P(C_3 | \text{voto}) = P(C_3 \succ C_1 | \text{voto})P(C_3 \succ C_2 | \text{voto}) = \frac{v^4}{(v^2 + e^2)^2}.$$

Ripetendo la stessa analisi per  $C_2$  e  $C_1$  si ricava

$$P(C_2 | \text{voto}) = P(C_2 \succ C_1 | \text{voto})P(C_2 \succ C_3 | \text{voto}) = \frac{v^{14}}{v^{14} + e^{14}} \times \frac{e^2}{v^2 + e^2}$$

e

$$P(C_1 | \text{voto}) = P(C_1 \succ C_2 | \text{voto})P(C_1 \succ C_3 | \text{voto}) = \frac{e^{14}}{v^{14} + e^{14}} \times \frac{e^2}{v^2 + e^2}.$$

Ora se  $v$  è prossimo ad 1 e quindi  $e \rightarrow 0$ , si vede che  $C_3$  è il candidato che, verosimilmente, è il migliore mentre qualora  $v = \frac{1}{2} + \varepsilon$  e quindi  $e = \frac{1}{2} - \varepsilon$  non è chiaro chi sia il migliore. Questa sensibilità rispetto al valore incognito di  $v$  può aver spinto Condorcet a cercare di battere altre strade ([27], p. 1238).

### 7.3 Laplace e la probabilità

Siamo così giunti al momento di introdurre Pierre Simon Laplace (1749-1827), figura di prima grandezza nella storia del calcolo delle probabilità, che fu uno scienziato molto influente nella Francia napoleonica. Egli fu introdotto negli ambienti scientifici parigini da D'Alembert, che gli procurò anche un incarico di insegnante presso una scuola militare e deve la sua fama scientifica soprattutto alla *Mécanique Celeste*, opera in cinque volumi pubblicati tra il 1799 ed il 1825. Molto importante fu il suo ruolo nello sviluppo della fisica francese. La sua opera principale di calcolo delle probabilità, la *Théorie analytique des probabilités*, la cui prima edizione apparve nel 1812 [19], divenne il testo di riferimento sulla disciplina per buona parte del XIX secolo. Alla *Théorie analytique* fu preme-  
so, a partire dall'edizione del 1814 un *Essai philosophique sur les probabilités*,

che era stato pubblicato separatamente<sup>51</sup> nel 1812 e che riassume il contenuto della *Théorie*, chiarendo l'impostazione concettuale di Laplace, i cui contributi alla teoria della probabilità abbracciano le diverse ramificazioni della disciplina. L'*Essai* trova la sua origine in un corso di matematiche elementari che Laplace aveva tenuto presso l'École Normale nel 1795. Alcuni passi di quest'opera sono diventati molto celebri, a partire dalla dichiarazione iniziale in cui si afferma che le più importanti questioni che si affrontano nella vita non sono che dei problemi di probabilità e che il probabile permea l'intera nostra conoscenza, che può dirsi certa solo nelle scienze matematiche. Il primo principio ad essere enunciato è il *principio di ragione sufficiente*: nulla può venire ad esistere senza una causa che lo produca. Poiché la conoscenza dell'uomo è imperfetta, i legami causali che legano tra loro gli eventi sfuggono ad un'indagine diretta e solo un'intelligenza perfetta che conoscesse ad un certo istante

tutte le forze che animano la natura e le condizioni rispettive di tutti gli esseri che la compongono e se fosse così estesa da sottomettere ad analisi tutti questi dati, sarebbe in grado di abbracciare in un'unica formula i movimenti dei corpi più grandi dell'universo come quelli dell'atomo più leggero: niente sarebbe incerto per questa intelligenza ed il futuro, come il passato, sarebbe presente ai suoi occhi.<sup>52</sup> ([19], p. ii)

Sussiste per Laplace una ineludibile componente “negativa” nella probabilità che rimarrà sempre legata in parte alla nostra ignoranza sul realizzarsi di un certo evento. L'obiettivo di ogni studio dei fenomeni casuali consiste

a ridurre ogni evento dello stesso genere ad un certo numero di casi ugualmente possibili, cioè, tali che noi siamo ugualmente indecisi in merito alla loro esistenza e a determinare il numero di casi favorevoli all'evento di cui si cerca la probabilità. Il rapporto di questo numero rispetto a quello di tutti i casi possibili è la misura di questa probabilità che non è altro che una frazione il cui numeratore è il numero di casi favorevoli mentre il denominatore è il numero di tutti i casi possibili.<sup>53</sup> ([19], p. iv)

Laplace insiste sulla necessità di scomporre un evento in porzioni equipossibili e sottolinea le cautele da usare quando ciò non sia fatto, ricollegandosi (II principio) all'esempio che D'Alembert utilizzava per far vedere la fragilità della probabilità matematica: qual è la probabilità di ottenere croce almeno una volta

<sup>51</sup>Nell'edizione del 1812 della *Théorie analytique* al posto dell'*Essai* vi era un'introduzione dedicata a Napoleone.

<sup>52</sup>toutes les forces dont la nature est animée, et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans le même formule, les mouvements des plus grands corps de l'univers et ceux du plus léger atome: rien ne serait incertain pour elle, et l'avenir comme le passé, serait présent à ses yeux.

<sup>53</sup>à réduire tous les événements du même genre, à un certain nombre de cas également possibles, c'est-à-dire, tels que nous soyons également indécis sur leur existence; et à déterminer le nombre des cas favorables à l'événement dont on cherche la probabilité. Le rapport de ce nombre à celui de tous les cas possibles, est la mesure de cette probabilité qui n'est ainsi qu'une fraction dont le numérateur est le nombre des cas favorables, et dont le dénominateur est le nombre de tous les cas possibles.

su due lanci di una moneta bilanciata. D'Alembert aveva concluso che i casi possibili erano tre: testa al primo lancio e croce al secondo; testa al primo ed al secondo lancio; croce al primo lancio. Verificatosi questo caso, per D'Alembert era inutile procedere e così concludeva che la probabilità di ottenere croce almeno una volta su due lanci fosse  $\frac{2}{3}$  e non  $\frac{3}{4}$ . Laplace sottolinea espressamente che all'evento: croce al primo lancio va associata una probabilità doppia rispetto alle altre eventualità per cui, pesando correttamente l'evento, si ritorna ad ottenere l'esatto valore della probabilità. Laplace fa seguire altri principii alla sua esposizione: dalla regola per trovare la probabilità del verificarsi di due eventi indipendenti (III principio) di cui si serve per mostrare come la veridicità di un fatto, tramandato successivamente da una persona ad un'altra, risulti alla fine poco attendibile; la regola della probabilità di eventi dipendenti (IV principio), che porta all'idea di probabilità condizionata al verificarsi di un altro evento, che conduce a sua volta alla formulazione verbale della formula di Bayes (Principii V-VIII). Laplace introdusse poi l'idea di speranza matematica, che fornì l'occasione di parlare del problema di S. Pietroburgo e dell'approccio di Daniel Bernoulli, basato sulla teoria dell'utilità (Principii IX e X).

Raccolti i dieci principii su cui si basa la teoria della probabilità, Laplace passa ad illustrare alcuni dei metodi analitici ritenuti più adatti allo studio delle probabilità: oltre al calcolo combinatorio, egli sottolinea l'importanza delle equazioni alle differenze, già applicate alla probabilità da Lagrange, e delle funzioni generatrici, cui egli diede un notevole impulso.

Infine, Laplace enumerò le principali applicazioni del calcolo delle probabilità: da quella tradizionale dei giochi d'azzardo alla probabilità che una testimonianza sia veritiera e che le decisioni prese da un'assemblea siano conformi al vero. Questi ultimi due ambiti sono particolarmente importanti nell'economia del pensiero di Laplace e riportano in primo piano da un lato il rapporto tra probabilità e questioni teologiche, dall'altro il legame tra probabilità e giurisprudenza. Inoltre, la probabilità investe le indagini di natura statistica, in particolare per lo studio della regolarità di certi rapporti, come quello del numero di neonati maschi rispetto alle neonate femmine. La teoria degli errori e le applicazioni alle rendite vitalizie contribuiscono a completare il panorama sterminato delle applicazioni del calcolo delle probabilità la cui centralità nel metodo scientifico appare in abbondanza.

La prima parte della *Théorie analytique* è dedicata all'esposizione degli strumenti matematici di maggior utilità nell'affrontare lo studio della probabilità ed è molto tecnica. Diamo solo un esempio di applicazione, relativo al problema della durata del gioco, vedendone la formalizzazione in termini di equazioni alle differenze e accennando al metodo di soluzione con le funzioni generatrici. Vi sono due giocatori,  $A$  e  $B$ , le cui probabilità di vincere una partita sono, rispettivamente,  $p$  e  $q$ . Ad ogni partita, il vincitore prende un gettone all'avversario; inizialmente  $A$  ha  $a$  gettoni, mentre  $B$  ne possiede  $b$ . Si vuole determinare la probabilità che  $A$  prenda tutti i gettoni di  $B$  al più in  $n$  partite. Laplace introdusse come incognita  $y_{x,x'}$ , la probabilità che  $A$  ha di aggiudicarsi il capitale di

$B$  quando egli possiede  $x$  gettoni e mancano  $x'$  partite per vincere il gioco.<sup>54</sup> Giocando un'ulteriore partita, questa incognita si cambierà in  $y_{x+1,x'-1}$  se  $A$  vince la partita o in  $y_{x-1,x'-1}$ , se è  $B$  a vincerla. Poiché il primo evento ha probabilità  $p$  ed il secondo ha probabilità  $q$ , la funzione  $y_{x,x'}$  verifica la seguente equazione alle differenze parziali

$$y_{x,x'} = py_{x+1,x'-1} + qy_{x-1,x'-1}. \quad (7.9)$$

La funzione generatrice è una funzione  $u(t, t')$  tale che il suo sviluppo in serie abbia  $y_{x,x'}$  come coefficiente di  $t^x t'^{x'}$ :

$$u(t, t') = \sum_{x,x'} y_{x,x'} t^x t'^{x'}.$$

Quindi  $y_{x+1,x'-1}$  ed  $y_{x-1,x'-1}$  sono i coefficienti di  $t^{x+1}t'^{x'-1}$  e di  $t^{x-1}t'^{x'-1}$ . L'equazione (7.9) si può allora rileggere come equazione per  $u(t, t')$ :

$$u(t, t') = u(t, t') \left( p \frac{t'}{t} + qt t' \right).$$

La soluzione di questa equazione ed il passaggio a ritroso per trovare  $y_{x,x'}$  occupano alcune fitte pagine della *Théorie analytique* e contengono passaggi di dubbio rigore. Una versione rigorosa del metodo di Laplace si può trovare, per esempio, nella monografia di Feller ([11], Cap. XIV. 3).

A conclusione di questa breve panoramica sull'impostazione di Laplace al calcolo delle probabilità, possiamo osservare, d'accordo con Sheynin [24], storico di questa disciplina, che Laplace non la studiò come un ramo della matematica, ma come una disciplina che si serve di strumenti matematici, anche molto raffinati, ma che non ha le caratteristiche di astrazione necessaria a portarla sullo stesso piano di altri rami della matematica. D'altronde Poisson, che fu allievo di Laplace, ne descriveva il lavoro in questi termini (cfr. [24], p. 179)

l'analisi era uno strumento che [Laplace] piegava alle più svariate applicazioni ma sempre subordinando il metodo speciale alla base stessa di ogni questione.<sup>55</sup>

## 7.4 Laplace e la probabilità delle cause

Una decina di anni dopo la comparsa del saggio di Bayes, nel 1774 Laplace pubblicò un lavoro dal titolo: *Sur la probabilité des causes par les événements* dove, pur ignaro di quanto fatto da Bayes, si pose problemi di natura simile. Laplace era pienamente consapevole delle potenzialità di questi studi che, ai suoi occhi, permettevano alla probabilità di allargare molto il proprio raggio di azione:

<sup>54</sup>Ricordiamo che il gioco è composto da  $n$  partite al più.

<sup>55</sup>l'analyse mathématique était un instrument qu'il pliait aux applications les plus variées, mais toujours en subordonnant la méthode speciale au fonds même de chaque question.

mi propongo di determinare la probabilità delle cause attraverso gli eventi, argomento nuovo per molti aspetti e che merita tanto più di essere coltivato in quanto è soprattutto da questo punto di vista che la scienza degli azzardi può essere utile nella vita civile.<sup>56</sup> ([16], p. 28)

Il prototipo dei problemi da trattare è quello di stabilire, data un'urna contenente biglietti bianchi e biglietti neri in un rapporto sconosciuto e sapendo di avere estratto un biglietto bianco, la probabilità che questo rapporto abbia il valore  $\frac{p}{q}$ . Il metodo di Laplace si basa su un principio:

PRINCIPIO. Se un evento può essere prodotto da un numero  $n$  di cause distinte, le probabilità che queste cause esistano, dato che l'evento si è realizzato, stanno tra loro come le probabilità dell'evento, dato che si siano verificate queste cause e la probabilità di ciascuna di esse è uguale alla probabilità dell'evento dato il verificarsi di questa causa, divisa per la somma di tutte le probabilità dell'evento, data che si sia realizzata ciascuna di queste cause.<sup>57</sup> ([16], p. 29)

Laplace non spende una parola per commentare questo Principio ma passa subito agli esempi che dovrebbero chiarirne il significato. Se  $\{A_1, A_2, \dots, A_n\}$  è l'insieme delle cause tra loro incompatibili dell'evento  $E$ , le due parti del Principio di Laplace si traducono in

$$\frac{P(A_i|E)}{P(A_j|E)} = \frac{P(E|A_i)}{P(E|A_j)} \quad (7.10)$$

e in

$$P(A_i|E) = \frac{P(E|A_i)}{\sum_{k=1}^n P(E|A_k)}. \quad (7.11)$$

Ora, se applichiamo le formule già usate da Bayes

$$\begin{aligned} P(A_i|E)P(E) &= P(A_i \cap E) & P(A_j|E)P(E) &= P(A_j \cap E) \\ P(E|A_i)P(A_i) &= P(A_i \cap E) & P(E|A_j)P(A_j) &= P(A_j \cap E), \end{aligned} \quad (7.12)$$

otteniamo

$$\frac{P(A_i|E)}{P(A_j|E)} = \frac{P(E|A_i) P(A_i)}{P(E|A_j) P(A_j)} \quad (7.13)$$

che coincide con (7.10) solo quando tutte le cause hanno la stessa probabilità, ovvero se è verificata l'ipotesi di uniformità a priori delle cause. Similmente, (7.11) coincide con la consueta *formula di Bayes* (cfr. p. es. [14], p. 62)

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{k=1}^n P(E|A_k)P(A_k)} \quad (7.14)$$

<sup>56</sup>Je me propose de déterminer la probabilité des causes par les événements, matière neuve à bien des égards et qui mérite d'autant plus d'être cultivée que c'est principalement sous ce point de vue que la science des hasards peut être utile dans la vie civile.

<sup>57</sup>PRINCIPE. Si un événement peut être produit par un nombre  $n$  de causes différentes, les probabilités de l'existence de ces causes prises de l'événement sont entre elles comme les probabilités de l'événement prises de ces causes, et la probabilité de l'existence de chacune d'elles est égale à la probabilité de l'événement prise de cette cause, divisée par la somme de toutes les probabilités de l'événement prises de chacune de ces causes.

se le cause hanno tutte la stessa probabilità a priori.

Qualche anno dopo, Laplace ritornò su questo principio dimostrandolo nell'articolo XV della lunga memoria *Sur les probabilités* [17], pubblicata nel 1781. Laplace considerò un evento  $E_1$  che può risultare solo dall'azione di  $n$  cause incompatibili ed esaustive<sup>58</sup>  $A_1, A_2, \dots, A_n$  ed introdusse le probabilità<sup>59</sup>

$$x_i = P(A_i|E_1) \quad \text{e} \quad a_i = P(E_1|A_i). \quad (7.15)$$

Laplace considerò un secondo evento  $E_2$ , simile al primo (*semblable au premier*)  $E_1$ , e scrisse

$$P(E_2|E_1) = \sum_{i=1}^n P(E_2|A_i)P(A_i|E_1) = \sum_{i=1}^n a_i x_i; \quad (7.16)$$

la probabilità che due eventi  $E_2$  ed  $E_3$ , entrambi simili al primo, si realizzino quando questo si realizza venne da lui scritta come

$$P(E_2 \cap E_3|E_1) = \sum_{i=1}^n a_i^2 x_i \quad (7.17)$$

e così via di seguito. Una seconda espressione della probabilità che  $E_1$  si realizzi è data da

$$P(E_1) = \sum_{k=1}^n P(E_1|A_k)P(A_k) = \frac{1}{n} \sum_{k=1}^n a_k, \quad (7.18)$$

dove Laplace implicitamente ha usato l'ipotesi di uniformità a priori delle cause

$$P(A_1) = P(A_2) = \dots = P(A_n) = \frac{1}{n}. \quad (7.19)$$

Similmente, le probabilità che due, tre, o più eventi simili al primo si verifichino sono scritte, rispettivamente, come

$$P(E_1 \cap E_2) = \frac{1}{n} \sum_{i=1}^n a_i^2 \quad P(E_1 \cap E_2 \cap E_3) = \frac{1}{n} \sum_{i=1}^n a_i^3, \dots \quad (7.20)$$

Pertanto, grazie all'equazione (7.18), le probabilità che, essendosi verificato il primo evento  $E_1$  se ne verifichi un secondo, un terzo, e così via, tutti simili al primo, sono scritte da Laplace come

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n a_i} \quad (7.21)$$

<sup>58</sup> $A_i \cap A_j = \emptyset, \forall i, j = 1, \dots, n, (i \neq j)$ ;  $P(E_1) = \sum_{k=1}^n P(E_1|A_k)P(A_k)$ .

<sup>59</sup>La notazione per le probabilità condizionate è abbastanza recente. I primi lavori in cui si sentì il bisogno di distinguere con una notazione propria questo concetto risalgono all'ultimo ventennio del XIX secolo: nel 1880 Hugh Mac Coll (1837-1909) utilizzò  $A_B$  per indicare che  $A$  è vero se  $B$  lo è; Felix Hausdorff utilizzò  $P_B(A)$  in un lavoro del 1901 e la sua notazione fu accolta nel manuale di Czuber, pubblicato poco dopo [8], e si trova anche in lavori di Kolmogorov e De Finetti. Markov propose  $P(A, B)$  nel 1900, notazione che, come la precedente, sopravvisse abbastanza a lungo nell'uso. Keynes propose  $P\left(\frac{A}{B}\right)$  nel suo manuale del 1921 [15]. La notazione attuale sembra essere dovuta ad Harold Jeffreys che ne fece uso a partire dal 1931, dopo avere usato la notazione  $P(A; B)$ .

$$P(E_2 \cap E_3 | E_1) = \frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1)} = \frac{\sum_{i=1}^n a_i^3}{\sum_{i=1}^n a_i} \dots \quad (7.22)$$

Per confronto con le espressioni trovate in precedenza, Laplace forma un sistema di  $n$  equazioni nelle incognite  $x_i$ :

$$\left\{ \begin{array}{l} \sum_{i=1}^n a_i x_i = \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n a_i} \\ \sum_{i=1}^n a_i^2 x_i = \frac{\sum_{i=1}^n a_i^3}{\sum_{i=1}^n a_i} \\ \sum_{i=1}^n a_i^{n-1} x_i = \frac{\sum_{i=1}^n a_i^n}{\sum_{i=1}^n a_i} \\ \sum_{i=1}^n x_i = 1 \end{array} \right.$$

che si verificano essere soddisfatte da

$$x_i = P(A_i | E) = \frac{P(E | A_i)}{\sum_{k=1}^n P(E | A_k)} = \frac{a_i}{\sum_{k=1}^n a_k}$$

che coincide con (7.11). Questa deduzione è lacunosa nelle giustificazioni offerte a supporto dei vari passaggi e possiamo ricostruirla in questo modo ([9], pp. 281-283). Anzitutto

$$P(E_2 | E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\sum_{i=1}^n P(E_1 \cap E_2 | A_i) P(A_i)}{\sum_{i=1}^n P(E_1 | A_i) P(A_i)} \quad (7.23)$$

che diventa, nell'ipotesi di uniformità delle cause (7.19),

$$P(E_2 | E_1) = \frac{\sum_{i=1}^n P(E_1 \cap E_2 | A_i)}{\sum_{i=1}^n P(E_1 | A_i)} = \frac{\sum_{i=1}^n P(E_1 \cap E_2 | A_i)}{\sum_{i=1}^n a_i}.$$

Per procedere, occorre ora supporre che gli eventi  $E_2$  ed  $E_1$  siano *condizionatamente* indipendenti, rispetto alle cause  $A_i$ , cioè che valga

$$P(E_1 \cap E_2 | A_i) = P(E_1 | A_i) P(E_2 | A_i) \quad \forall i = 1, \dots, n. \quad (7.24)$$

Sotto questa ipotesi avremo ancora

$$P(E_2 | E_1) = \frac{\sum_{i=1}^n P(E_1 | A_i) P(E_2 | A_i)}{\sum_{i=1}^n a_i} = \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n a_i}$$

dove, nell'ultimo passaggio, si è interpretata la condizione che gli eventi  $E_1$  ed  $E_2$  siano *simili* come  $P(E_1 | A_i) = P(E_2 | A_i)$ , quale che sia la causa  $A_i$  considerata. D'altra parte, ripartendo da (7.23) ed utilizzando nuovamente l'indipendenza condizionata di  $E_1$  ed  $E_2$ , abbiamo

$$\begin{aligned} P(E_2 | E_1) &= \frac{\sum_{i=1}^n P(E_1 | A_i) P(E_2 | A_i) P(A_i)}{P(E_1)} = \frac{\sum_{i=1}^n P(E_1 \cap A_i) P(E_2 | A_i)}{P(E_1)} = \\ &= \sum_{i=1}^n P(A_i | E_1) P(E_2 | A_i) = \sum_{i=1}^n a_i x_i : \end{aligned} \quad (7.25)$$

uguagliando le espressioni ottenute per  $P(E_2|E_1)$  si ricava la prima delle equazioni di Laplace. Per procedere con gli altri eventi simili ad  $E_1$ , occorre estendere via via l'ipotesi di indipendenza condizionata, coinvolgendo un numero maggiore di eventi.

L'ipotesi di *uniformità* delle cause, già utilizzata tacitamente più volte, venne formulata all'inizio dell'art. XVII di [17]:

Quando non vi è alcun dato *a priori* sulla possibilità di un evento, bisogna ipotizzare come ugualmente probabili tutte le possibilità, da zero fino all'unità.<sup>60</sup> ([17], p. 419)

Alcuni anni più tardi, Laplace ritornò sull'ipotesi di uniformità delle cause. Sia  $x \in [0, 1]$  la probabilità a priori di una causa  $A$  ed  $y(x)$  la probabilità del realizzarsi dell'evento  $E$  che dipende da  $A$ . La probabilità che  $x \in [\theta, \theta']$ , sapendo che  $E$  si è realizzato è viene scritta da Laplace come

$$P(x \in [\theta, \theta']|E) = \frac{\int_{\theta}^{\theta'} y(x)dx}{\int_0^1 y(x)dx},$$

supponendo ancora costante la distribuzione (continua) di probabilità delle cause. Affrontando il caso generale, Laplace asserì:

Se i valori di  $x$ , considerati indipendentemente dal risultato osservato, non sono tutti ugualmente possibili, ma la loro probabilità è espressa da una funzione  $z$  di  $x$ , basterà cambiare (...)  $y$  in  $yz$ , ciò che equivale a supporre tutti i valori di  $x$  come equiprobabili e a considerare il risultato osservato come formato da due risultati indipendenti, le cui probabilità sono  $y$  e  $z$ . È dunque possibile ricondurre tutti i casi a quello in cui si supponga la stessa possibilità ai differenti valori di  $x$  e, per questo motivo, noi adotteremo quest'ipotesi nelle ricerche successive.<sup>61</sup> ([18], p. 303)

Laplace punta dunque l'attenzione sull'analogia formale tra la formula di Bayes completa e quella nell'ipotesi di uniformità delle cause, proponendo in sostanza un cambio di variabile per passare da un caso all'altro, senza prestare troppa attenzione all'impatto di una distribuzione a priori delle cause piuttosto che un'altra. Nella *Théorie analytique* ([19], p. 182) Laplace presenterà la formula di Bayes anche nel caso di distribuzione a priori non uniforme delle cause.

Già nella memoria del 1774, Laplace aveva considerato diversi esempi ad ampio spettro, per illustrare le potenzialità del metodo basato sul "Principio" enunciato sopra. Ne consideriamo qui due, lasciando la discussione di un terzo esempio al Capitolo 9, dedicato alla teoria degli errori.

<sup>60</sup>Lorsqu'on n'a aucune donnée *a priori* sur la possibilité d'un événement, il faut supposer toutes les possibilités, depuis zéro jusqu'à l'unité, également probables.

<sup>61</sup>Si les valeurs de  $x$ , considérées indépendamment du résultat observé, ne sont pas toutes également possibles, mais que leur probabilité soit exprimée par une fonction  $z$  de  $x$ , il suffira de changer (...)  $y$  dans  $yz$ , ce qui revient à supposer toutes les valeurs de  $x$  comme également possibles et à considérer le résultat observé comme étant formé de deux résultats indépendants, dont les probabilités sont  $y$  et  $z$ . On peut donc ramener de cette manière tous les cas à celui où l'on suppose une égale possibilité aux différentes valeurs de  $x$  et, par cette raison, nous adopterons cette hypothèse dans les recherches suivantes.

PROBLEMA I. Se un'urna racchiude un'infinità di biglietti bianchi e neri in un rapporto incognito e se si estraggono  $p+q$  biglietti di cui  $p$  sono bianchi e  $q$  sono neri; si chiede la probabilità che, estraendo un nuovo biglietto, esso sarà bianco.<sup>62</sup> ([16], p. 31)

Nella soluzione di questo problema, Laplace chiarisce anzitutto che il rapporto tra biglietti bianchi e numero totale di biglietti può essere un numero *frazionario* qualsiasi  $x \in (0, 1)$ , locuzione ambigua che, presa in sé, non implica la razionalità di  $x$ . Se  $x$  fosse noto, la probabilità di estrarre  $p$  biglietti bianchi e  $q$  neri è proporzionale ad  $x^p(1-x)^q$ : Laplace non considerò il coefficiente binomiale associato, forse perché la sua presenza in questo problema si rivela inutile, come vedremo tra poco. Indicata con  $A_x$  l'ipotesi: l'urna contiene la frazione  $x$  di biglietti bianchi e con  $E_1$  l'evento dato dall'estrazione di  $p+q$  biglietti, di cui  $p$  bianchi, allora Laplace, grazie al principio (7.11) scrisse

$$P(A_x|E_1) = \frac{P(E_1|A_x)}{\sum P(E_1|A_x)}$$

che già richiederebbe qualche spiegazione, visto che ora a denominatore è una serie infinita di termini, il cui significato è opaco. Moltiplicando numeratore e denominatore per  $dx$  e trasformando la somma infinita in un integrale, Laplace riscrisse la formula precedente come

$$P(A_x|E_1) = \frac{x^p(1-x)^q dx}{\int_0^1 x^p(1-x)^q dx}. \quad (7.26)$$

Da qui si vede che il coefficiente binomiale  $E_{p,q}$  omissso in precedenza non sarebbe comunque comparso in questa formula, semplificandosi a numeratore ed a denominatore, come già avveniva negli esempi di Bayes. Ora, chiamiamo  $E_2$  l'evento che consiste nell'estrazione di un ulteriore biglietto bianco. Laplace osserva che, se si conoscesse il vero valore di  $x$ , la probabilità  $p(E_2|A_x)$  di estrarre un ulteriore biglietto bianco sarebbe proprio  $x$ , perché il numero di biglietti è infinito e quindi si può supporre che l'estrazione avvenga *con* reimbussolamento: dunque, poiché dall'equazione (7.25) si ha

$$P(E_2|E_1) = \sum_x P(E_2|A_x)P(A_x|E_1),$$

sostituendo il valore di  $P(A_x|E_1)$  dato dall'equazione (7.26) nell'intervallo  $[0, 1]$  si ottiene

$$P(E_2|E_1) = \frac{\int_0^1 x^{p+1}(1-x)^q dx}{\int_0^1 x^p(1-x)^q dx}.$$

---

<sup>62</sup>PROBLÈME I. Si une urne renferme une infinité de billet blancs et noirs dans un rapport inconnu, et que l'on en tire  $p+q$  billet dont  $p$  soient blancs et  $q$  soient noirs; on demande la probabilité qu'en tirant un nouveau billet de cette urne il sera blanc.

Il calcolo degli integrali si può effettuare per parti, essendo gli esponenti  $p$  e  $q$  interi:

$$\begin{aligned}\int_0^1 x^p(1-x)^q dx &= \frac{x^{p+1}}{p+1}(1-x)^q \Big|_0^1 + \frac{q}{p+1} \int_0^1 x^{p+1}(1-x)^{q-1} dx \\ &= \frac{q}{p+1} \int_0^1 x^{p+1}(1-x)^{q-1} dx.\end{aligned}$$

Iterando la procedura si ottiene

$$\int_0^1 x^p(1-x)^q dx = \frac{q!}{(p+1)(p+2)\cdots(p+q)} \int_0^1 x^{p+q} dx = \frac{p!q!}{(p+q+1)!}.$$

Valutando allo stesso modo l'integrale a numeratore, si ottiene la *probabilità intera* (*probabilité entière*) di estrarre ancora un biglietto bianco

$$P(E_2|E_1) = \frac{p+1}{p+q+2}$$

che è nota, partire da John Venn (1834-1923), come *regola di successione* ([26], Chapt. VIII) e che fu bersaglio di feroci critiche come di notevoli studi.

Laplace applicò la formula (7.26) all'analisi di diversi problemi relativi ai giochi di azzardo, nei quali entra un parametro incognito, come il problema di S. Pietroburgo, nel quale le imperfezioni inevitabili nella moneta fanno sì che la probabilità di ottenere testa in un lancio non sia esattamente  $\frac{1}{2}$  ma  $\frac{1}{2} \pm \varpi$  o come il problema della suddivisione della posta che dipende dalla probabilità di vittoria di un giocatore in ciascuna partita.

Rispetto a Bayes e Price, Laplace è più ellittico ed il suo uso del continuo è, come per De Moivre, finalizzato a semplificare l'analisi del problema, sostituendo delle serie infinite con degli integrali. Mentre Bayes *non* dimostra la formula di Bayes nella forma discreta, Laplace la pone a fondamento dell'intera teoria nella forma semplificata in cui a tutte le cause compete la stessa probabilità.

## 7.5 Un problema demografico

Laplace applicò il "Principio" della probabilità inversa a problemi con ricadute concrete nella società. In particolare, egli considerò un problema demografico che aveva incuriosito diversi studiosi sin dall'inizio del XVIII secolo: la lieve ma persistente prevalenza di neonati di sesso maschile rispetto a quelli di sesso femminile.

Abbiamo visto come già John Graunt avesse fatto cenno a questo problema che fu trattato con più dettagli in un lavoro del medico e scrittore scozzese John Aburthnot (1667-1735), che tradusse in inglese il *De Ratiociniis* di Huygens, in un breve lavoro pubblicato nel 1712 [1], che viene considerato come uno dei primi test di ipotesi. In effetti Aburthnot osservò come maschi e femmine fossero all'incirca nello stesso numero e si domandò se un tale equilibrio fosse imputabile al caso o fosse la traccia (*footstep*) della divina provvidenza. Per rispondere

a questa domanda considerò un dado a due facce (*a die with two sides*), cioè una moneta equilibrata con *M* (*male*) ed *F* (*female*) sulle facce e, con il consueto sviluppo del binomio, si chiese quale fosse la probabilità che, lanciando il dado un numero pari di volte, si ottenesse *M* ed *F* un ugual numero di volte. Poiché questa probabilità, come pure quella che vi sia un lieve sbilanciamento a favore dei maschi,<sup>63</sup> è molto bassa al crescere del numero di lanci, Aburthnot concludeva che

è molto improbabile (se tutto fosse retto solo dal caso) che essi [i termini da considerare] non raggiungano le estremità [dello sviluppo del binomio].<sup>64</sup>

Tornando a Laplace, egli considerò le statistiche di grandi metropoli come Parigi e Londra: a Parigi, dal 1745 al 1770 erano nati 251527 maschi e 241945 femmine, corrispondenti ad un rapporto  $\frac{105}{101}$  tra neonati maschi e femmine; a Londra, i dati dal 1664 fino al 1757 indicavano 737629 neonati maschi e 698958 femmine, con un rapporto di  $\frac{19}{18}$ , leggermente superiore a quello riscontrato a Parigi. Lo squilibrio non era solo cumulativo ma si riscontrava ogni anno, lasciando spazio ad un interrogativo naturale:

benché la differenza sia poco considerevole, sarebbe un fatto ben straordinario che essa fosse un effetto del caso ed è molto più naturale supporre che, in Francia ed in Inghilterra, la natura favorisca maggiormente la nascita dei maschi che quella delle femmine.<sup>65</sup> ([17], p. 429)

Laplace sapeva bene che in piccoli villaggi della Francia si poteva riscontrare, almeno su lassi temporali più brevi, una tendenza opposta, ma i dati delle città maggiori avevano dalla loro parte la forza dei numeri:

A dire il vero, le osservazioni sulle nascite avvenute in alcune piccole città della Francia durante un periodo di quattro o cinque anni sembrano indicare una minor facilità per la nascita dei maschi rispetto a quella per le femmine; è però molto probabile che, su un piccolo numero di nati, come quattro o cinquecento, ci siano più femmine che maschi, anche se la facilità alla nascita di questi ultimi è più grande; in questo delicato genere di ricerche occorre adoperare numeri molto più grandi, vista soprattutto l'esiguità della differenza tra la facilità di nascita dei maschi rispetto a quella delle femmine e sarà lecito cercare la causa di quel fenomeno solo quando ci sarà una solida garanzia che il numero di nascite osservato in

<sup>63</sup>Per Aburthont, come già per Graunt, l'equilibrio sostanziale tra neonati di un sesso e dell'altro è anche una dimostrazione della inaccettabilità della poligamia.

<sup>64</sup>it is very improbable (if mere Chance govern'd) that they would never reach as far as the Extremities.

<sup>65</sup>quoique la différence soit peu considérable, il serait assez extraordinaire que cela fût l'effet du hasard, et il est bien plus naturel de penser que, en France et en Angleterre, la nature favorise plus la naissance des garçons que celle des filles.

un luogo indichi, con probabilità molto grande, che la nascita dei maschi sia meno possibile di quelle delle femmine.<sup>66</sup> ([17], p. 429)

Si tratta di una dichiarazione metodologica che evidenzia la necessità di valutare quanto un dato sia *significativo* dal punto di vista statistico per potervi costruire una teoria che ne spieghi le ragioni. Il problema del sesso dei nascituri è, per Laplace, la palestra ideale su cui applicare la probabilità delle cause: si suppone di sapere che, su  $p + q$  neonati,  $p$  siano maschi e  $q$  femmine; si vuole trovare la probabilità  $P$  che, su un numero di  $n + m$  nascite future,  $m$  siano di maschi ed  $n$  di femmine. Si tratta di riprodurre lo schema già usato per dedurre la regola di successione. Detta  $x \in [0, 1]$  la probabilità che un nascituro sia maschio, sia  $A_x$  l'ipotesi che l'urna cui il problema è ridotto contiene una frazione  $x$  di neonati maschi. Conoscendo questo dato, l'evento  $E$  che consta nella nascita di  $p$  neonati maschi su  $p + q$  ha probabilità

$$P(E|A_x) = \lambda x^p (1-x)^q \quad \text{con} \quad \lambda = \binom{p+q}{p}$$

e quindi vale ancora l'espressione (7.26) per  $P(A_x|E)$ . D'altra parte, se  $e$  è l'evento per cui nasceranno  $m$  maschi su  $m + n$  neonati, si avrà

$$P(e|A_x) = \gamma x^m (1-x)^n \quad \text{con} \quad \gamma = \binom{m+n}{m}.$$

Siccome  $x$  è incognita, Laplace suppone che sia distribuita uniformemente e ricava, passando al continuo per trasformare somme in integrali,

$$P = P(e|E) = \sum_x P(e|A_x)P(A_x|E) = \frac{\gamma \int_0^1 x^{p+m} (1-x)^{q+n} dx}{\int_0^1 x^p (1-x)^q dx}.$$

Calcolati gli integrali ed applicata la formula di Stirling ripetutamente, si giunge a

$$P = \gamma \frac{(q+n)^{q+n+\frac{1}{2}} (p+q)^{p+q+\frac{1}{2}} (p+m)^{p+m+\frac{1}{2}}}{(p)^{p+\frac{1}{2}} (q)^{q+\frac{1}{2}} (p+q+m+n)^{p+q+m+n+\frac{1}{2}}}$$

che, nel limite in cui sia  $m$  che  $n$  sono molto più piccoli rispetto a  $p$  e  $q$ , conduce a

$$P = \gamma \frac{p^m q^n}{(p+q)^{m+n}}. \quad (7.27)$$

<sup>66</sup>A la vérité, les naissances observées pendant quatre ou cinq ans dans quelques petites villes de France semblent y indiquer une moindre facilité pour la naissance des garçons que pour celle des filles; mais il est très possible que, sur un petit nombre des naissances, tel que quatre o cinq cents, al y ait plus filles que des garçons, quoique la facilité de la naissance de ceux-ci soit plus grande; il faut employer à cette recherche délicate de beaucoup plus grands nombres, vu surtout le peu de différence qui existe entre les facilités des naissances des garçons et des filles, et ce n'est que lorsqu'on sera bien assuré que le nombre observé des naissances dans un lieu quelconque indique, avec une très grande probabilité, que les naissances des garçons y sont moins possibles que celles des filles, qu'il sera permis de rechercher la cause de ce phénomène.

Il valore di  $P$  ottenuto in (7.27) è lo stesso, osserva Laplace, che si sarebbe ottenuto supponendo che le probabilità di nascita di un maschio e quelle di una femmina fossero nel rapporto di  $p$  su  $q$ , cioè nel rapporto *osservato* nel passato. In questo senso, i dati accumulati nel passato forniscono un'indicazione su quanto potrà avvenire in futuro. Ora, l'esito dell'analisi che ha condotto a (7.27) viene riassunto da Laplace in termini del teorema di Bernoulli:

Se si indica con  $\theta$  una quantità molto piccola e con  $P$  la probabilità che la possibilità della nascita di un maschio sia compresa tra  $\frac{p}{p+q} - \theta$  e  $\frac{p}{p+q} + \theta$ , il valore di  $P$  differirà tanto meno dalla certezza, cioè da 1, quanto più grandi saranno  $p$  e  $q$  ed è possibile far crescere  $p$  e  $q$  a tal punto che la differenza tra  $P$  ed 1 sia inferiore ad ogni grandezza assegnata, per quanto  $\theta$  possa essere piccolo.<sup>67</sup> ([17], pp. 420-421)

Poiché però né  $p$  né  $q$  sono infiniti, occorre stimare l'entità del divario tra  $P$  e  $\frac{p}{p+q}$  e Laplace valuta questa differenza attraverso ad una serie diretta ma un po' lunga di passaggi che egli applica al problema in esame calcolando la probabilità che  $x$ , probabilità che un nascituro sia maschio, sia inferiore ad  $\frac{1}{2}$ . Ottenuto un valore di circa  $10^{-42}$ , Laplace trasse la conclusione che

si può ritenere altrettanto certo quanto una qualunque altra verità morale, che la differenza osservata a Parigi tra le nascite di maschi e di femmine sia dovuta ad una maggiore facilità di nascita di maschi.<sup>68</sup> ([17], pp. 431-432)

## 7.6 Poisson e la probabilità inversa

Siméon-Denis Poisson (1781-1840) fu, in ordine cronologico, l'ultimo degli allievi di Laplace e le opere in cui più si è occupato del calcolo delle probabilità risalgono al suo ultimo decennio di vita. Fra queste spiccano una memoria [21] centrata sul problema demografico già affrontato da Laplace ed una monografia incentrata sulle applicazioni giuridiche della probabilità [22]. Nella lunga memoria—*Mémoire sur la proportion des naissances des filles et des garçons*—del 1830 Poisson si soffermò a lungo sul teorema di Bernoulli-De Moivre per discutere, oltre alla distribuzione normale, anche il limite che conduce alla distribuzione oggi nota con il suo nome. Anche se non direttamente collegato al contenuto del capitolo, diamo un cenno alla deduzione di Poisson. Al solito, sia  $A$  un evento di probabilità  $p$  costante e sia  $q = 1 - p$  la probabilità dell'evento complementare  $B$  in una prova. Detto  $n$  il numero di prove, la probabilità  $X$  che

<sup>67</sup>Si l'on désigne par  $\theta$  une quantité très fort petite et par  $P$  la probabilité que la possibilité de la naissance d'un garçon est comprise dans les limites  $\frac{p}{p+q} - \theta$  et  $\frac{p}{p+q} + \theta$ , la valeur de  $P$  différera d'autant moins de la certitude ou de l'unité que  $p$  et  $q$  seront de plus grands nombres, et l'on peut tellement faire croître  $p$  et  $q$  que la différence de  $P$  à l'unité soit moindre qu'aucune grandeur donnée, quelque petit que  $\theta$  soit d'ailleurs.

<sup>68</sup>on peut regarder comme aussi certain qu'aucune autre vérité morale, que la différence observée à Paris entre les naissances des garçons et celles des filles est due à une plus grande possibilité dans la naissance des garçons.

A non si verifichi più di  $x$  volte su  $n$ , cioè che  $B$  si verifichi almeno  $n - x := m$  volte viene calcolata scomponendo questo evento in più eventi incompatibili:

- Evento  $E_0$ .  $B$  si verifica  $m$  volte nelle prime  $m$  prove. Questo evento ha probabilità  $q^m$  di realizzarsi;
- Evento  $E_1$ .  $B$  si verifica  $m$  volte nelle prime  $m+1$  prove ed  $A$  si verifica una volta, che non dovrà essere l'ultima per non ricadere nel caso precedente. Poiché dunque  $A$  può realizzarsi in una delle prime  $m$  prove, la probabilità di questo evento è  $mpq^m$ .
- Evento  $E_2$ .  $B$  si verifica  $m$  volte nelle prime  $m+2$  prove ed  $A$  si verifica due volte, senza che possa verificarsi alla prova  $(m+2)$ -esima, altrimenti si rientrerebbe nei casi precedenti. Dunque  $A$  deve verificarsi due volte in  $m+1$  prove e la probabilità che spetta a questo caso è  $\frac{(m+1)m}{2}p^2q^m$
- .....
- Evento  $E_x$ .  $B$  si verifica  $m$  volte nelle  $m+x=n$  prove ed  $A$  si verifica  $x$  volte, senza che possa verificarsi all'ultima prova, per non ricadere in uno dei casi precedenti. Dunque  $A$  può verificarsi  $x$  volte in  $m+x-1$  prove e la probabilità che spetta a questo caso è

$$\frac{(m+x-1)(m+x-2)\cdots(m+1)m}{x!}p^xq^m.$$

Sommando i risultati si ottiene

$$X = \sum_{k=0}^x p(E_k) = q^{n-x} \sum_{k=0}^x \binom{n-x+k-1}{k} p^k.$$

L'approssimazione alla De Moivre, che Poisson riottiene utilizzando i metodi analitici di Laplace, richiede che non solo  $n$ , ma anche  $x$  ed  $n-x$  siano molto grandi e di conseguenza, se si suppone che né  $p$  né  $q$ , siano troppo piccoli, che anche  $pn$  e  $qn$  siano molto grandi. Quando invece  $p \ll 1$ ,  $n \gg 1$  ma in modo che  $pn = \omega$  resti finito e comunque  $\frac{x}{n} \ll 1$  abbiamo dapprima

$$X = q^{n-x} \left[ 1 + (n-x)p + \frac{(n-x)(n-x+1)}{2}p^2 + \cdots + \right. \\ \left. + \cdots + \frac{(n-x)(n-x+1)\cdots(n-2)(n-1)}{x!}p^x \right]$$

ed osservando che  $np = \omega$  mentre sia  $px = \frac{\omega x}{n}$  che  $p$  sono molto inferiori ad 1, possiamo scrivere anzitutto, per  $n \gg 1$

$$q^n = (1-p)^n = \left(1 - \frac{\omega}{n}\right)^n \simeq e^{-\omega} \quad \text{e} \quad q^{-x} = \left(1 - \frac{\omega}{n}\right)^{-x} \simeq 1 + \frac{x}{n}\omega \simeq 1$$

$$X = \left[ \sum_{k=0}^x \frac{\omega^k}{k!} \right] e^{-\omega}$$

in particolare, la probabilità che l'evento  $A$  non si presenti mai nelle  $n$  prove si ottiene ponendo  $x = 0$  e vale  $e^{-\omega}$ .

Nella seconda parte della memoria [21], Poisson introdusse il modello delle urne per risolvere problemi di probabilità inversa, su cui ritornò con maggior chiarezza nel Capitolo II delle *Recherches* e per questo seguiremo questa seconda esposizione.

Qui Poisson considerò un evento  $E$  il cui realizzarsi può essere attribuito *solo* ad un numero finito  $m$  di cause  $C_1, \dots, C_m$  mutuamente esclusive e che, prima del realizzarsi o meno di  $E$ , sono ritenute *equiprobabili*:

$$P(C_1) = P(C_2) = \dots = P(C_m).$$

Questa situazione di equilibrio viene rotta, per così dire, dal realizzarsi di  $E$ . Assegnate  $p_i := P(E|C_i)$ , per trovare  $\pi_i = P(C_i|E)$  Poisson considerò un problema che riteneva *equivalente* al precedente:

Ora, quale che sia l'evento  $E$  è possibile assimilarlo, per fissare le idee, all'estrazione di una pallina bianca da un'urna contenente palline di quel colore e palline nere.<sup>69</sup> ([22], p. 82)

Dette  $U_1, U_2, \dots, U_m$  queste urne, in numero pari alle cause  $C_i$  dell'evento  $E$ , Poisson suppose che la frazione di palline bianche contenuta in  $U_i$  sia  $p_i$ . Nell'analogia, l'urna  $U_i$  gioca il ruolo della causa  $C_i$  ed occorre determinare la probabilità che una pallina bianca provenga da un'urna specifica  $U_i$ . Poisson riduce le probabilità  $p_i$  allo stesso denominatore, scrivendo

$$p_i = \frac{\alpha_i}{\mu} \quad \forall i = 1, \dots, m : \quad (7.28)$$

egli pertanto assume che sia  $\mu$  che i vari  $\alpha_i$  siano interi, contravvenendo in parte all'estensione del concetto di probabilità agli irrazionali cui aveva fatto allusione in precedenza nelle *Recherches*. Immaginando dunque che ciascuna urna contenga lo stesso numero  $\mu$  di palline bianche o nere, Poisson considerò una nuova urna  $U$  in cui vengono riversate tutte le palline, avendo però cura di numerare con 1 quelle provenienti dall'urna  $U_1$ , ..., con  $m$  quelle provenienti dall'urna  $U_m$ . La probabilità  $\pi_i$  coincide con la probabilità di estrarre da  $U$  una pallina bianca numerata con  $i$ . Ora, quest'ultima probabilità vale

$$\pi_i = \frac{\alpha_i}{\sum_{k=1}^m \alpha_k} = \frac{p_i}{\sum_{k=1}^m p_k}, \quad (7.29)$$

dove, nell'ultimo passaggio, abbiamo utilizzato l'equazione (7.28). L'equazione (7.29) è la formula di Bayes nel caso di cause a priori equiprobabili.

Nell'art. 34 delle *Recherches*, Poisson considerò in dettaglio il caso in cui le cause  $C_i$  *non* hanno la stessa probabilità *a priori*. Poisson ricorse ancora alla

<sup>69</sup>Or, quel que soit l'événement  $E$ , on peut l'assimiler, pour fixer les idées, à l'arrivée d'une boule blanche, extraite d'une urne qui contenait des boules de cette couleur et des boules noires.

“finzione” delle urne  $U_i$  come surrogato delle cause  $C_i$  procedendo a raggrupparle in  $\ell$  famiglie  $\mathcal{F}_1, \dots, \mathcal{F}_\ell$ . Se  $s$  è il numero complessivo di urne e la famiglia  $\mathcal{F}_k$  è formata da  $a_k$  urne tutte con la *stessa* composizione, allora

$$s = \sum_{k=1}^{\ell} a_k.$$

Se  $p_i$  è la probabilità di estrarre una pallina bianca da un’urna che appartiene alla famiglia  $\mathcal{F}_i$ , avremo

$$\pi_i = \frac{a_i \alpha_i}{\sum_{k=1}^{\ell} a_k \alpha_k} = \frac{a_i p_i}{\sum_{k=1}^{\ell} a_k p_k}$$

e siccome  $P(C_i) =: q_i = \frac{a_i}{s}$ , la formula precedente diventa

$$\pi_i = P(C_i|E) = \frac{q_i p_i}{\sum_{k=1}^{\ell} p_k q_k} = \frac{P(E|C_i)P(C_i)}{\sum_{k=1}^{\ell} P(E|C_k)P(C_k)}. \quad (7.30)$$

Poisson sottolinea la necessità di avere a disposizione la formula di Bayes completa, soprattutto in vista delle applicazioni giuridiche, già nel preambolo alle *Recherches* in cui aveva brevemente tratteggiato la storia recente di queste applicazioni, riconoscendo a Condorcet il primato di aver applicato i risultati di Bayes a questo genere di problemi e criticando l’ipotesi di uniformità delle cause formulata da Laplace:

È corretto dire che fu di Condorcet l’idea ingegnosa di far dipendere la soluzione [dei problemi giudiziari] dal principio di Blayes [sic!], considerando successivamente la colpevolezza e l’innocenza dell’accusato come una causa sconosciuta del giudizio pronunciato, che dunque è il fatto osservato, da cui occorre dedurre la probabilità di questa causa. La correttezza di questo principio si dimostra con pieno rigore; la sua applicazione alla questione di cui ci occupiamo non può lasciare alcun dubbio; ma per questa applicazione, Laplace formulò un’ipotesi che non è affatto incontestabile: egli suppone che la probabilità che un giurato non si sbaglia possa assumere tutti i valori ugualmente possibili dalla certezza, rappresentata dall’unità, fino all’indifferenza, che nel calcolo corrisponde alla frazione  $\frac{1}{2}$  e si riferisce ad una uguale probabilità di errore e di esattezza. L’illustre matematico fonda la sua ipotesi su questo: che l’opinione di un giurato è attratta più verso la verità che verso l’errore, cosa che in generale occorre ammettere. Esistono però infinite leggi di errore che soddisfano questa condizione, senza essere costretti a supporre che la probabilità che un giurato non si sbaglia non possa mai scendere sotto a  $\frac{1}{2}$  e che sopra questo limite tutti i valori siano ugualmente possibili. L’ipotesi particolare di Laplace non potrebbe essere dunque giustificata *a priori*.<sup>70</sup> ([22], pp. 2-3)

<sup>70</sup>il est juste de dire que c’est à Condorcet qu’est due l’idée ingénieuse de faire dépendre la solution, du principe de Blayes, en considérant successivement la culpabilité et l’innocence

Vediamo l'applicazione più semplice presentata da Poisson, in cui si considera un fatto *non* inconfutabile che ha, a priori, probabilità  $q$  di essere vero ed  $1 - q$  di non esserlo. Supponiamo che un testimone sia chiamato a deporre in merito e che la probabilità che egli non ci inganni, involontariamente o di proposito, sia  $p$ . Supponiamo che il testimone attesti la verità del fatto: come cambierà  $q$  in seguito a questa testimonianza? L'evento osservato è l'attestazione della verità di un fatto da parte del testimone. Nell'ipotesi che il fatto sia vero, la probabilità che il testimone dica il vero è  $p$ . Indicati con  $V$  ed  $F$  le ipotesi: il fatto è vero oppure falso; con  $v, f$  gli eventi: il testimone asserisce che il fatto è vero oppure che è falso, abbiamo

$$P(v|V) = p \quad P(v|F) = 1 - p$$

oltre che

$$P(V) = q \quad P(F) = 1 - q$$

Applicando (7.30) si ottiene allora

$$r := P(V|v) = \frac{pq}{pq + (1-p)(1-q)}$$

e dunque

$$r - q = \frac{q(1-q)(2p-1)}{pq + (1-p)(1-q)} \quad (7.31)$$

rappresenta l'incremento o diminuzione della credibilità del fatto a seguito della testimonianza. Essa è positiva se  $p > \frac{1}{2}$ , negativa se  $p < \frac{1}{2}$  e si annulla quando  $p = \frac{1}{2}$  cioè quando vi sono uguali possibilità che il testimone sia veritiero o mentisca. Se  $q = \frac{1}{2}$  allora  $r = p$ , cioè la veridicità del fatto

non dipende da altro se non dalla veracità e dai lumi del testimone.<sup>71</sup>  
([22], p. 99)

Altri limiti interessanti sono quelli in cui  $q$  è prossimo a 0 mentre  $p$  tende ad 1 ma in modo che  $x = \frac{q}{1-p} \ll 1$ . Questo caso corrisponde alla situazione in cui il fatto appare in sé inverosimile, perché contrario a leggi naturali suffragate

---

de l'accusé, comme une cause inconnue du jugement prononcé, qui est alors le fait observé, duquel s'agit de déduire la probabilité de cette cause. L'exactitude de ce principe se démontre en toute rigueur; son application à la question qui nous occupe, ne peut non plus laisser aucun doute; mais pour cette application, Laplace fait une hypothèse qui n'est point incontestable: il suppose que la probabilité qu'un juré ne se trompera pas est susceptible de tous les degrés également possibles, depuis la certitude, représentée par l'unité, jusqu'à l'indifférence, qui répond dans le calcul à la fraction  $\frac{1}{2}$ , et se rapport à une égale chance d'erreur et de vérité. L'illustre géomètre fonde son hypothèse sur ce que l'opinion d'un juré a sans doute plus de tendance vers la vérité que vers l'erreur; ce qu'on doit admettre en général. Mais il existe une infinité des lois différentes de probabilité des erreurs qui satisfont à cette condition, sans qu'on soit obligé de supposer que la chance qu'un juré ne se trompera pas, ne puisse jamais descendre au-dessous de  $\frac{1}{2}$ , et qu'au-dessus de cette limite, toutes ses valeurs sont également possibles. La supposition particulière de Laplace ne saurait donc être justifiée *a priori*.

<sup>71</sup>ne dépend plus que de la véracité et des lumières du témoin.

da molte esperienze, mentre il testimone che ne afferma la verità è degno della massima fiducia. In questo caso si vede che deve essere

$$r - q = x + o(x) \quad (7.32)$$

per cui la testimonianza cambia di poco il nostro giudizio sulla scarsa verosimiglianza dell'evento: appare ragionevole non dare credito alla testimonianza e supporre che il testimone, in buona fede, si inganni. L'opera di Poisson contiene molti altri esempi su cui non possiamo dilungarci. L'applicazione della probabilità al campo delle scienze giuridiche, grazie al teorema di Bayes, iniziata da Condorcet e Laplace e proseguita da Poisson segna un altro momento di riflessione critica della disciplina che si sviluppò negli anni immediatamente successivi alle *Recherches*.

## 7.7 La regola di successione finita

Abbiamo già parlato più volte della regola di successione, mostrando la deduzione di Laplace nel caso continuo e secondo uno schema in cui le palline nell'urna sono infinite. Uno schema diverso fu considerato a partire dalla fine del '700 da Pierre Prévost (1751-1839) e da Simone Antoine Jean Lhuilier (1750-1840) che proposero il seguente problema ([23], p. 117):

Un'urna contenga due tipi di biglietti (che dirò bianchi e neri), in un rapporto incognito. Si estraiga successivamente un certo numero di biglietti senza reinserire il biglietto estratto nell'urna. Sapendo il numero di biglietti dei due tipi che sono stati estratti, si chiede la probabilità che, estraendo altri biglietti allo stesso modo, ne escano un numero preassegnato dei due tipi.<sup>72</sup>

Il primo punto da osservare è che si tratta di una estrazione senza rimborso, ed in questo senso lo schema di Prévost e L'Huilier rappresenta una novità in questo genere di problemi. Anziché seguire la dimostrazione dei due studiosi svizzeri, illustro le tappe salienti di quella fornita dal matematico italiano Antonio Bordini (1788-1860) in un lavoro [6] pubblicato nel 1843 dove il problema è riletto in un contesto differente:

Accade talvolta, sul terminare di un esame, che l'individuo, facente l'esame stesso, non è contento delle risposte da lui medesimo già date; ovvero, che l'esaminatore, od in generale chi deve votare sull'esito dell'esame, non ha ancora acquistata quella cognizione dell'interrogato, che è necessaria per dargli scientemente il voto; per cui l'uno o l'altro, od entrambi, desiderano, e generalmente convengono ben anco, qualora sia permesso, di continuare l'esame sulla materia rimanente.

<sup>72</sup>Soit une urne contenant des billets de deux espèces (que j'appellerai blancs et noirs), dans un rapport inconnu. Soit tiré successivement un certain nombre de ces billets, sans remettre dans l'urne, à chaque extraction, le billet tiré. Connoissant le nombre des billets de chaque espèce qui ont été tirés, on demande la probabilité que tirant de la même manière de nouveaux billets, en nombre donné, il y a des nombres donnés de ces deux espèces.

Bordoni, già a partire da una monografia pubblicata nel 1837 [5], intendeva applicare il calcolo delle probabilità per stimare il grado di preparazione *effettiva* di uno studente, conoscendo il numero di risposte corrette da lui date in un esame. Bordoni supponeva che le domande ammissibili per l'esame fossero un numero assegnato  $a$ , di cui  $b$  note al candidato mentre sulle altre egli non aveva preparazione alcuna. Le domande erano disposte in un'urna da cui ne veniva estratto, senza reimbussolamento, un certo numero  $m$  proposto al candidato. Supponendo che egli riuscisse a rispondere correttamente ad  $r$  domande, il rapporto  $\frac{r}{m}$  misura il successo dell'esame. Bordoni era interessato a conoscere cosa si potesse concludere sul rapporto  $\frac{b}{a}$ , che misura la *reale* preparazione dello studente, a partire dalla conoscenza di  $\frac{r}{m}$ . In sostanza, egli assimilava un esame ad una successiva estrazione di palline bianche (le domande alle quali il candidato sapeva rispondere correttamente) e nere (quelle la cui risposta egli ignorava completamente). Senza addentrarci a discutere la rigidità di questa ipotesi, vediamo il metodo seguito da Bordoni per risolvere il problema. Supponendo dunque che il candidato avesse risposto correttamente ad  $r$  quesiti su  $m$ , si vuole trovare la probabilità che, su  $m'$  ulteriori domande, sempre estratte a sorte, ne conosca altre  $r'$ . Bordoni ([6] p. 197) indicò con  $q_x$

la probabilità, che fra gli  $m'$  vi siano questi  $r'$ , ammesso  $x$  il numero di quelli fra tutti gli  $a$  da lui medesimo conosciuti.

cioè

$$q_x = p\left(\frac{r'}{m'} \mid \frac{x}{a}\right).$$

L'obiettivo è quello di calcolare

$$Q = p\left(\frac{r'}{m'} \mid \frac{r}{m}\right). \quad (7.33)$$

Posto ancora  $P_x = p\left(\frac{x}{a} \mid \frac{r}{m}\right)$ , probabilità che il candidato conosca effettivamente  $x$  delle  $a$  domande ammissibili, noto che sia l'esito della prima parte dell'esame, Bordoni scrisse

$$Q = \sum_{x=c}^{a-n-n'} P_x q_x$$

dove  $c = r + r'$  rappresenta il numero minimo di domande di cui il candidato conosce la risposta esatta,  $n = m - r$  e  $n' = m' - r'$  sono il numero di risposte *non* date dal candidato nelle due parti dell'esame. Dietro l'uso di questa formula vi sono due osservazioni. Anzitutto, supponendo noto  $x$ , dopo la prima estrazione saranno rimaste  $x' = x - r$  domande note al candidato nell'urna su un totale  $a' = a - m$  domande ammissibili. Le domande di cui il candidato non conosce la domanda sono diminuite di  $n = m - r$  unità. Dunque  $q_x$  rappresenta la probabilità che il candidato ha di rispondere ad  $r'$  nuove domande su  $m'$  proposte, sapendo che l'urna da cui sono estratte ne contiene  $a'$ , di cui  $x'$  a lui note. D'altra parte  $P_x$  rappresenta la probabilità che l'urna abbia questa composizione. Ora, per il significato di  $Q$ , si vede che solo i valori di  $x$  selezionati nella

(7.33) corrispondono a valori non nulli di  $q_x$ . Il problema è dunque spostato al calcolo di  $q_x$  e  $P_x$ . Per il primo termine, determiniamo anzitutto ([5], pp. 98-99) il numero di combinazioni che si possono formare con  $x'$  oggetti, combinati a gruppi di  $r'$  e di  $a' - x'$  oggetti, combinati a gruppi di  $n'$ . Moltiplicando tra loro questi numeri otterremo il numero  $y_{r'}$  di modi in cui è possibile estrarre, tra le  $a'$  domande rimaste dopo la prima parte dell'esame,  $r'$  note al candidato ed  $m' - r'$  di cui ignora la risposta:

$$y_{r'} = \binom{x'}{r'} \binom{a' - x'}{n'}.$$

D'altra parte il numero di modi con cui scegliere  $m'$  domande su  $a'$  ammissibili è

$$A := \binom{a'}{m'}$$

e quindi, semplificando alcuni fattoriali,

$$q_x = \frac{y_{r'}}{A} = \frac{m'!}{r'!n'!} \frac{(x', x' - r' + 1)(a' - x', a' - x' - n' + 1)}{(a', a' - m' + 1)}$$

dove abbiamo impiegato della notazione di Bordoni

$$(a, a - b) := a(a - 1) \cdots (a - b + 1)(a - b)$$

per il fattoriale decrescente.

Per calcolare  $P_x = p\left(\frac{x}{a} \mid \frac{r}{m}\right)$  Bordoni si servì della formula di Bayes

$$P_x = \frac{p_x}{\sum_{x=r}^{a-n} p_x}$$

dove  $p_x = p\left(\frac{r}{m} \mid \frac{x}{a}\right)$  rappresenta la probabilità che uno studente conosca la risposta ad  $r$  domande sulle  $m$  poste nella prima parte dell'esame, noto che sia il numero  $x$  di risposte note sul totale delle domande ammissibili: Bordoni ha qui tacitamente ammesso che la probabilità a priori che uno studente conosca  $x$  delle  $a$  domande sia distribuita uniformemente. Il calcolo di  $p_x$  si effettua sulla falsariga di quello che ha condotto a  $q_x$  e fornisce

$$p_x = \frac{m!}{r!n!} \frac{(x, x - r + 1)(a - x, a - x - n + 1)}{(a, a - m + 1)}.$$

I calcoli restanti delle somme finite furono svolti da Bordoni ricorrendo a formule di integrazione finita e conducono al risultato

$$Q = \frac{(m + 1, n + 1)}{r!} \frac{(m', n' + 1)}{r'!} \frac{(r + r')!}{(m + m' + 1, n + n' + 1)}$$

che, questo è l'aspetto importante, *non* dipende da  $a$ . Inoltre, ponendo  $m' = r' = 1$  ed  $n' = 0$  si ottiene per  $Q$  l'espressione

$$Q = \frac{r + 1}{m + 2}$$

che coincide con la regola di Laplace, pur ottenuta a partire da ipotesi diverse.



# Bibliografia

- [1] J. Aburthnot: An Argument for Divine Providence, taken from the Constant Regularity observed in the Births of both Sexes. *Philosophical Transactions of the Royal Society of London*, **27**, 186-190, (1710).
- [2] T. Bayes, R. Price: An essay towards solving a problem in the Doctrine of Chance. By the Late Rev. Mr Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418, (1763).
- [3] J. Bertrand: *Calcul des probabilités*. Gauthier-Villars, Paris, (1889).
- [4] D. Black: *The Theory of Committees and Elections*. Kluwer, Dordrecht, (1958).
- [5] A. Bordoni: *Sopra gli esami scolastici*, Giusti, Milano, (1837).
- [6] A. Bordoni: Sugli esami ossia sul merito di un esaminato. *Memorie dell'Imperiale Regio Istituto Lombardo*, **1**, 189-216, (1843).
- [7] J.-A.-N. De Caritat, marquis de Condorcet: *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris, (1785).
- [8] E. Czuber: *Wahrscheinlichkeitsrechnung*. I Band. Teubner, Leipzig und Berlin, (1908).
- [9] A. I. Dale: *A History of Inverse Probability. From Thomas Bayes to Karl Pearson*. II Edition. Springer, New York, (1999).
- [10] P. Dupont: Un joyau dans l'histoire des sciences: le mémoire de Thomas Bayes de 1763. *Rendiconti del Seminario Matematico dell'Università Politecnica di Torino*, **37**, 105-138, (1979).
- [11] W. Feller: *An Introduction to Probability Theory and its Applications*, Volume I. Third Edition, Revised Printing. Wiley & Sons, New York, (1968).
- [12] W.V. Gehrlein: *Condorcet's paradox*. Springer, Berlin, (2006).

- [13] D.A. Gillies: Was Bayes a Bayesian? *Historia Mathematica*, **14**, 325-346, (1987).
- [14] B.V. Gnedenko: *Teoria della Probabilità*. II Edizione. Editori Riuniti, Edizioni MIR, Roma, (1992).
- [15] J.M. Keynes: *A treatise on probability*. Mac Millan & co., London, (1921).
- [16] P. S. de Laplace: Sur la probabilité des causes par les événements. *Mémoires de l'Académie royale des Sciences de Paris*, **6**, 621-656, (1774). In *Œuvres de Laplace*. Vol. VIII, Gauthier-Villars, Paris, 27-65, (1891).
- [17] P. S. de Laplace: Mémoire sur les probabilités. *Mémoires de l'Académie royale des Sciences de Paris*, **6**, , (1778). In *Œuvres de Laplace*. Vol. IX, Gauthier-Villars, Paris, 383-485, (1893).
- [18] P. S. de Laplace: Mémoire sur les approximations des formules qui sont fonctions de très grand nombres. *Mémoires de l'Académie royale des Sciences de Paris*, 1-88, (1782); 423-467 (1783). In *Œuvres de Laplace*. Vol. X, Gauthier-Villars, Paris, 209-338, (1894).
- [19] P.-S. de Laplace: *Théorie analytique des probabilités*. II Édition. Courcier, Paris, (1814).
- [20] J. Loveland: Buffon, the Certainty of Sunrise, and the Probabilistic Reductio ad Absurdum. *Archive for History of Exact Sciences*, **55**, 465-477, (2001).
- [21] S.-D. Poisson: Mémoire sur la proportion des naissances des filles et des garçons. *Mémoires de l'Académie des Sciences de l'Institut de France*, **9**, 239-308, (1830).
- [22] S.-D. Poisson: *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris, Bachelier, (1837).
- [23] P. Prévost, S.A. L'Huilier: Sur les probabilités. *Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Berlin, 1796*, 117-142, (1799).
- [24] O. B. Sheynin: P.S. Laplace's work on probability. *Archive for History of Exact Sciences*, **16**, 137-182, (1976).
- [25] I. Todhunter: *A History of the Mathematical Theory of Probability. From the Time of Pascal to that of Laplace*. MacMillan and Co., Cambridge and London (1865).
- [26] J. Venn: *The logic of chance*. III Edition. London, MacMillan & co., (1888).
- [27] H.P. Young: Condorcet's Theory of Voting. *The American Political Science Review*, **82**, 1231-1244, (1988).

## Capitolo 8

# La teoria degli errori

### 8.1 I primi passi della teoria

Nella Terza Giornata del *Dialogo sopra i due massimi sistemi del mondo*, pubblicato nel 1632 [11], Galileo presentò alcune considerazioni sugli errori accidentali nelle osservazioni astronomiche, a proposito delle conclusioni che l'astronomo Scipione Chiaramonti (1565-1652) aveva tratto sulla distanza dalla Terra della *Nova*, una stella luminosissima, improvvisamente comparsa nella costellazione di Cassiopea nel 1572. Chiaramonti sosteneva, grazie a calcoli eseguiti sulla scorta di osservazioni astronomiche effettuate da diversi astronomi, che la *Nova* fosse più vicina della Luna e non andasse collocata tra le stelle fisse. La questione non era di poco conto perché solo la collocazione sublunare della stella era compatibile con l'incorruttibilità delle sfere celesti sostenuta da Aristotele.

Combinando le osservazioni dei vari autori esaminate si poteva concludere che tutte erano affette, in maggior o minor misura, da errore e che andavano scartate sia quelle che collocavano la *Nova* all'interno della Terra, sia quelle che la ponevano oltre il firmamento, il cielo delle stelle fisse. Quando le osservazioni non fornivano risultati fisicamente inaccettabili

se noi vogliamo pur da tante fatiche ritrar qualche frutto, bisogna ridursi alle correzioni, emendando tutte l'osservazioni (...) e tra queste solamente andar ricercando, per via de più probabili e più numerosi riscontri, se non il sito particolare e giusto, cioè la sua vera distanza dal centro della Terra, almeno di venire in cognizione se ella fu tra gli elementi o pur tra i corpi celesti. ([11], pp. 371-372)

Per non incorrere in grossolane semplificazioni, come aveva fatto Chiaramonti, occorre avere chiare alcune proprietà degli errori cui vanno soggette le misurazioni. Salviati, l'accusatore di Chiaramonti nel *Dialogo*, domandò allora a Simplicio

E prima io vi domando, se gli astronomi nell'osservare con loro strumenti, e cercar, verbigratia, quanta sia l'elevazione di una stella sopra l'orizzonte,

possono deviar dal vero tanto nel più quanto nel meno, (...) o pure se l'errore non può mai esser se non d'un genere, cioè che, errando, sempre pecchino nel soverchio e non mai nel meno, o sempre nel meno né già mai nel soverchio. ([11], p. 372)

Accettato che gli errori possono discostarsi dal valor vero sia in difetto che in eccesso, Salviati richiama l'attenzione di Simplicio sull'impatto che un piccolo errore di misurazione può avere sull'osservabile che si vuole misurare. In particolare, un piccolo errore nelle osservazioni che già danno un valore considerevole alla distanza della stella, potrebbe portare questa distanza molto più lontano. Ora, poiché

è da credere che gli osservatori abbiano errato più presto di poco che di assai ([11], p. 397)

il numero delle osservazioni compatibili con la posizione della *Nova* nel firmamento accrescevano il loro peso. Accettato che gli astronomi menzionati da Chiaramonti erano persone diligenti e constatata comunque la presenza di errori in tutte le osservazioni, come comportarsi di fronte a risultati discordanti, a quale o a quali osservazioni prestar maggior credito? Afferma Salviati

per poter dalle loro osservazioni ritrar quel più di notizia che sia possibile, conveniente cosa è che noi gli applichiamo le minori e più vicine emende e correzioni che si possa, purché elle bastino a ritirar l'osservazioni dall'impossibilità alla possibilità. ([11], pp. 370-371)

Anche se Galileo non specifica cosa intenda per applicare le correzioni minori, da un'analisi dei suoi dati si può desumere che egli intendesse con ciò la minimizzazione dei valori assoluti delle deviazioni rispetto al valor vero ipotetico ([14], pp. 149-160). Concludendo la sua analisi egli osservò che la somma delle correzioni da apportare alle dieci coppie di osservazioni favorevoli a Chiaramonti perché esse dessero un valore coerente con la distanza della *Nova* dalla Terra era molto superiore a quella da portare alle dieci osservazioni a lui più sfavorevoli perché queste fornissero una posizione della *Nova* tra le stelle fisse, cioè nel firmamento. Da questo risultato egli traeva il convincimento che la propria ipotesi fosse più credibile di quella di Chiaramonti.

Il problema della teoria degli errori di cui vogliamo tracciare lo sviluppo storico si può riassumere con le parole del geodeta italiano Paolo Pizzetti (1860-1918) che aprono un lungo lavoro di rassegna sull'argomento

Il problema *della combinazione più vantaggiosa delle osservazioni* si presenta ogniqualvolta un sistema di ricerche sperimentali, aventi per oggetto la conoscenza di date *incognite fisiche*, fornisce all'osservatore *un numero di dati più che sufficiente* per la completa determinazione delle incognite. L'osservatore deve allora, per procedere al calcolo delle incognite, o escludere una parte dei dati che l'esperienza gli ha forniti, ovvero combinare questi dati in guisa che, pur tenendo conto di ciascuno di essi, i valori delle incognite restino determinati in modo unico e non dubbio. ([23], pp. 117-118)

Questa classe di problemi ebbe un notevole impulso nella seconda metà del XVIII secolo, sulla scorta di alcune ricerche dagli importanti risvolti pratici: la determinazione della longitudine terrestre in mare aperto e la determinazione dell'esatta forma della superficie terrestre. La mancanza di un metodo per determinare la longitudine in mare aperto era ritenuto, a torto o a ragione, uno dei maggiori pericoli nella navigazione e per questo, tutti gli stati che disponevano di ingenti flotte si interessarono a sovvenzionare ricerche che portassero alla soluzione del problema ([37], p. 12). Ridotto all'essenziale, il problema richiede di determinare l'ora locale, desumibile dalla posizione del sole, della luna o delle stelle e l'ora, allo stesso istante, in una posizione di riferimento, come può essere il meridiano di Greenwich: una differenza di quattro minuti corrisponde ad un grado di longitudine ad est o ad ovest, a seconda che l'ora locale sia in ritardo o in anticipo rispetto a quella di riferimento. Questo problema fu studiato, in particolare, dall'astronomo tedesco Tobias Mayer (1723-1762) con un metodo basato sulla determinazione della posizione di alcuni crateri lunari e di altre caratteristiche della superficie lunare ([10], pp. 11-15). Ad un certo punto Mayer dei suoi studi si trovò a dover determinare tre incognite in funzione di tre parametri affetti da errori sperimentali, avendo a disposizione un sistema di 27 equazioni lineari che ripartì in tre gruppi, ciascuno composto da nove equazioni, a seconda dei valori dei coefficienti moltiplicativi di una delle incognite, la latitudine lunare del cratere Manuzio ([36], pp. 16-25).

Il problema di riconoscere se la Terra fosse appiattita più ai poli che all'equatore è legato ad una celebre controversia che vide coinvolto l'astronomo italiano, naturalizzato francese, Dominique Cassini e che fu risolta da una spedizione in Lapponia guidata da Pierre Louis de Maupertuis (1698-1759) che confermò l'ipotesi newtoniana secondo cui la Terra è appiattita ai poli, come conseguenza della rotazione attorno al proprio asse. Considerando un arco di meridiano sufficientemente piccolo, la teoria prevedeva, nel caso in cui la forma della Terra fosse quella di un ellissoide oblato, una relazione affine tra la lunghezza  $a$  di un arco di  $1^\circ$  di meridiano centrato alla latitudine  $\theta$  e la funzione  $\sin^2 \theta$ :

$$a = z + y \sin^2 \theta, \quad (8.1)$$

dove  $z$  ed  $y$  rappresentano, rispettivamente, la lunghezza del grado di meridiano all'equatore e la differenza tra la lunghezza del grado di meridiano misurato al polo e all'equatore. Tramite le osservazioni dei valori di  $a$  e  $\theta$ , occorre risalire a  $y$  e  $z$  per poi calcolare, come misura dell'ellitticità della Terra, il rapporto  $\frac{3z}{y}$ .

Il gesuita dalmata Ruggero Boscovich (1711-1787), nato a Ragusa, oggi Dubrovnik, che insegnò anche presso l'Università di Pavia, con il confratello britannico Christopher Maire (1697-1767), ricevette da papa Benedetto XIV, il cardinale Prospero Lambertini, l'incarico di compiere rilievi geodetici per migliorare l'accuratezza della mappe dello stato Pontificio. Le misure comprendevano quella del grado di meridiano nei pressi di Roma e furono condotte tra il 1750 ed il 1752. Nel 1755 i risultati furono pubblicati nel *De literaria expeditione per pontificiam ditionem ad dimetiendos duos meridiani gradus et corrigendam mappam geographicam* ma i dettagli del metodo seguito da Boscovich per determinare i

parametri del modello furono pubblicati solo nel 1760 in una collocazione editoriale a dir poco inconsueta: un commento ad un poema in esametri latini di un conterraneo di Boscovich, Benedikt Stay [18]. Fortunatamente, la spiegazione del metodo, puramente geometrico, venne inserita come nota al termine della versione francese del *De literaria expeditione*, nel 1770 [4], dove fu notato da Laplace. Boscovich disponeva di  $n = 5$  misure  $(p_i, a_i)$ , rappresentate dai punti  $a, b, c, d, e$  nel piano cartesiano in Figura 8.1, le cui ascisse sono i valori di  $p := 10^4 \sin^2 \theta$  mentre in ordinata compaiono i valori di  $a$ . Il fattore  $10^4$  indica che i valori di  $\sin^2 \theta$  erano approssimati a quattro cifre decimali e dunque  $p$  aveva un valore intero. In assenza di errori di misura, una qualsiasi coppia di valori di  $p_i$  ed  $a_i$  permetterebbe di risolvere il sistema lineare (8.1) nelle incognite  $y$  e  $z$  che dovrebbero rendere una identità tutte le equazioni restanti di tipo

$$a_i = z + y \sin^2 \theta_i.$$

Questa verifica, effettuata per tutte le  $10 = \binom{5}{2}$  coppie di equazioni che si possono costruire con le misure a disposizione, non forniva invece una risposta univoca, per effetto degli errori  $\delta a_i$  sulle misurazioni di  $a_i$ , più rilevanti di quelli da cui era affetta la variabile  $p$  (cfr. Fig. 8.1). Dovrà allora essere

$$\delta a_i = z + y \sin^2 \theta_i - a_i$$

e Boscovich pose due condizioni sugli errori  $\delta a_i$ : anzitutto

$$\sum_{i=1}^n \delta a_i = 0 \tag{8.2}$$

che afferma sostanzialmente come si ritengano ugualmente probabili errori positivi e negativi. La seconda condizione impone che

$$\sum_{i=1}^n |\delta a_i| = \text{minimo}. \tag{8.3}$$

La condizione (8.2) equivale a dire che gli errori sono a media nulla e dunque la posizione del centro di massa  $G$  dei punti  $(p_i, a_i)$  è la stessa del centro di massa dei punti  $(p_i, a_i + \delta a_i)$ . Per rendere operativa la condizione (8.3) Boscovich considerò la retta verticale  $SGT$  che viene ruotata in verso orario attorno a  $G$ . Ora, quando questa retta è verticale,  $\sum_{i=1}^n |\delta a_i|$  è massima, perché infinita, ma quando inizia a ruotare in senso orario, essa diminuisce, almeno finché incontra il primo punto sperimentale: nel nostro caso  $e$ . Allo stesso modo, dopo che la retta ha incontrato l'ultimo punto— $c$  nella figura—la quantità (8.3) riprende a crescere fino a tornare al valore  $+\infty$  quando  $SGT$  torna ad essere verticale. Quando  $SGT$  si trova all'interno della nuvola di punti  $a, b, c, d$  ed  $e$  non è immediato ottenere l'andamento di (8.3) perché i vari addendi hanno comportamento diverso, a seconda che i punti corrispondenti siano già stati superati o meno dalla retta mobile  $SGT$ . Boscovich osserva che le *variazioni* subite dai  $\delta a_i$

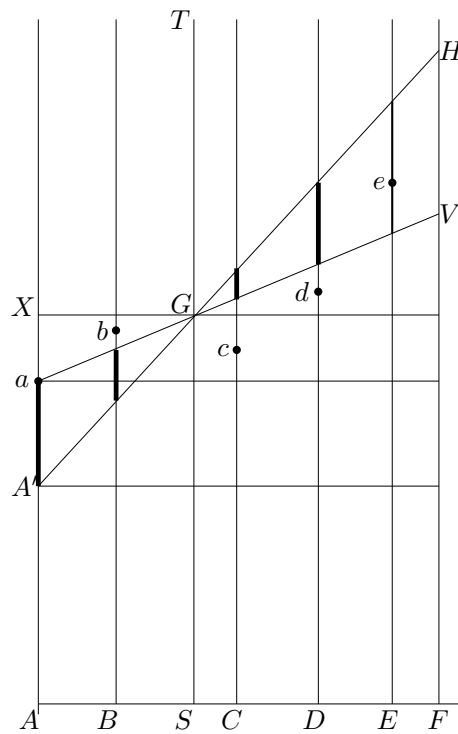


Figura 8.1: Illustrazione del metodo di Boscovich per trovare la retta che soddisfi alla condizione (8.3). Quando la retta mobile  $SGT$ , ruotando intorno a  $G$ , passa dalla posizione  $A'H$  alla posizione  $aV$ , gli errori  $\delta a_i$  subiscono delle variazioni, evidenziate con un tratto più spesso. Per similitudine, queste variazioni sono proporzionali alle lunghezze dei segmenti  $AS$ ,  $BS$ ,  $CS$ ,  $DS$  ed  $ES$ .

quando si considerano due rette per  $G$  vicine tra loro, sono *proporzionali* alle lunghezze dei segmenti  $AS, BS, CS, DS, ES$ , per la similitudine tra i triangoli aventi vertici in  $G$  e lati opposti le variazioni stesse. Ora, si determini l'ordine col quale la retta  $SGT$  incontra i punti  $a, b, c, d$  ed  $e$ , cioè  $e, a, c, d, b$ , e si considerino le lunghezze dei quei segmenti, tra  $AS, BS, CS, DS, ES$ , corrispondenti ai punti superati da  $SGT$  nella sua rotazione: finché tale somma non supera la metà della somma delle lunghezze di tutti i segmenti, occorre procedere nella rotazione mentre ci si arresta in caso contrario. Superato  $E$ , il segmento  $ES$  è inferiore a tale quantità e dunque occorre continuare a ruotare  $SGT$  fino ad intercettare il punto  $a$ : poiché  $AS + ES$  supera  $\frac{1}{2}(AS + BS + CS + DS + ES)$  è inutile procedere e la retta passante per  $a$  e per  $G$  è quella cercata. Osserviamo, in conclusione, che Boscovich non aggiunse sviluppi analitici che furono considerati da Laplace nel 1783 [17] e che egli si occupò diffusamente dei modi per ridurre gli errori di misura negli strumenti di osservazione [27].

Un approccio diverso fu seguito da Thomas Simpson (1710-1761), che si occupò a più riprese di calcolo delle probabilità e di matematica attuariale. In particolare, importa considerare per la teoria degli errori due suoi lavori: un articolo apparso nel 1756 [33] ed un capitolo di un volume pubblicato nel 1757 [34]. Simpson considerò una distribuzione di errori  $-v, -(v-1), \dots, -1, 0, 1, \dots, v-1, v$  che possono presentarsi con probabilità proporzionali rispettivamente a

$$1 \times r, 2r, 3r, \dots, vr, (v+1)r, vr \dots, 3r, 2r, 1 \times r$$

in modo da formare una distribuzione triangolare *discreta* e simmetrica intorno all'errore nullo cui compete la massima probabilità. Simpson si pose il problema di stabilire la probabilità che, prendendo la media aritmetica di un numero  $t$  assegnato di osservazioni, l'errore non superi  $\frac{m}{t}$ . Per questo egli si servì del metodo delle funzioni generatrici e costruì la funzione razionale<sup>1</sup>

$$f(r) := r^{-v} + 2r^{-v+1} + 3r^{-v+2} + \dots + (v+1)r^0 + \dots + 3r^{v-2} + 2r^{v-1} + r^v =$$

$$r^{-v} \left( \sum_{k=0}^v r^k \right)^2 = r^{-v} \frac{(1-r^{v+1})^2}{(1-r)^2}.$$

Questa quantità viene elevata alla  $t$ -esima potenza, così da avere il prodotto

$$f^t(r) = r^{-vt} (1-r^w)^n (1-r)^{-n}$$

dove si è posto  $n = 2t$  e  $w = v+1$ . Di questa funzione occorre calcolare il coefficiente di  $r^m$ , che rappresenta il numero di casi in cui l'eccesso del numero di errori positivi su quello degli errori negativi in  $n$  osservazioni è  $m$ . Attraverso calcoli diretti ma un po' laboriosi, Thomson ottenne un'espressione compatta di questo coefficiente da cui poté ricavare le probabilità

$$P \left( \frac{\sum_{i=1}^t \varepsilon_i}{t} = \frac{m}{t} \right) \quad \text{e} \quad P \left( \frac{\sum_{i=1}^t \varepsilon_i}{t} \leq \frac{m}{t} \right)$$

<sup>1</sup>Per convincersi del fatto che  $f(r) = r^{-v} \left( \sum_{k=0}^v r^k \right)^2$  si può sviluppare il quadrato della somma con il metodo usato da Bernoulli per trovare le frequenze dei punteggi ottenibili lanciando più dadi, discusso nella Sezione 1.6.

dove  $\varepsilon_i$  è l'errore da cui è affetta la misura  $i$ -esima. Nel volume del 1757 Simpson riconsiderò la stessa distribuzione triangolare per gli errori in una misurazione singola, estesa però al continuo: si tratta del primo esempio di distribuzione continua degli errori.<sup>2</sup> Gli stessi metodi di Simpson vennero usati, oltre dieci anni dopo, da Joseph-Louis Lagrange (1736-1813) in un lavoro di ampio respiro [15] in cui egli considerò diverse possibili distribuzioni degli errori, sia discrete che continue: in particolare, la distribuzione triangolare di Simpson vi compare, senza che però Lagrange lo citi. Lagrange considerò, tra gli altri, un esempio molto semplice, in cui si possono commettere solo errori di misura pari a 0, 1 e  $-1$  in modo che i casi in cui si possa avere errore nullo in una misura sono  $a$  e quelli in cui l'errore è unitario sono  $2b$ :  $b$  con errore  $+1$  ed altri  $b$  con errore  $-1$ <sup>3</sup>. Quando, in particolare  $a = b$ , la probabilità di non commettere errore in una singola misurazione è  $\frac{a}{a+2b} = \frac{1}{3}$  mentre, prendendo 2, 3, 4, 5 o 6 osservazioni e facendo la media aritmetica dei valori misurati, questa probabilità vale  $\frac{3}{9}$ ,  $\frac{7}{27}$ ,  $\frac{19}{81}$ ,  $\frac{51}{243}$ ,  $\frac{141}{729}$ , ... . Osserva Lagrange:

Si vede (...) che la probabilità di avere un errore nullo diminuisce a misura che si prende un numero maggiore di osservazioni cosicché, se si volesse stimare il vantaggio che si può avere nel prendere la media tra più osservazioni, tramite la differenza delle probabilità che l'errore sia nullo sul risultato medio e quella che sia nullo su ciascun risultato particolare, si troverebbe, nel caso trattato qui, che il vantaggio sarebbe sempre negativo, cioè si tratterebbe di uno svantaggio, che aumenterebbe col numero di osservazioni; da questo sembrerebbe potersi concludere che sarebbe meglio limitarsi ad un'unica osservazione, piuttosto che prendere la media tra più osservazioni: vi è però una considerazione essenziale da fare in proposito, dalla quale risulta che è sempre meglio nella pratica moltiplicare le osservazioni quanto più possibile.<sup>4</sup> ([15], p. 177)

L'osservazione di Lagrange è che non conta il confronto con il valore *puntuale* della probabilità di avere errore nullo sulla media, quanto quello con la probabilità di avere risultati compresi in un *intervallo* centrato intorno all'errore nullo. Lagrange chiarisce con un esempio: se il valore privo di errori fosse  $\varrho$ ,

<sup>2</sup>Per ulteriori dettagli sul lavoro di Simpson, si veda [32].

<sup>3</sup>In altri termini, le probabilità di avere errori 0 e  $\pm 1$  sono:

$$p(0) = \frac{a}{a+2b} \quad p(\pm 1) = \frac{b}{a+2b}.$$

<sup>4</sup>On voit (...) que la probabilité que l'erreur soit nulle diminue à mesure que l'on prend un plus grand nombre d'observations, de sorte que si l'on voulait estimer l'avantage qu'il peut y avoir à prendre le milieu entre plusieurs observations, par l'excès de la probabilité que l'erreur soit nulle dans le résultat moyen, sur celle que l'erreur soit aussi nulle dans chaque résultat particulier, on trouverait, dans le cas dont il s'agit ici, que l'avantage serait toujours négatif, c'est-à-dire qu'il se changerait en désavantage, lequel irait même en augmentant plus il y aurait d'observations; d'où il semble que l'on pourrait conclure que, dans ce cas, il vaudrait mieux s'en tenir à une observation unique, que de prendre le milieu entre plusieurs observations: mais il y a une considération essentielle à faire sur cette matière, de laquelle il résulte qu'il est toujours plus avantageux dans la pratique de multiplier des observations autant que l'on peut.

una singola osservazione potrebbe condurre a tre valori:  $\varrho - 1$ ,  $\varrho$  e  $\varrho + 1$ . Se si prendono due osservazioni dello stesso tipo e si fa la media aritmetica dei risultati ottenuti, lo spettro dei valori possibili è

$$\varrho - 1, \quad \varrho - \frac{1}{2}, \quad \varrho, \quad \varrho + \frac{1}{2}, \quad \varrho + 1$$

sicché vi sono errori più piccoli, pari a  $\pm \frac{1}{2}$ , che entrano in scena cambiando le conclusioni:

Dunque, per quanto la probabilità che l'errore sia nullo possa essere più piccola quando si prende il risultato medio di più osservazioni che non quando si prenda quello di ciascuna osservazione in particolare, tuttavia se si cerca la probabilità che l'errore non superi  $\frac{1}{2}$  o  $\frac{1}{3}$ , ..., si troverà che questa probabilità è maggiore nel primo che nel secondo caso.<sup>5</sup> ([15], p. 185)

Per dare sostanza a quanto asserito, Lagrange riprende l'esempio di prima e mostra come le probabilità di discostarsi per non più di  $\frac{1}{2}$  dal valore vero aumentano col numero di osservazioni. Tra i vari esempi di distribuzioni ipotetiche degli errori, Lagrange considerò anche distribuzioni *non* simmetriche ma in alcuni punti della sua discussione gli argomenti non sono chiari o contengono errori (p. 175 di [23]; pp. 596-598 di [21]). Lo statistico inglese Karl Pearson (1857-1936) sottolineò tuttavia [21] come Lagrange nel Problema VI di [15] si fosse avvicinato molto ad ottenere la distribuzione  $\chi^2$ , che sarebbe poi stata studiata nel 1875 dal geodeta tedesco Friedrich Helmert (1843-1917) e, soprattutto, dallo stesso Pearson nel 1900.

Un altro autore di questo periodo che è interessante menzionare per i contributi alla teoria degli errori è Johann Heinrich Lambert che pubblicò nel 1760 la *Photometria*, opera di ottica in cui Lambert propose metodi per misurare l'intensità della luce. Lambert non dà per scontato che la media aritmetica sia sempre la scelta migliore ed, anzi, espressamente menziona un caso in cui occorre prendere la media *geometrica* dei risultati delle misure, anziché quella aritmetica ([14], p. 131). Per Lambert però, quando gli errori positivi o negativi si bilanciano, la media diventa il valore vero, purché si abbia un numero infinito di osservazioni, grazie al teorema di Bernoulli. Quando gli esperimenti sono in numero finito  $N$ , Lambert si preoccupa di studiare l'impatto della misurazione che ha errore massimo. Se  $Q$  è il valore vero da misurare, la media  $M$  su  $N$  misurazioni è legata a  $Q$  da

$$M_N = \frac{\sum_{i=1}^N (Q + \varepsilon_i)}{N} = Q + \frac{\sum_{i=1}^N \varepsilon_i}{N},$$

gli errori sono indicati con  $\varepsilon_1 < \varepsilon_2 < \varepsilon_3 < \dots < \varepsilon_{N-1} < \varepsilon_N$ , con  $\varepsilon_N \gg |\varepsilon_i|$  per tutti gli  $i \neq N$ . Anche se errori grandi sono più rari, è bene scartare questi

<sup>5</sup>Ainsi, quoique la probabilité que l'erreur soit nulle puisse être plus petite lorsqu'on prend le résultat moyen de plusieurs observations que lorsqu'on prend le résultat de chaque observation en particulier, cependant, si on cherche la probabilité que l'erreur ne surpasse pas  $\frac{1}{2}$ , ou  $\frac{1}{3}$ , ..., on trouvera que cette probabilité sera plus grande dans le premier cas que dans le second.

risultati, per ridurre tra  $Q$  e la media. Infatti, senza l'osservazione fuori scala, affetta da errore  $\varepsilon_N$ , si avrebbe una media

$$M_{N-1} = Q + \frac{\sum_{i=1}^{N-1} \varepsilon_i}{N-1} = Q + \frac{A}{N-1},$$

avendo cura di definire  $A := \sum_{i=1}^{N-1} \varepsilon_i$ . Ora, poiché  $\varepsilon_N$  è l'errore massimo si ha

$$\frac{A}{N-1} = \frac{\sum_{i=1}^{N-1} \varepsilon_i}{N-1} < \frac{(N-1)\varepsilon_N}{N-1} = \varepsilon_N$$

da cui segue

$$(N-1)\varepsilon_N - A > 0$$

ovvero, aggiungendo  $AN$  ad ambo i termini,

$$(N-1)(A + \varepsilon_N) > AN$$

e quindi

$$\frac{A}{N-1} < \frac{A + \varepsilon_N}{N} \quad \text{ovvero} \quad M_{N-1} - Q < M_N - Q$$

che dimostra come la media si sia accostata maggiormente al valore vero  $Q$  scartando il valore molto maggiore degli altri.

Lambert si pose poi il problema di fornire un metodo per determinare un valore medio di una quantità che, misurata più volte, è stata oggetto di errori che si sono presentati con frequenze *diverse*. La scelta di questo valore medio deve essere tale

che risulti massima la probabilità che essa [la media] abbia una minima discrepanza rispetto al valore vero, tra tutte le possibili scelte.<sup>6</sup> ([14], p. 140)

Lambert costruì una curva la cui ascissa rappresenta la quantità da misurare, con il segmento  $AC$  (*cfr.* Fig. 8.2) che rappresenta il vero valore, mentre  $BC$  e  $CD$  sono gli errori massimi da cui una misura può essere affetta e  $CQ$ ,  $CP$ ,  $CR$  e  $CS$  sono gli altri errori che si sono presentati con frequenze (*vices*), rispettivamente,  $QM$ ,  $PN$ ,  $RL$  e  $SK$ , ordinate che appartengono alla curva continua  $BNMLKD$ .

Questi segmenti sono, più precisamente, le frequenze vere, che Lambert distingue da quelle osservate. Per avere coincidenza tra queste frequenze occorrerebbe per Lambert che l'esperimento fosse ripetuto infinite volte. Si supponga che in un esperimento condotto un numero di volte pari a  $n+m+\ell+k$  si siano presentati:  $n$  volte il valore  $AP$ ,  $m$  volte il valore  $AQ$ ,  $\ell$  volte il valore  $AR$  e  $k$  volte il valore

<sup>6</sup>ut maxime probabile sit, eam a vera omnium minime discrepare

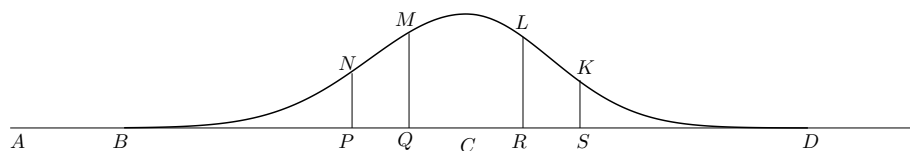


Figura 8.2: Profilo della curva degli errori ipotizzato da Lambert nella *Phorometria*.

AS. Lambert considera le frequenze vere corrispondenti  $PN$ ,  $QM$ ,  $RL$  ed  $SK$  e, implicitamente, sviluppa il quadrinomio

$$(PN + QM + RL + SK)^{n+m+\ell+k}$$

di cui determina il generico termine

$$\frac{(n + m + \ell + k)!}{n!m!\ell!k!} PN^n QM^m RL^\ell SK^k$$

che rappresenta il numero diverso di volte in cui gli errori scelti si possono manifestare in  $n + m + \ell + k$  osservazioni. Ora Lambert, lavorando a valori di  $n$ ,  $m$ ,  $\ell$  e  $k$  fissati, suppone che gli errori osservati possano variare o, meglio, che  $CQ$  vari e di conseguenza anche gli altri errori  $CP$ ,  $CR$  e  $CS$ , ipotizzati tutti equidistanti tra loro. Il criterio da seguire per determinare  $PN$ ,  $QM$ ,  $RL$  ed  $SK$  è il seguente:

Poiché in generale la probabilità massima compete al caso che compare con massima frequenza, ne consegue che deve essere<sup>7</sup>

$$PN^n QM^m RL^\ell SK^k = \text{numero massimo.} \quad (8.4)$$

([14], p. 143)

Per questo motivo, Lambert può essere visto come colui che introdusse per la prima volta un argomento di *massima verosimiglianza* [17]. Infatti, ([7], p. 9) la verosimiglianza  $L(H|R)$  di un'ipotesi  $H$  sulla base di un insieme di dati  $R$  è proporzionale alla probabilità  $p(R|H)$  di osservare i dati  $R$ , supponendo verificata l'ipotesi  $H$ :

$$L(H|R) \propto p(R|H).$$

Per procedere, Lambert calcolò la derivata logaritmica di (8.4) ottenendo la condizione

$$n \frac{d(PN)}{PN} + m \frac{d(QM)}{QM} + \ell \frac{d(RL)}{RL} + k \frac{d(SK)}{SK} = 0.$$

<sup>7</sup>Cum vero iam in genere iste casus maxime sit probabilis, qui omnium frequentissime occurrit, consequens est, debere esse

$$PN^n QM^m RL^\ell SK^k = \text{numero maximo.}$$

L'operazione si può intendere in questo modo (*cfr.* Fig. 8.2). Siccome i punti  $P$ ,  $Q$ ,  $R$  ed  $S$  sono punti del grafico di una funzione  $\varphi(x)$  che ha un massimo in un punto incognito  $C$ , corrispondente al valore  $x = x_C$  dell'ascissa, contata a partire da  $A$ , dette  $x_P$ ,  $x_Q$ ,  $x_R$  ed  $x_S$  le ascisse dei punti corrispondenti alle misurazioni, (8.4) si può riscrivere come

$$\varphi^n(x_P - x_C)\varphi^m(x_Q - x_C)\varphi^\ell(x_R - x_C)\varphi^k(x_S - x_C)$$

e la condizione di stazionarietà diventa

$$n \frac{\varphi'(x_P - x_C)}{\varphi(x_P - x_C)} + m \frac{\varphi'(x_Q - x_C)}{\varphi(x_Q - x_C)} + \ell \frac{\varphi'(x_R - x_C)}{\varphi(x_R - x_C)} + k \frac{\varphi'(x_S - x_C)}{\varphi(x_S - x_C)} = 0$$

dove  $\varphi'$  indica la derivata rispetto all'argomento di  $\varphi(x)$ . Questa condizione assume un significato geometrico se si osserva che, per una funzione  $y(x)$ ,  $\frac{y(x)}{y'(x)}$  è la lunghezza, a meno del segno, del segmento dell'asse delle ascisse compreso tra il generico punto  $x$  e l'intersezione con l'asse delle ascisse della tangente in  $(x, y(x))$  alla curva: tale segmento era noto come *sottotangente* della funzione  $y(x)$ . Dunque, se  $\nu$ ,  $\mu$ ,  $\lambda$  e  $\kappa$  sono le sottotangenti alla curva degli errori  $\varphi(x)$ , la stazionarietà imporrebbe

$$\frac{n}{\nu} + \frac{m}{\mu} + \frac{\ell}{\lambda} + \frac{k}{\kappa} = 0.$$

Come commento, Lambert osservò che, se  $\varphi(x)$  fosse stata simmetrica rispetto al punto  $C$ , il valore medio ottenuto con questo metodo coinciderebbe piuttosto spesso (*saepius*) con la media aritmetica. Come esempio, se vi fossero due sole osservazioni e fosse  $n = \ell = 1$ ,  $m = k = 0$ , la stazionarietà si ridurrebbe a

$$\frac{1}{\mu} + \frac{1}{\lambda} = 0$$

che richiede che  $AC$  sia la media aritmetica tra  $AR$  e  $AQ$ .

Anche Daniel Bernoulli si occupò di teoria degli errori mettendo in discussione la scelta sistematica della media aritmetica di più osservazioni come valore più attendibile. Egli assunse il seguente principio guida per scegliere il modo migliore con cui combinare i risultati delle osservazioni ([2], pp. 5-6):

Ritengo poi che occorra scegliere, tra gli infiniti modi in cui gli errori delle osservazioni possono aver luogo, quello che è dotato del massimo grado di probabilità, relativamente all'intero complesso delle osservazioni.<sup>8</sup>

Riferendosi poi all'esempio di un arciere che scocca una freccia diretta verso un bersaglio idealizzato in un segmento verticale, in termini suggestivi Bernoulli parla del bersaglio come di un centro di forza che attrae le frecce ed espone il problema principale della teoria degli errori in questi termini: ([2], p. 7)

<sup>8</sup>Existimo autem, quod inter infinitos modos, quibus observationum errores contingere potuerunt, is seligendus sit, qui maximo probabilitatis gradu, pro integro observationum complexu, donatus fuerit.

occorre determinare la posizione più probabile del bersaglio dalla conoscenza delle posizioni di un certo numero di colpi.<sup>9</sup>

Volendo trovare la distribuzione di probabilità degli errori (*scala probabilitatum*), Bernoulli ripete come altri prima di lui la richiesta che la distribuzione sia simmetrica rispetto alla posizione del valore vero e che le probabilità di commettere errori piccoli siano inferiori rispetto a quelle di commettere errori rilevanti. Egli insiste però sulla richiesta che la probabilità di commettere errori superiori ad un certo limite deve essere nulla e sugli effetti che questa condizione ha sulla forma della legge degli errori ([2], pp. 8-9):

Infine, poiché gli errori massimi da ambo le parti fungono da limiti tra quanto può e quanto non può accadere, è bene che l'ultima particella della scala di probabilità tenda rapidamente da ambo le parti verso la linea su cui sono collocati i punti osservati, in modo che le sue tangenti siano pressoché perpendicolari a questa linea ed in modo che la stessa scala indichi come impossibile il realizzarsi di un passaggio oltre i limiti ipotizzati; questa condizione tuttavia non richiede pieno rigore, purché i limiti agli errori non siano stati posti con troppa confidenza.<sup>10</sup>

La serie di proprietà richieste alla scala di probabilità è compatibile con la scelta di una semi-ellisse o di una semicirconferenza centrata sull'asse delle ascisse cui Bernoulli si attiene per semplicità osservando che

un'ellisse costruita sul medesimo asse servirà allo scopo, per quanto allungata o compressa essa sia, ciò che indica come non dobbiamo essere troppo preoccupati di avere un'accurata descrizione della scala.<sup>11</sup>

Scelta una semicirconferenza come curva degli errori, il raggio  $r$  è dato dal massimo errore che può essere commesso mentre la posizione del centro va localizzata con un argomento di massima verosimiglianza. Se le misure ottenute sono disposte in ordine crescente a valgono  $A, A + a, A + b, A + c, \dots$ , detta  $x$  la distanza da  $A$  del centro della circonferenza, la probabilità di osservare  $A$  è proporzionale a  $\sqrt{r^2 - x^2}$ , quella di osservare  $A + a$ , punto che dista  $x - a$  dal centro della semicirconferenza è proporzionale a  $\sqrt{r^2 - (x - a)^2}$ , e così via per gli altri valori osservati. La probabilità di osservare tutti quanti i valori ricavati dalle misure è allora proporzionale a

$$L(x) := \sqrt{r^2 - x^2} \sqrt{r^2 - (x - a)^2} \sqrt{r^2 - (x - b)^2} \sqrt{r^2 - (x - c)^2} \dots$$

<sup>9</sup>determinanda est positio metae quam maxime probabilis ex cognitis aliquot ictuum locis.

<sup>10</sup>Denique ex eo, quod ambae aberrationes maximae tanquam limites censentur inter id, quod contingere potest et quod fieri nequit, oportet ut ultima scalae particula, ab utroque latere, praecipitanter tendat ad lineam, in qua puncta observata locantur, sic ut tangentes in punctis extremis fiant ad eandem lineam propemodum perpendiculares et ut ipsa scala indicet, transgressum fieri vix posse ultra limites suppositos; nec tamen haec conditio omnem requireret rigorem, si modo haud nimis confidenter erroribus limites posueris.

<sup>11</sup>utcumque elongata vel compressa fuerit ellipsis super eodem axe constructa, idem praestabit officium, quod indicat non esse, ut nimis simus solliciti de accurata scalae descriptione

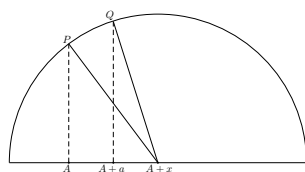


Figura 8.3: Profilo semicircolare ipotizzato da Daniel Bernoulli come possibile curva degli errori.

e la massimizzazione di questa funzione fornirà il valore di  $x$  cui compete la massima probabilità e che andrà assunto come valore più attendibile della grandezza misurata. Bernoulli in effetti massimizzò  $L^2(x)$  ed illustrò, con qualche esempio, delle situazioni in cui la regola di massimizzazione di  $L(x)$  conduceva a valori diversi dalla media aritmetica delle osservazioni. Bernoulli si mostrava molto cauto sulle ipotesi più idonee da assumere per affrontare la soluzione del problema ed osservava come, nel caso di tre sole osservazioni, la massimizzazione di  $L(x)$  portasse già ad un'equazione algebrica di quinto grado. Notiamo infine come egli fosse ben consapevole dell'arbitrarietà nella scelta della semicirconferenza come legge degli errori e propose come alternativa quella di una parabola, senza però una significativa riduzione delle difficoltà computazionali.

## 8.2 Primi contributi originali di Laplace

Laplace si occupò di teoria degli errori accidentali già nel lavoro sulla probabilità inversa [16] del 1774, esaminato nel Capitolo 8. Qui egli considerò questo problema (*Problème III*):

Determinare la media che occorre prendere di tre osservazioni date di uno stesso fenomeno.<sup>12</sup> ([16], p. 42)

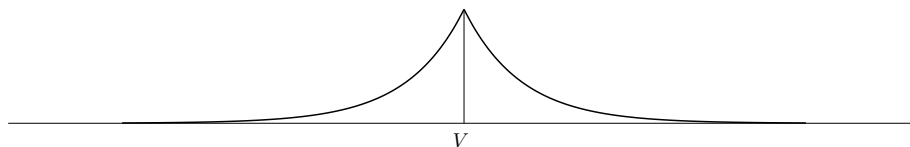


Figura 8.4: Profilo della curva degli errori  $\varphi(x)$  trovato da Laplace nel 1774 a partire dalla formula di Bayes.

Laplace considerò un fenomeno non specificato che avviene ad un certo istante di tempo, misurato in secondi. Le tre osservazioni fissano il presentarsi del fenomeno a tre istanti diversi:  $a$ ,  $b$  e  $c$  con  $a$  e  $b$  separati da  $p$  secondi e  $b$  e  $c$  da  $q$  secondi. Si domanda quale punto dell'asse dei tempi si possa assumere come valore dell'istante in cui si è verificato il fenomeno. Laplace indicò con  $V$  il vero istante in cui il fenomeno in esame ha luogo ed enunciò le proprietà principali della funzione  $\varphi(x)$  che rappresenta la probabilità che una misura sperimentale si allontani di  $x$  da  $V$ , sostanzialmente analoghe a quelle già incontrate in altri autori:

1. La curva  $y = \varphi(x)$  deve essere divisa in parti del tutto simili (*entièrement semblables*) dalla retta passante per il punto  $V$  dell'asse delle ascisse, dal momento che gli scarti dal vero per eccesso o per difetto sono egualmente probabili;
2.  $y = \varphi(x)$  tende asintoticamente all'asse delle ascisse dal momento che la probabilità di allontanarsi dal vero di una quantità infinita è nulla;
3. L'area compresa tra il grafico di  $y = \varphi(x)$  e l'asse delle ascisse ha valore 1.

Detta  $x$  la distanza temporale di  $a$  da  $V$ , Laplace scrisse la probabilità di ottenere i valori  $a$ ,  $b$  e  $c$  nella forma (*cf.* Fig. 8.5)

$$y(x) := \varphi(x)\varphi(p-x)\varphi(p+q-x),$$

<sup>12</sup>Déterminer le milieu que l'on doit prendre entre trois observations données d'un même phénomène.

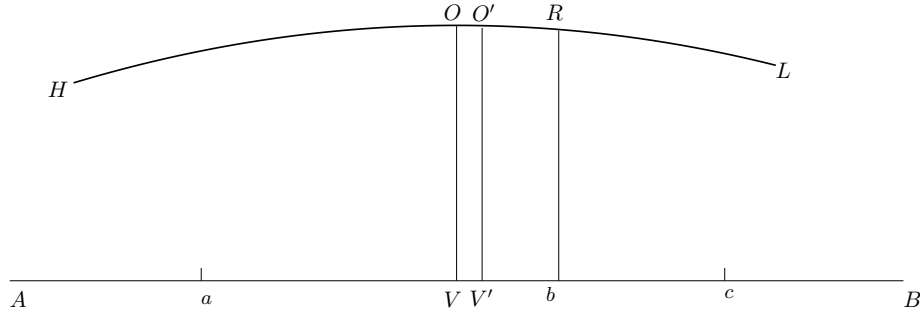


Figura 8.5: Profilo della curva  $y(x)$  che fornisce la probabilità affinché tre osservazioni, separate da  $p$  e  $q$  secondi tra loro, abbiano luogo se il vero valore è  $V$ .

supponendo tra loro indipendenti le misure. Se l'istante vero in cui il fenomeno si realizza fosse  $V'$ , a distanza temporale  $x'$  da  $a$ , questa probabilità si trasformerebbe in

$$y(x') := \varphi(x')\varphi(p-x')\varphi(p+q-x').$$

Le funzioni così costruite rappresentano, rispettivamente, le probabilità che le osservazioni avvengano agli istanti  $a$ ,  $b$ ,  $c$ , nell'ipotesi che  $V$  o  $V'$ , rispettivamente, siano i veri istanti in cui il fenomeno si verifica:

$$\frac{\varphi(x)\varphi(p-x)\varphi(p+q-x)}{\varphi(x')\varphi(p-x')\varphi(p+q-x')} = \frac{P(a, b, c|V)}{P(a, b, c|V')} = \frac{P(p, q|V)}{P(p, q|V')}.$$

Grazie al Postulato che Laplace aveva posto alla base dello studio della probabilità inversa che abbiamo illustrato nella sezione 8.2, si ha anche

$$\frac{\varphi(x)\varphi(p-x)\varphi(p+q-x)}{\varphi(x')\varphi(p-x')\varphi(p+q-x')} = \frac{P(V|a, b, c)}{P(V'|a, b, c)} = \frac{P(V|p, q)}{P(V'|p, q)}.$$

Laplace considerò due possibili medie: la prima (*milieu de probabilité*) è l'istante tale che le probabilità che  $V$  accada prima o dopo di esso sono uguali. La sua ascissa  $\eta$  verifica l'uguaglianza

$$\int_{-\infty}^{\eta} y(x)dx = \int_{\eta}^{+\infty} y(x)dx.$$

La seconda media—(*milieu d'erreur* o *milieu astronomique*)—è tale da minimizzare la somma degli errori in valore assoluto che si commettono prendendo tale valore come valore vero, moltiplicati per le rispettive probabilità. L'ascissa  $\mu$  di questa seconda media è tale da minimizzare la funzione

$$g(\mu) := \int_{-\infty}^{\infty} |x - \mu|y(x)dx.$$

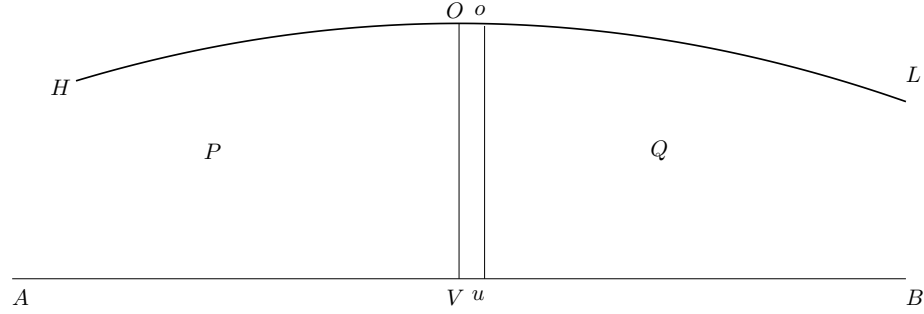


Figura 8.6: Coincidenza tra media di probabilità e media astronomica.

Laplace (*cf.* Fig. 8.6) impartì a questa media, indicata con  $V$ , una variazione  $Vu = d\mu$  dell'ascissa ed indicò con  $P$  e  $Q$  i centri di massa delle porzioni  $VOH$  e  $uoL$ , le cui distanze da  $OV$  sono indicate con  $z'$  e  $z$ , rispettivamente.

Se  $M$  ed  $N$  sono le aree sottese da  $uoL$  e  $VOH$ , la funzione  $g(\mu)$  è

$$g(\mu) = Nz + Nz' + \frac{1}{2}ydx^2$$

nella prima posizione di  $V$ , mentre è

$$g(\mu + d\mu) = M(z - d\mu) + N(z' + d\mu) + \frac{1}{2}y d\mu^2.$$

quando  $V$  si è spostato in  $u$ , variando di  $d\mu$  la sua ascissa. Infatti abbiamo

$$\begin{aligned} g(\mu) &:= \int_{-\infty}^{\infty} |x - \mu|y(x)dx = \\ &= \int_{-\infty}^{\mu} (\mu - x)y(x)dx + \int_{\mu}^{\mu+d\mu} (x - \mu)y(x)dx + \int_{\mu+d\mu}^{\infty} (x - \mu)y(x)dx \\ &= Nz' + \frac{1}{2}d\mu \cdot y(\mu)d\mu + Nz = Nz' + Nz + \frac{1}{2}y(\mu)(d\mu)^2 \\ g(\mu + d\mu) &:= \int_{-\infty}^{\infty} |x - \mu - d\mu|y(x)dx = \\ &= \int_{-\infty}^{\mu} (\mu + d\mu - x)y(x)dx + \int_{\mu}^{\mu+d\mu} (\mu + d\mu - x)y(x)dx + \int_{\mu+d\mu}^{\infty} (x - \mu - d\mu)y(x)dx \\ &= Nz' + Nd\mu - \frac{1}{2}d\mu \cdot y(\mu)d\mu + d\mu y(\mu)d\mu + Nz - Md\mu = \\ &= N(z' + d\mu) + M(z - d\mu) + \frac{1}{2}y(\mu)(d\mu)^2 \end{aligned}$$

Imponendo la condizione di stazionarietà  $g(\mu + d\mu) = g(\mu)$ , Laplace ottenne  $(M - N)d\mu = 0$  che mostra come le due definizioni di media debbano coincidere, dovendo essere  $M = N$ .

Non potendo scegliere  $\varphi(x)$  costante perché in contraddizione con le richieste fatte alle leggi di errore ammissibili, Laplace suppose che le variazioni subite dalle differenze di  $\varphi(x)$  dovessero seguire la stessa legge di  $\varphi(x)$ , sulla base del principio di ragione insufficiente:

poiché non abbiamo alcun motivo di supporre una legge per le ordinate ed un'altra per le loro differenze.<sup>13</sup> ([16], pp. 45-46)

Grazie a questa opinabile ipotesi, Laplace pose

$$\frac{d\varphi(x+dx)}{d\varphi(x)} = \frac{\varphi(x+dx)}{\varphi(x)}$$

da cui segue che  $\frac{d\varphi(x+dx)}{\varphi(x+dx)} = \frac{d\varphi(x)}{\varphi(x)}$  che si può vedere come condizione affinché il differenziale di  $\ln \varphi(x)$  sia *costante*, per cui

$$\frac{d\varphi(x)}{dx} = -m\varphi(x)$$

dove  $m > 0$  è una costante ancora incognita. Poiché  $\varphi(x)$  deve essere simmetrica rispetto all'asse  $x = 0$  in virtù del requisito 1, Laplace conclude che  $\varphi(x) = be^{-m|x|}$  dove la costante  $b$  è determinata grazie alla condizione 3 come  $b = \frac{m}{2}$ , da cui segue la proposta di legge degli errori

$$\varphi(x) = \frac{m}{2}e^{-m|x|}. \quad (8.5)$$

Con questa legge, Laplace trova l'ascissa  $\mu$  del valore vero  $V$  imponendo che

$$\int_{-\infty}^{\mu} y(t)dt = \int_{\mu}^{\infty} y(t)dt$$

dove

$$y(t) = \frac{m^3}{8}e^{-m[|t|+|p-t|+|p+q-t|]}.$$

Il calcolo richiede qualche attenzione perché l'espressione di  $y(t)$  muta a seconda della posizione del punto generico, a causa dei valori assoluti, per i quali Laplace non ha una notazione propria. La posizione di  $\mu$ , nell'ipotesi  $p > q$ , è

$$\mu = p + \frac{1}{m} \ln \left( 1 + \frac{1}{3}e^{-mp} - \frac{1}{3}e^{-mq} \right)$$

che si riduce alla media aritmetica dei tre valori  $0$ ,  $p$  e  $p+q$ , cioè  $\mu = \frac{2p+q}{3}$ , solo nel caso in cui  $m \ll 1$ . Per ottenere il valore di  $m$ , Laplace fa ancora ricorso al principio della probabilità inversa per cui il rapporto tra le probabilità che esso abbia il valore  $m$  piuttosto che  $m'$ , sapendo che le osservazioni hanno distanze tra loro  $p$  e  $q$  secondi, è uguale al rapporto tra le probabilità che, noti i valori

<sup>13</sup>comme nous n'avons aucune raison de supposer une autre loi aux ordonnées qu'à leurs différences.

del parametro,  $m$  oppure  $m'$ , le tre osservazioni abbiano tra loro le distanze  $p$  e  $q$  secondi. La complessa analisi di questa condizione conduce Laplace ad un'equazione algebrica di grado 15 ed è inficiata da un errore che rende non valida la conclusione [35]. La legge proposta da Laplace è legata alla *mediana* di un insieme di osservazioni e non alla media aritmetica e, a dispetto delle fragili basi su cui poggia la sua deduzione, ha trovato e trova molte applicazioni in statistica dove è nota col nome di distribuzione di Laplace o distribuzione esponenziale doppia. Alcuni anni più tardi, nel 1781, Laplace abbandonerà la legge degli errori proposta nel 1774.

### 8.3 La nascita del metodo dei minimi quadrati

Nel 1805, Adrien-Marie Legendre (1752-1833) pubblicò un'opera, *Nouvelles méthodes pour la détermination des orbites des comètes* [20] contenente una breve appendice, quattro pagine in tutto, in cui proponeva il metodo dei minimi quadrati per determinare  $m$  incognite  $x, y, z, \dots$  legate tra loro da equazioni lineari del tipo

$$E_i = a_i + b_i x + c_i y + f_i z + \dots \quad i = 1, \dots, n$$

in cui i coefficienti  $a_i, b_i, c_i, f_i \dots$  sono noti, anche per effetto di misurazioni, ed  $E_i$  rappresentano gli errori legati alle misure delle variabili

che bisogna determinare tramite la condizione che il valore di<sup>14</sup>  $E$  si riduca, per ciascuna equazione, ad una quantità nulla o molto piccola.<sup>15</sup> ([20], p. 72)

Idealmente, in assenza di errori, si dovrebbe avere  $E_i = 0$  per tutti  $i$ , ma ciò non è.

Legendre osservò anche come il problema non avrebbe alcuna difficoltà se il numero delle equazioni e quello delle incognite coincidessero mentre il problema si pone nel caso praticamente importante in cui  $m < n$ . Egli considerò

$$\sum_{i=1}^n E_i^2 = E$$

e, siccome

$$\frac{\partial E}{\partial x} = 2 \sum_{i=1}^n E_i \frac{\partial E_i}{\partial x} = 0 \quad \frac{\partial E_i}{\partial x} = a_i,$$

si ottengono  $m$  condizioni di stazionarietà

$$\frac{\partial E}{\partial x} = \sum a_i b_i + x \sum b_i^2 + y \sum b_i c_i + z \sum b_i f_i + \dots = 0,$$

<sup>14</sup>Si legga  $E_i$  al posto di  $E$  in questo punto.

<sup>15</sup>qu'il faut déterminer par la condition que la valeur de  $E$  se réduise, pour chaque équation, à une quantité ou nulle ou très-petite

$$\frac{\partial E}{\partial y} = \sum a_i c_i + x \sum b_i c_i + y \sum c_i^2 + z \sum c_i f_i + \dots = 0$$

fino ad esaurire tutte le  $m$  variabili. I valori ottenuti per le incognite  $x, y, \dots$  risolvono il problema proposto anche se Legendre non dimostrò in generale che si trattava di un minimo della funzione  $E$ . Affermò correttamente che si tratta di un minimo nel caso particolare in cui c'è una sola incognita  $x$  e più misure  $a_i$  perché allora la quantità

$$E = \sum_{i=1}^n (x - a_i)^2$$

è in effetti minimizzata da

$$\bar{x} = \frac{a_1 + a_2 + \dots + a_n}{n}$$

dal momento che

$$\sum_{i=1}^n (x - a_i)^2 = \sum_{i=1}^n (x - \bar{x} + \bar{x} - a_i)^2 = \sum_{i=1}^n (\bar{x} - a_i)^2 + n(x - \bar{x})^2 + 2(x - \bar{x}) \sum_{i=1}^n (\bar{x} - a_i) \geq \sum_{i=1}^n (\bar{x} - a_i)^2$$

per la nota proprietà della media aritmetica  $\sum_{i=1}^n (\bar{x} - a_i) = 0$ . Legendre considerò, come esempio, la determinazione delle coordinate  $(x, y, z)$  di un punto  $P$  che vengono ottenute più volte fornendo una successione di terne  $(a_i, b_i, c_i)$ . Costruendo

$$\sum_{i=1}^n [(x - a_i)^2 + (y - b_i)^2 + (z - c_i)^2],$$

Legendre trovò come soluzione le coordinate del centro di massa degli  $n$  punti osservati  $(a_i, b_i, c_i)$ :

$$x = \frac{a_1 + a_2 + \dots + a_n}{n} \quad y = \frac{b_1 + b_2 + \dots + b_n}{n} \quad z = \frac{c_1 + c_2 + \dots + c_n}{n}.$$

Legendre, che applicò il metodo dei minimi quadrati anche ad un problema sostanzialmente equivalente a quello considerato da Boscovich, *non* giustificò dal punto di vista probabilistico il metodo proposto.

## 8.4 Minimi quadrati e legge degli errori: Robert Adrain

Una prima giustificazione venne data nel 1808 dal matematico statunitense, irlandese di nascita, Robert Adrain (1775-1843) il cui lavoro però non fu notato dalla comunità matematica che intorno al 1870, molto tempo dopo la sua morte. Adrain fondò una rivista, *The Analyst*, che non ebbe una lunga vita, sulle cui pagine propose per la soluzione un problema proposto da un docente dell'università della Pennsylvania, Robert Patterson, in cui si chiedeva di trovare l'area di una regione piana poligonale *chiusa*  $ABCDE$  che, per effetto degli errori nelle misure delle lunghezze dei lati e delle inclinazioni rispetto ad una direzione data, non si chiudeva, sicché  $E \neq A$ :

si richiede il calcolo dell'area di questa regione chiusa nell'ipotesi più probabile circa questo errore.<sup>16</sup> (*The Analyst*, (1808), p. 42)

Una prima risposta al quesito di Patterson, che aveva messo in palio 10 dollari da destinare a chi avesse per primo risolto il problema, fu presentata da Nathanael Bowditch (1773-1838), studioso di matematica applicata, in particolare di astronomia, che tradusse e commentò in inglese i primi quattro volumi della *Mécanique Celeste* di Laplace. Egli partì da due principii:

- a) gli errori da cui sono affette misure di lunghezza sono da ritenersi proporzionali alle lunghezze dei segmenti misurati;
  - b) sulle inclinazioni dei lati del poligono si possono commettere errori uguali.
- Bowditch non giustificò queste ipotesi ma fornì un metodo geometrico ed uno

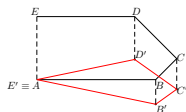


Figura 8.7: Metodo di Bowditch per correggere gli errori di misura in modo da recuperare la chiusura di un poligono.

<sup>16</sup>it is required to compute the area of this enclosure on the most probable supposition of this error.

aritmetico per correggere gli errori nelle misure. Il metodo geometrico consisteva nel formare il quoziente

$$r := \frac{|AE|}{|AB| + |BC| + |CD| + |DE|}$$

e nel tracciare, dai vertici  $B$ ,  $C$ ,  $D$ , ed  $E$ , i segmenti  $BB' = r|AB|$ ,  $CC' = r(|AB| + |BC|)$ ,  $DD' = r(|AB| + |BC| + |CD|)$  ed  $EE' = r(|AB| + |BC| + |CD| + |DE|) = |AE|$  paralleli ad  $AE$  e diretti nel verso che porta  $E$  a sovrapporsi ad  $E'$ , cioè ad  $A$ . I punti  $A$ ,  $B'$ ,  $C'$ ,  $D'$  ed  $E' = A$  ora trovati rappresentano i vertici del contorno che si richiude su se stesso (Fig. 9.7).

Pubblicata la soluzione di Bowditch, che si aggiudicò il compenso di 10 dollari, Adrain scrisse una nota abbastanza estesa sullo stesso soggetto, dove compare subito la probabilità:

La questione di cui propongo la soluzione è questa: siano  $Ab$   $Bb$   $AB$  il vero valore di una qualsiasi quantità, di cui  $Ab$  è la misura ottenuta tramite osservazione od esperimento e  $Bb$  è l'errore; qual è l'espressione della probabilità che l'errore  $Bb$  si presenti quando si misura  $AB$ ?<sup>17</sup> ([1], p. 93)

Adrain fornì due dimostrazioni della legge degli errori, di cui qui riportiamo la seconda, supportata da un argomento geometrico, considerando i possibili errori  $x$  sulla lunghezza ed  $y$  sull'orientazione per un dato segmento  $AB$ . Gli errori  $(x, y)$  sono interpretati come coordinate cartesiane ortogonali, centrate in  $B$ .<sup>18</sup> In primo luogo Adrain rappresentò errori *equiprobabili* sulla lunghezza ( $Bb$ ) e

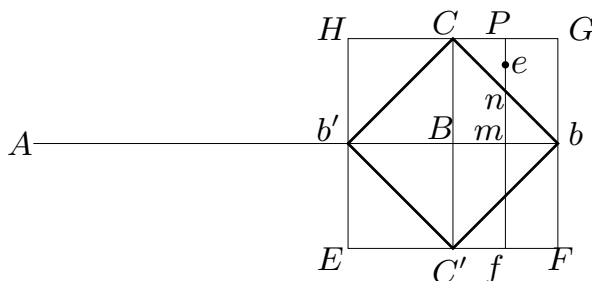


Figura 8.8: Metodo geometrico di Adrain per ottenere la legge degli errori.

sull'inclinazione ( $BC$ ) cui debbono corrispondere altri due errori equiprobabili tra loro e con i precedenti ( $Bb'$ ) e ( $BC'$ ). I punti  $bCb'C'$  individuano un quadrato cui fa riscontro il quadrato  $EFHG$  di lato  $bb' = CC'$ . Adrain intendeva trovare la curva di uguale probabilità per gli errori, cioè tale che

<sup>17</sup>The question which I propose to resolve is this: Supposing  $Ab$   $Bb$   $AB$  to be the true value of any quantity, of which the measure by observation or experiment is  $Ab$ , the error being  $Bb$ ; what is the expression of the probability that the error  $Bb$  happens in measuring  $AB$ ?

<sup>18</sup>A ben vedere i segmenti  $bb'$  e  $CC'$  sono adimensionali e hanno lunghezza proporzionale agli errori (relativi) che si commettono sulle lunghezze o sulle inclinazioni.

la probabilità di finire in ogni punto della curva deve essere la stessa che si ha di finire in  $b$ ,  $C$ ,  $b'$  o  $C'$ .<sup>19</sup> ([1])

Adrain voleva mostrare come tale curva dovesse restringere nella regione compresa tra i due quadrati  $EFGH$  e  $bCb'C'$ . Preso un punto  $m$  interno al segmento  $Bb$  ed eretta la perpendicolare  $Pf$ , siccome la probabilità di avere un errore  $Bm$  in lunghezza è inferiore a quella di avere un errore 0 in lunghezza, dal momento che, anche per Adrain, ad errori maggiori fanno riscontro minori probabilità, mentre la probabilità dell'errore sull'orientazione in  $P$  non differisce da quella in  $C$ , il punto  $P$  si trova su una curva cui compete una probabilità di errore totale *minore* di quella relativa alla curva passante per  $b$ ,  $C$ ,  $b'$  e  $C'$ . Lo stesso discorso vale per i punti di  $GF$  diversi da  $b$  perché, coinvolgono un errore sugli angoli che è maggiore di quello in  $b$ , mentre l'errore sulla lunghezza del segmento è la stessa. Similmente, sui punti del segmento  $mn$ , dove  $n$  è l'intersezione di  $mP$  con  $Cb$ , l'errore in lunghezza ha la stessa probabilità di verificarsi che in  $m$ . Pur essendo  $Bb = Bm + mn$ , è più probabile commettere un certo errore totale *ripartito* sulle due misure, angolari e di lunghezza, che non tutto concentrato sulla sola lunghezza, come accade in  $b$ . Ripetendo lo stesso ragionamento per tutti i punti interni al quadrato  $bCb'C'$ , Adrain concluse che questi punti appartenevano a curve di probabilità con valori *maggiori* di quella passante per i vertici  $b$ ,  $C$ ,  $b'$  o  $C'$  che dunque rimane nella regione  $bGC$ , deve essere continua e deve essere formata da quattro porzioni simmetriche rispetto agli assi coordinati:

Infine, la curva deve essere la più semplice possibile tra quelle che soddisfano tutte le condizioni precedenti e dunque deve essere la circonferenza di un cerchio avente centro in  $B$ .<sup>20</sup> ([1])

Si tratta di una deduzione gratuita, motivata solo da comodità di calcolo. Infatti, se  $X(a, x)$  ed  $Y(b, y)$  esprimono le (densità di) probabilità che gli errori  $x$  ed  $y$  si manifestino sulle grandezze  $a$  e  $b$  e se ne assume l'indipendenza, la probabilità che questi errori si presentino simultaneamente è  $X(a, x)Y(b, y)$ . Adrain massimizza il logaritmo  $\ln X(a, x) + \ln Y(b, y)$  di questa probabilità, sulla curva di equazione  $x^2 + y^2 = \text{costante}$ . Se si parametrizza questo vincolo come  $(x(t), y(t))$ , in modo che sia

$$x\dot{x} + y\dot{y} = 0 \quad \dot{y} = -\frac{x}{y}\dot{x}, \quad (8.6)$$

poiché

$$\frac{d}{dt}(\ln X(a, x) + \ln Y(b, y)) = \dot{x} \frac{\partial \ln X(a, x)}{\partial x} + \dot{y} \frac{\partial \ln Y(b, y)}{\partial y},$$

la condizione di stazionarietà diventa, in virtù di (8.6),

$$\dot{x} \left( \frac{\partial \ln X(a, x)}{\partial x} - \frac{x}{y} \frac{\partial \ln Y(b, y)}{\partial y} \right) = 0 \quad (8.7)$$

<sup>19</sup>the probability of terminating on any point of the curve may be the same as the probability of terminating on any one of the four points  $b$ ,  $C$ ,  $b'$ ,  $C'$ .

<sup>20</sup>Lastly, the curve must be the simplest possible having all the preceding conditions, and must consequently be the circumference of a circle having its centre in  $B$ .

da cui si ottiene, se  $\dot{x} \neq 0$ ,

$$\frac{1}{x} \frac{\partial \ln X(a, x)}{\partial x} = \frac{1}{y} \frac{\partial \ln Y(b, y)}{\partial y} = -n \quad (8.8)$$

dove  $n$  è una costante positiva e la distribuzione esponenziale viene ottenuta come prima. Integrando, la probabilità di commettere errori  $x$  ed  $y$  sulle variabili  $a$  e  $b$  è proporzionale a

$$e^{-\frac{n}{2}(x^2+y^2)}.$$

Quando gli errori equiprobabili lungo i due assi non hanno ugual valore ma sono tali che  $\frac{Bb}{bC} = \frac{1}{p}$ , Adrain prese come curva di egual probabilità un'ellisse di centro  $B$  e la curva degli errori è proporzionale a

$$e^{-\frac{n}{2}\left(x^2 + \frac{y^2}{p}\right)}.$$

La probabilità di osservare errori  $x$  ed  $y$  è massima quando l'esponente è minimo e, in questo modo, Adrain diede una giustificazione probabilistica del metodo dei minimo quadrati, completando il suo lavoro con alcune applicazioni, tra le quali figurano le stesse utilizzate da Legendre, anche se non è chiaro quando egli venne a conoscenza del contenuto dei *Nouvelles méthodes*.

Questa dimostrazione della legge degli errori ha aspetti in comune con quella che sarà fornita nel 1850 dall'astronomo britannico John Herschel (1792-1871) e criticata da Robert Leslie Ellis (1817-1859), personaggio poliedrico che, oltre ad essere stato un matematico, fu curatore delle opere di Francis Bacon. Herschel, convinto che vi fosse una sola legge degli errori  $f$ , la determinava ricorrendo ad un esempio specifico; un sasso lasciato cadere con l'intenzione di colpire un bersaglio: la distanza  $r$  dal punto di caduta dal centro del bersaglio rappresenta l'errore della misura. Introdotte coordinate cartesiane  $x$  ed  $y$  con centro nel punto da colpire, Herschel suppose che la probabilità di errore fosse funzione solo di  $r$ ,  $f(r^2) = f(x^2 + y^2)$ , visto che gli errori in direzioni distinte hanno la stessa probabilità. Per determinare la forma di  $f$ , Herschel scomponeva la posizione del punto di impatto lungo gli assi coordinati ed osservava che il verificarsi di un errore  $r$

è dunque un evento composto di cui queste [le proiezioni sugli assi] sono i costituenti semplici e quindi la sua probabilità sarà il prodotto delle loro probabilità, prese separatamente.<sup>21</sup> ([9], p. 325)

Si tratta della stessa ipotesi di Adrain che condusse Herschel all'equazione funzionale per  $f$

$$f(x^2)f(y^2) = f(x^2 + y^2). \quad (8.9)$$

L'ipotesi di indipendenza fu però criticata decisamente da Ellis:

---

<sup>21</sup>is therefore a compound event of which they are the simple constituents, therefore its probability will be the product of their separate probabilities.

a meno di non dimostrare che una deviazione  $y$  avvenga con la stessa frequenza relativa quando  $x$  ha un valore piuttosto che un altro, non siamo autorizzati a dire che la probabilità del verificarsi delle deviazioni  $x$  ed  $y$  sia il prodotto delle probabilità delle singole deviazioni.<sup>22</sup> ([9], p.326)

A dispetto delle critiche ricevute, non solo da Ellis, la deduzione della legge degli errori a partire dall'equazione funzionale (8.9) ebbe un certo successo: Maxwell se ne servì per la prima deduzione della legge di distribuzione delle velocità molecolari in un gas perfetto, salvo poi abbandonarla per privilegiare argomenti che mettevano in risalto il ruolo degli urti tra molecole.

## 8.5 Gauss e la legge degli errori

Il terzo fondatore del metodo dei minimi quadrati fu Carl Friedrich Gauss (1777-1855) che ne trattò per la prima volta in un testo a stampa nel 1809, a conclusione del volume *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, dedicato allo studio delle orbite dei pianeti [12]. A differenza di Legendre, Gauss legò subito il metodo alla probabilità. Egli considerò un numero  $\mu$  di funzioni  $V_1, V_2, \dots, V_\mu$  dipendenti da un numero  $\nu < \mu$  di incognite  $p, q, r, s, \dots$ , mentre  $M_1, \dots, M_\mu$  indicano i valori delle funzioni determinati a seguito di un'osservazione. Le ipotesi di Gauss sono le seguenti:

- non vi è motivo di supporre a priori un'osservazione eseguita meglio di un'altra ed occorre ritenere che errori che abbiano lo stesso valore siano equiprobabili;
- la probabilità che un errore  $\Delta$  abbia valori compresi nell'intervallo  $[\Delta, \Delta + d\Delta]$  è

$$\varphi(\Delta)d\Delta$$

dove  $\varphi$  è una funzione *pari*, con un massimo in  $\Delta = 0$

- questo massimo è *unico* e la funzione  $\varphi(\Delta)$  si deve annullare quando  $\Delta$  eccede una soglia critica;
- invece di lavorare con una funzione su un dominio limitato, si suppone  $\varphi(\Delta)$  definita su tutto l'asse reale con la condizione che tenda asintoticamente all'asse delle ascisse quando  $|\Delta| \gg 1$ .

Come corollari, Gauss evidenzia che la probabilità di trovare un errore nell'intervallo finito  $[D, D']$  è

$$\int_D^{D'} \varphi(\Delta)d\Delta$$

<sup>22</sup>unless it can be shown that a deviation  $y$  occurs with the same comparative frequency when  $x$  has one value as when it has another, we are not entitled to say that the probability of the concurrence of two deviations  $x$  and  $y$  is the product of the probabilities of each.

e che

$$\int_{-\infty}^{\infty} \varphi(\Delta) d\Delta = 1.$$

Con queste osservazioni, Gauss può considerare la funzione

$$\Omega := P(M_1 - V_1, \dots, M_\mu - V_\mu | p, q, r, \dots) = \varphi(M_1 - V_1) \varphi(M_2 - V_2) \cdots \varphi(M_\mu - V_\mu) \quad (8.10)$$

• il principio della media aritmetica:

Come assioma si è soliti assumere che, se una certa quantità viene determinata tramite più osservazioni immediate, eseguite nelle stesse circostanze e con uguale attenzione, la media aritmetica di tutti i valori osservati mostrerà il valore più probabile, se non con rigore assoluto, almeno con con approssimazione tale da potervi aderire nel modo più sicuro possibile.

Gauss considera allora il caso particolare in cui  $V_1 = V_2 = \cdots = V_\mu = p$  ed assume la media aritmetica dei valori misurati  $M_1, M_2, \dots, M_\mu$  come valore per  $p$ :  $p = \frac{1}{\mu}(M_1 + M_2 + \cdots + M_\mu)$ . La condizione di massima probabilità diventa allora:

$$\varphi'(M_1 - p) + \varphi'(M_2 - p) + \cdots + \varphi'(M_\mu - p) = 0.$$

Poiché questa condizione deve valere per tutti i sistemi di valori  $M_1, M_2, \dots, M_\mu$ , Gauss considerò il caso particolare in cui

$$M_2 = M_3 = \cdots = M_\mu = M_1 - \mu N$$

cosicché

$$M_1 - p = (\mu - 1)N \quad \text{e quindi } M_i - p = -N, \text{ se } i \neq 1.$$

La condizione di stazionarietà si riduce a

$$\varphi'[(\mu - 1)N] = (1 - \mu)\varphi'(-N) = (\mu - 1)\varphi'(N)$$

da cui Gauss concluse che

$$\frac{\varphi'(\Delta)}{\Delta} = \frac{1}{\Delta} \frac{d \ln \varphi(\Delta)}{d\Delta} = k$$

dove  $k$  è una costante. Integrando, si ottiene

$$\ln \varphi(\Delta) = \frac{k}{2} \Delta^2 + C,$$

dove  $C$  è una costante o, in modo equivalente,

$$\varphi(\Delta) = \chi e^{\frac{k}{2} \Delta^2}, \quad \chi = e^C.$$

Poiché le ipotesi sulla natura della funzione  $\varphi$  impongono  $k < 0$ , Gauss pose  $\frac{k}{2} = -h^2$  e, grazie all'integrale di Laplace

$$\int_{-\infty}^{\infty} e^{-h^2 \Delta^2} d\Delta = \frac{\sqrt{\pi}}{h}, \quad (8.11)$$

Gauss poté finalmente scrivere “la” legge di distribuzione degli errori come

$$\varphi(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}. \quad (8.12)$$

Gauss chiarì il significato del parametro  $h$

come misura della precisione delle osservazioni.<sup>23</sup> ([12], p. 213)

L’argomento adoperato si può riassumere in questi termini: se si considera, oltre alla curva (8.12), un’altra curva di parametro  $h_1$

$$\varphi_1(x) = \frac{h_1}{\sqrt{\pi}} e^{-h_1^2 x^2}$$

ci si può attendere che, nel primo caso, l’errore sarà compreso nell’intervallo  $(-\delta, \delta)$  con probabilità

$$\int_{-\delta}^{\delta} \varphi(x) dx$$

mentre nel secondo caso la probabilità di un errore compreso nell’intervallo  $(-\delta_1, \delta_1)$  sarà data da

$$\int_{-\delta_1}^{\delta_1} \varphi_1(x) dx :$$

imponendo l’uguaglianza di queste probabilità, sarà più precisa la distribuzione cui compete una maggior compattazione degli errori attorno al valore  $x = 0$ . Operando le trasformazioni  $hx = t$  nel primo integrale e  $h_1x = t$  nel secondo, si perviene agli integrali

$$\int_{-h\delta}^{h\delta} \Phi(t) dt \quad \text{e} \quad \int_{-h_1\delta_1}^{h_1\delta_1} \Phi(t) dt, \quad \text{con} \quad \Phi(t) = \frac{1}{\sqrt{\pi}} e^{-t^2}$$

che sono uguali solo se

$$h\delta = h_1\delta_1$$

per cui a valori maggiori di  $h$  corrisponde una maggior precisione delle misurazioni. Riscrivendo

$$\Omega(v_1, \dots, v_\mu) = h^\mu \pi^{-\frac{\mu}{2}} e^{-(v_1^2 + v_2^2 + \dots + v_\mu^2)},$$

in cui entrano gli scarti  $v_1, v_2, \dots, v_\mu$ , Gauss osservò che il massimo di  $\Omega$  si realizza quando l’esponente è minimo, che è la stessa condizione imposta nel metodo dei minimi quadrati. Gauss riteneva di poter concludere dicendo che il principio della media aritmetica equivaleva a postulare la validità del metodo dei minimi quadrati:

<sup>23</sup>tamquam mensura praecisionis observationum.

E dunque il sistema più probabile dei valori delle incognite  $p, q, r, s$  ecc. sarà quello in cui i quadrati delle differenze tra i valori osservati e quelli calcolati delle funzioni  $V_1, V_2, V_3$  ecc. daranno la somma minima, se pure si debba presumere lo stesso grado di precisione in tutte le osservazioni.

Questo principio, che offre un uso molto frequente nelle applicazioni della matematica alla filosofia naturale, deve valere come assioma alla stessa stregua con cui si adotta la media aritmetica tra più valori osservati come valore più probabile.<sup>24</sup> ([12], p. 213)

L'ipotesi di equivalenza delle osservazioni venne superata introducendo dei pesi opportuni. Gauss considerò anche il metodo dei minimi quadrati indipendentemente dal calcolo delle probabilità in una sezione (§186) nella quale citò Legendre, senza mancare però di fare osservare come egli si fosse già servito del metodo a partire dal 1795, un decennio prima della pubblicazione di Legendre. Questi, a sua volta, reagì in maniera piuttosto piccata senza però che Gauss sentisse il dovere di intervenire nuovamente sulla questione. La deduzione della legge degli errori di Gauss dipende dunque in modo sostanziale dal principio della media aritmetica, sulla cui generalità però vi possono essere dei dubbi: abbiamo infatti visto come Laplace, già nel 1774, avesse formulato un criterio alternativo che portava alla mediana di una serie di osservazioni come “valor medio” più affidabile. Di fatto, Gauss selezionò la classe di funzioni di distribuzione  $\varphi(\Delta)$  da prendere in considerazione, semplificando l'equazione funzionale da risolvere. Quanto a questa equazione, abbiamo visto che egli dedusse da un'equazione del tipo

$$f(nx) = nf(x) \quad (8.13)$$

valida per ogni valore di  $x$  e per  $n$  intero, la soluzione  $f(x) = kx$ . Ora, l'equazione (8.13) si deduce dall'equazione funzionale di Cauchy

$$f(x+y) = f(x) + f(y)$$

e si può dimostrare che la soluzione generale è  $f(x) = kx$  non è del tutto immediata: in Appendice a questo capitolo è proposta la dimostrazione fornita da Louis-Augustin Cauchy (1789-1857) nel *Cours d'Analyse* pubblicato nel 1821. Possiamo sintetizzare l'aspetto critico della derivazione ancora con le parole di Robert Ellis ([8], p.207):

In realtà, siamo del tutto certi che, per diverse classi di osservazioni, la legge di probabilità debba cambiare e non abbiamo un'evidenza diretta che in una classe essa coincida con la forma assegnatale. Pertanto, una

<sup>24</sup>Systema itaque maxime probabile valorum incognitarum  $p, q, r, s$  etc. erit, in quo quadrata differentiarum inter functionum  $V_1, V_2, V_3$  etc. valores observatos et computatos summam minimam efficiunt, siquidem in omnibus observationibus idem praecisionis gradus praesumendus est.

Hocce principium, quod in omnibus applicationibus mathesis ad philosophiam naturalem usum frequentissimum offert, ubique axiomatis loco eodem iure valere debet, quo medium arithmeticum inter plures valores observatos eiusdem quantitatis tamquam valor maxime probabilis adoptatur.

delle due: o la regola della media aritmetica si appoggia ad un pregiudizio del tutto illusorio o, se ha un fondamento valido, il ragionamento (...) deve essere sbagliato (...). Ora, (...) è chiaro che non c'è alcun motivo per supporre che, dal momento che la media aritmetica fornisce il risultato vero quando il numero delle osservazioni cresce *sine limite*, esso debba fornire il risultato più probabile, quando il numero di osservazioni è finito.<sup>25</sup>

L'ultima frase si riferisce alla presenza della distribuzione Gaussiana nel teorema centrale del limite, cui dedicheremo la prossima sezione.

## 8.6 Il teorema centrale del limite

Nel 1810 Laplace pubblicò una memoria [18], dal titolo: *Mémoire sur les approximations des formules qui sont fonctions de très grand nombres et sur leur application aux probabilités*, di grande importanza perché contiene la prima dimostrazione del teorema che György Pólya [26] chiamerà *teorema centrale del limite* (*Zentral Grenzwertsatz*) per il ruolo fondamentale che gioca in tutta la teoria della probabilità. La memoria di Laplace non si occupava del metodo dei minimi quadrati ma, letta la *Theoria motus* di Gauss, in cui la distribuzione normale degli errori giocava un ruolo fondamentale, Laplace vi aggiunse una nota [19] con l'intento di fondare su basi più solide il metodo dei minimi quadrati. Procedendo con ordine, la memoria del 1810 fu presentata da Laplace come un notevole progresso analitico perché egli era riuscito a superare difficoltà tecniche che si erano invece rivelate insormontabili in precedenza. Il problema da cui prese le mosse Laplace era ancora di meccanica celeste:

Si supponga che tutte le inclinazioni rispetto all'eclittica siano ugualmente possibili, da zero all'angolo retto; si richiede la probabilità che l'inclinazione media di  $n$  orbite sia compresa tra limiti assegnati.<sup>26</sup> ([18], p. 305)

Laplace voleva rendersi conto se un eventuale addensamento delle inclinazioni delle orbite delle comete note o dei pianeti del sistema solare attorno ad un valor medio fosse meramente casuale o fosse da attribuire ad una causa specifica. Tradotto in altri termini, Laplace considerò  $n$  variabili aleatorie distribuite uniformemente e voleva trovare la distribuzione della somma. Non ci dilunghiamo su questa parte della memoria di Laplace perché le stesse tecniche verranno riutilizzate nella sezione VI, dedicata specificamente alla teoria degli errori dove,

<sup>25</sup>In fact, we are perfectly sure that in different classes of observations the law of probability of error must vary, and we have no direct proof that in any class it coincides with the form assigned to it. Therefore one of two things must be true, either the rule of the arithmetical mean rests on a mere illusory prejudice, or, if it has a valid foundation, the reasoning (...) must be incorrect (...). Now (...) it is clear that there is no reason for supposing that because the arithmetical mean would give the true result if the number of observations were increased sine limite, it must give the most probable result, the number of observations being finite.

<sup>26</sup>On suppose toutes les inclinaisons à l'écliptique également possibles depuis zéro jusqu'à l'angle droit, et l'on demande la probabilité que l'inclinaison moyenne de  $n$  orbites sera compris dans des limites données.

rispetto al problema dell'inclinazione delle orbite, Laplace non ipotizzò a priori una distribuzione uniforme per gli errori da cui sono affette le singole misure:

Il problema risolto circa le inclinazioni è lo stesso di quello in cui si propone di determinare che l'errore medio di un numero  $n$  di osservazioni sarà compreso tra limiti assegnati, supponendo che gli errori delle singole osservazioni si possano distribuire in modo uniforme nell'intervallo [di ampiezza]  $h$ . Considereremo ora il caso generale in cui gli errori si distribuiscono secondo una legge qualsiasi.<sup>27</sup> ([18], p. 322)

Laplace suppose che gli errori delle singole misure seguissero una distribuzione  $\varphi(x)$  su cui egli non pose alcuna ipotesi di simmetria ma solo le condizioni affinché rappresentasse una densità di probabilità adatta al problema in esame. Gli errori possono essere positivi o negativi ed estendersi lungo un intervallo di lunghezza  $h$  che viene diviso in un numero molto grande (*un nombre infini*)  $k + k'$  di parti in modo che  $k$  parti stiano a sinistra dell'errore nullo e  $k'$  parti a destra. Il valore di  $\varphi\left(\frac{s}{k+k'}\right)$  è (proporzionale al)la probabilità di avere, in una qualsivoglia osservazione, un errore posto nel punto  $s$  della suddivisione. A questo punto egli si servì di una variante del metodo delle funzioni generatrici, in cui figurano funzioni esponenziali. Precisamente, egli considerò la funzione

$$\left[ \sum_{\ell=-k}^{k'} \varphi\left(\frac{\ell}{k+k'}\right) e^{\ell i\omega} \right]^n$$

ed osservò che il coefficiente di  $e^{s i\omega}$  rappresenta la probabilità che la *somma* degli errori in  $n$  osservazioni sia  $s$ . D'altra parte se si moltiplica

$$\sum_{\ell=-k}^{k'} \varphi\left(\frac{\ell}{k+k'}\right) e^{\ell i\omega}$$

prima per un certo esponenziale  $e^{-qi\omega}$  e poi si eleva il prodotto alla potenza  $n$ , il coefficiente di  $e^{s i\omega}$  in questa potenza è la probabilità che la somma degli errori sia  $s + nq$ . Laplace prende per  $q$  il numero delle suddivisioni comprese tra l'ascissa  $x = 0$ , dove l'errore è nullo, e l'ascissa del centro di massa della porzione di piano  $\mathcal{S}$  Compresa tra il grafico di  $\varphi(x)$  e l'asse delle ascisse. La somma da calcolare è

$$S(\omega) := \left[ \sum_{\ell=-k}^{k'} \varphi\left(\frac{\ell}{k+k'}\right) e^{(\ell-q)i\omega} \right]^n$$

<sup>27</sup>Le problème que nous avons résolu (...) relativement aux inclinaisons, est le même que celui dans lequel on se propose de déterminer la probabilité que l'erreur moyenne d'un nombre  $n$  d'observations sera compris dans des limites données, en supposant que les erreurs de chaque observation puissent également s'étendre dans l'intervalle  $h$ . Nous allons maintenant considérer le cas général dans lequel les facilités des erreurs suivent une loi quelconque.

che Laplace trattò passando al continuo grazie alle posizioni

$$\frac{\ell}{k+k'} = \frac{x}{h} \quad \frac{q}{k+k'} = \frac{q'}{h} \quad \frac{1}{k+k'} = \frac{dx}{h} \quad (8.14)$$

dove ora  $q'$  è l'ascissa del centro di massa della regione  $\mathcal{S}$ , normalizzata alla misura  $h$  del supporto della funzione  $\varphi$ . Poiché  $\frac{k+k'}{h}dx = 1$ , si può scrivere

$$S(\omega) = \left[ \sum_{\ell=-k}^{k'} \varphi \left( \frac{\ell}{k+k'} \right) e^{(\ell-q)i\omega} \right]^n = \left[ \sum_{\ell=-k}^{k'} \varphi \left( \frac{\ell}{k+k'} \right) e^{(\ell-q)i\omega} \frac{k+k'}{h} dx \right]^n$$

per cui, passando dalle somme agli integrali e sviluppando l'esponenziale  $e^{(\ell-q)i\omega}$ , si ottiene

$$S(\omega) = \frac{(k+k')^n}{h^n} \left[ \int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \varphi \left( \frac{x}{h} \right) dx + (k+k')i\omega \int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \frac{x-q'}{h} \varphi \left( \frac{x}{h} \right) dx - \frac{(k+k')^2}{2} \omega^2 \int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \frac{(x-q')^2}{h^2} \varphi \left( \frac{x}{h} \right) dx + \dots \right]^n.$$

La scelta del valore di  $q$  e quindi di  $q'$  si rivela particolarmente felice perché, grazie alle proprietà del centro di massa, si può concludere che

$$\int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \frac{x-q'}{h} \varphi \left( \frac{x}{h} \right) dx = 0$$

riducendo così  $S$  a

$$S(\omega) = \frac{(k+k')^n}{h^n} \mu^n \left[ 1 - \frac{\mu'}{2\mu} (k+k')^2 \omega^2 + \dots \right]^n$$

dove

$$\mu := \int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \varphi \left( \frac{x}{h} \right) dx \quad \mu' := \int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \frac{(x-q')^2}{h^2} \varphi \left( \frac{x}{h} \right) dx.$$

Ora, se la funzione  $S(\omega)$  viene moltiplicata per  $2 \cos r\omega = (e^{ir\omega} + e^{-ir\omega})$ , il termine indipendente da  $\omega$  nel nuovo prodotto rappresenta la probabilità  $P$  che la somma degli errori sia  $nq - r$  oppure  $nq + r$ . Per trovare questo termine, grazie all'ortogonalità delle funzioni  $\{e^{ik\omega}\}$  su  $[0, 2\pi]$ , occorre calcolare<sup>28</sup>

$$\frac{1}{2\pi} \int_0^{2\pi} 2S(\omega) \cos r\omega d\omega,$$

ottenendo così

$$P = \frac{(k+k')^n}{h^n} \frac{1}{\pi} \mu^n \int_0^{2\pi} \cos r\omega \left[ 1 - \frac{\mu'}{2\mu} (k+k')^2 \omega^2 + \dots \right]^n d\omega. \quad (8.15)$$

<sup>28</sup>Il  $\pi$  usato nella memoria di Laplace è uguale al nostro  $2\pi$ .

Posto

$$t := (k + k')\omega \quad (8.16)$$

Laplace ricavò

$$\begin{aligned} \ln \left[ 1 - \frac{\mu'}{2\mu}(k + k')^2\omega^2 + \dots \right]^n &= \\ &= n \ln \left( 1 - \frac{\mu'}{2\mu}(k + k')^2\omega^2 + \dots \right) = n \left( -\frac{\mu'}{2\mu}t^2 + \dots \right), \end{aligned}$$

senza alcuna discussione sulla liceità del passaggio che coinvolge uno sviluppo in serie di  $\ln(1 - x)$ , il cui raggio di convergenza è finito. L'espressione di  $P$  diventa allora

$$P := \left( \frac{(k + k')\mu}{h} \right)^n \frac{1}{\pi} \frac{1}{k + k'} \int_0^\infty \cos \left( \frac{rt}{k + k'} \right) e^{-\frac{\mu'}{2\mu}nt^2} [1 + Ant^2 + \dots]^n dt,$$

dove  $A$  è un coefficiente opportuno. Ora, dal momento che tutti gli errori delle singole osservazioni cadono nell'intervallo, di ampiezza  $h$ ,  $\left[ -\frac{kh}{k+k'}, \frac{k'h}{k+k'} \right]$ , deve aversi

$$\sum_{\ell=-k}^{k'} \varphi \left( \frac{\ell}{k + k'} \right) = \sum_{\ell=-k}^{k'} \varphi \left( \frac{\ell}{k + k'} \right) \frac{k + k'}{h} dx = 1$$

e, sostituendo alla somma un integrale, otteniamo

$$\int_{-\frac{kh}{k+k'}}^{\frac{k'h}{k+k'}} \varphi \left( \frac{x}{h} \right) dx = \frac{h}{k + k'}, \quad (8.17)$$

cioè  $\frac{(k+k')\mu}{h} = 1$ . Mantenendo i termini dominanti nello sviluppo precedente e ponendo  $\frac{r}{k+k'} = \ell\sqrt{n}$ , si ricava

$$P = \frac{1}{\pi(k + k')} \int_0^\infty \cos(\ell t\sqrt{n}) e^{-\frac{\mu'}{2\mu}nt^2} dt = \frac{1}{\sqrt{\pi}(k + k')} \sqrt{\frac{\mu}{\mu'n}} e^{-\frac{\mu'}{2\mu}\ell^2}$$

dove, nell'ultimo passaggio, Laplace si è servito di risultati dimostrati nella Sezione III di [18]. Per avere la probabilità che l'errore cada nell'intervallo  $[nq - r, nq + r]$  occorre integrare  $P$  in  $dr$ . Dunque la distribuzione di probabilità normale<sup>29</sup> compare come distribuzione limite dell'errore da cui è affetta la somma di variabili aleatorie identicamente distribuite e quindi la loro media aritmetica. È superfluo notare che l'analisi di Laplace manca del rigore necessario per renderla accettabile non solo ai nostri occhi ma anche a quelli di diversi matematici del XIX secolo. Ciò però non diminuisce la sua importanza, come osserverà lo statistico francese Irenée-Jules Bienaymé (1796-1878) ([2], p. 37):

<sup>29</sup>Il termine sarà introdotto nella letteratura nell'ultimo quarto del XIX secolo da statistici come l'inglese Francis Galton (1822-1911), il tedesco Wilhelm Lexis (1837-1914) e dal matematico americano Charles Sanders Peirce (1839-1914). Sarà a partire da un lavoro [22] dello statistico Karl Pearson che il termine sarà adottato nella letteratura.

I principii della dimostrazione di Laplace sono al riparo da ogni obiezione; solo gli strumenti analitici ne possono sollevare. Non possiedono tutto il rigore auspicabile e forse nelle applicazioni sarebbe opportuno discutere i casi particolare che potrebbero presentarsi.<sup>30</sup>

La memoria del 1810 contiene un'appendice [19] in cui Laplace mostra di essere nel frattempo venuto a conoscenza della *Theoria motus* di Gauss e, comprendendo l'importanza, pose a fondamento della teoria degli errori proprio il teorema centrale del limite.

## 8.7 Il *Theorema insigne* di Gauss

Oltre al lavoro del 1809, Gauss si occupò altre volte di teoria degli errori, visti anche i suoi interessi in astronomia e geodesia [20]. Tra i suoi lavori, esaminiamo la *Theoria combinationis observationum erroribus minime obnoxiae*, pubblicata nel 1823 [13], a partire da un teorema, che lui indicò come *theorema insigne*, e che consiste in una disuguaglianza del tipo di Bienaymé-Čebišev, che cioè stima, per una densità di probabilità  $\varphi(x)$  che deve soddisfare condizioni non restrittive e che per Gauss rappresenta sempre una legge ammissibile degli errori di misurazione, integrali della forma

$$\int_{-\lambda m}^{\lambda m} \varphi(x) dx =: \mu \quad (8.18)$$

quando  $\lambda$  è un parametro reale e  $m$  rappresenta lo scarto quadratico medio della distribuzione, avendo Gauss scelto di porre

$$m^2 := \int_{-\infty}^{+\infty} x^2 \varphi(x) dx.$$

Gauss chiama *m error medium metuendum*, l'errore medio da temere, mentre chiama errore probabile il valore di  $\varrho := \lambda m$  tale che

$$\int_{-\varrho}^{\varrho} \varphi(x) dx = \frac{1}{2}.$$

Gauss si pose dapprima il problema di valutare integrali di tipo (8.18) per alcune distribuzioni particolari  $\varphi(x)$ , tra cui la distribuzione normale, per poi dimostrare il teorema cui abbiamo fatto riferimento nel titolo di questa sezione ([13], pp. 9-10)

Benché la relazione tra  $\lambda$  e  $\mu$  dipenda dalla natura della funzione<sup>31</sup>  $\varphi(x)$ , è tuttavia possibile stabilire dei risultati generali. Vale a dire, quale che sia

<sup>30</sup>Les principes de la démonstration de Laplace sont à l'abri de toute objection; les moyens analytiques peuvent seuls en soulever quelques-unes. Il n'ont point toute la rigueur désirable, et peut-être, dans les applications, conviendrait-il de bien discuter les cas particuliers qui pourraient se présenter.

<sup>31</sup>Notiamo che, conformemente all'uso del tempo, l'argomento di una funzione non era posto tra parentesi.

questa funzione, purché solo sia disposta in modo tale da decrescere sempre o almeno non crescere, quando  $x$  cresce in valore assoluto, certamente  $\lambda$  sarà minore o al più uguale a  $\mu\sqrt{3}$ , ogni volta in cui  $\mu$  è minore di  $\frac{2}{3}$ ;

$\lambda$  non è maggiore di  $\frac{2}{3\sqrt{1-\mu}}$ , ogni volta in cui  $\mu$  è maggiore di  $\frac{2}{3}$ . Nel caso in cui  $\mu = \frac{2}{3}$  entrambi i limiti coincidono, supposto che  $\lambda$  non possa superare  $\sqrt{\frac{4}{3}}$ .<sup>32</sup>

Gauss considerò la classe di distribuzioni  $\varphi(x)$  dotate di un solo massimo in  $x = 0$  (distribuzioni unimodali) e a media nulla:

$$\int_{-\infty}^{\infty} x\varphi(x)dx = 0.$$

Per la dimostrazione, egli invertì l'integrale

$$y(x) = \int_{-x}^x \varphi(t)dt$$

ponendo  $x = \psi(y)$ , cosicché fosse

$$y = \int_{-\psi(y)}^{\psi(y)} \varphi(t)dt.$$

Derivando ambo i membri rispetto ad  $y$  si ottiene

$$1 = \varphi(\psi(y))\psi'(y) + \varphi(-\psi(y))\psi'(y) = [\varphi(\psi(y)) + \varphi(-\psi(y))]\psi'(y)$$

e quindi, ripristinando il significato di  $\psi(y)$ ,

$$\psi'(y) = \frac{1}{\varphi(x) + \varphi(-x)}.$$

Poiché, per definizione,  $y \in [0, 1]$  e  $\varphi(x)$  è unimodale, con massimo assoluto in  $x = 0$ ,  $\psi'(y)$  è non solo positiva, ma anche monotona crescente o, almeno, non decrescente, cosicché  $\psi''(y) > 0$ . Inoltre, essendo  $\psi(0) = 0$ , integrando l'uguaglianza

$$d(t\psi'(t)) = \psi'(t)dt + t\psi''(t)dt$$

nell'intervallo  $[0, y]$ , avremo

$$y\psi'(y) = \psi(y) + \int_0^y t\psi''(t)dt$$

<sup>32</sup>Quanquam relatio inter  $\lambda$  et  $\mu$  ab indole functionis  $\varphi x$  pendet, tamen quaedam generalia stabilire licet. Scilicet qualiscunque sit haec functio, si modo ita est comparata, ut ipsius valor, crescente valore absoluto ipsius  $x$ , sempre decrescat, vel saltem non crescat, certo erit  $\lambda$  minor vel saltem non maior quam  $\mu\sqrt{3}$ , quoties  $\mu$  est minor quam  $\frac{2}{3}$ ;

$\lambda$  non maior quam  $\frac{2}{3\sqrt{1-\mu}}$ , quoties  $\mu$  est maior quam  $\frac{2}{3}$ . Pro  $\mu = \frac{2}{3}$  uterque limes coincidit, puta  $\lambda$  nequit esse maior quam  $\sqrt{\frac{4}{3}}$ .

e quindi

$$y\psi'(y) - \psi(y) = \int_0^y t\psi''(t)dt \geq 0$$

per cui la funzione

$$g(y) = 1 - \frac{\psi(y)}{y\psi'(y)}$$

varia nell'intervallo  $[0, 1]$ . Posto  $f = g(\mu)$  cosicché, per (8.18),  $\psi(\mu) = \lambda m$ , avremo

$$f = 1 - \frac{\lambda m}{\mu\psi'(\mu)},$$

da cui si ottiene

$$\psi'(\mu) = \frac{\lambda m}{(1-f)\mu} \quad (8.19)$$

che viene preso come coefficiente angolare della retta

$$F(y) = \psi'(\mu)(y - \mu f) \quad (8.20)$$

che, per costruzione, è tangente al grafico di  $\psi(y)$  nel punto  $y = \mu$ , essendo

$$F(\mu) = \psi(\mu) \quad \text{e} \quad F'(\mu) = \psi'(\mu).$$

Poiché  $\psi(y)$  è convessa, si conclude che deve essere

$$\psi(y) \geq F(y),$$

che è significativa nell'intervallo di valori  $[\mu f, 1]$  per la variabile  $y$ . Si ha allora

$$\int_{\mu f}^1 (F(y))^2 dy \leq \int_{\mu f}^1 (\psi(y))^2 dy \leq \int_0^1 (\psi(y))^2 dy : \quad (8.21)$$

poiché

$$y = \int_{-x}^x \varphi(t)dt =$$

si ha anche

$$\frac{dy}{dx} = \varphi(x) + \varphi(-x) \quad (8.22)$$

e quindi, essendo  $x = \psi(y)$ , si conclude che

$$\int_{\mu f}^1 (F(y))^2 dy \leq \int_0^{+\infty} x^2 \varphi(x) dx + \int_0^{+\infty} x^2 \varphi(-x) dx = \int_{-\infty}^{+\infty} x^2 \varphi(x) dx = m^2$$

o, calcolando esplicitamente il primo integrale,

$$m^2 \geq \frac{\lambda^2 m^2 (1 - \mu f)^3}{3(1-f)^2 \mu^2}$$

da cui si ottiene

$$\lambda^2 \leq \frac{3\mu^2(1-f)^2}{(1-\mu f)^3}.$$

Gauss studiò allora il comportamento del membro di destra, che indicheremo con  $\alpha(f)$ , al variare di  $f \in [0, 1]$ . Con un calcolo diretto, si ricava

$$\alpha' = -\frac{3\mu^2(1-f)}{(1-\mu f)^4} [2 - 3\mu + \mu f].$$

Ora, se  $\mu \leq \frac{2}{3}$ ,  $\alpha'(f) < 0$  su tutto il segmento  $[0, 1]$  per cui il massimo è raggiunto in  $\alpha(0)$  e vale  $3\mu^2$  per cui in questo caso

$$\lambda \leq \mu\sqrt{3}.$$

Quando invece  $\mu > \frac{2}{3}$ , il massimo di  $\alpha$  è raggiunto quando  $2 - 3\mu + \mu f = 0$ , cioè quando  $f = 3 - \frac{2}{\mu}$ . Inserendo questo valore in  $\alpha$  si ottiene

$$\lambda \leq \frac{2}{3} \frac{1}{\sqrt{1-\mu}},$$

che completa la dimostrazione del teorema. I risultati possono essere letti come limitazioni su  $\mu$ , in funzione di  $\lambda$ . In particolare poi, quando  $\mu = \frac{1}{2}$  e quindi  $\lambda m$  costituisce l'errore probabile, si ha che  $\lambda \leq \sqrt{\frac{3}{4}}$  ed il massimo è raggiunto sulla densità

$$\varphi(x) = \begin{cases} \frac{1}{2a} & x \in [-a, a] \\ 0 & \text{altrove} \end{cases}$$

dove  $a$  è legato allo scarto quadratico medio  $m$  dalla relazione  $m = \frac{a}{\sqrt{3}}$ , densità che Gauss aveva studiato prima di dimostrare questo elegante risultato.

## 8.8 Il modello lineare di Gauss

Nella *Theoria combinationis* Gauss propose un'altra via per arrivare al metodo dei minimi quadrati, abbandonando il ricorso al principio della media aritmetica. È stato ipotizzato che questo atteggiamento, come l'introduzione di risultati validi per ogni distribuzione  $\varphi(x)$ , indichi un allontanamento da parte di Gauss dall'idea che sia possibile trovare la legge degli errori [28], data dalla distribuzione normale. Infatti, introducendo la funzione  $\varphi(x)$  Gauss aveva osservato ([13], p. 4)

A stento, o addirittura mai, sarà possibile nella pratica assegnare a priori questa funzione.<sup>33</sup>

Come anticipato all'astronomo Heirich Olbers (1758-1840) in una lettera del 22 Febbraio 1819 [24], Gauss propose un ulteriore fondamento del metodo dei

<sup>33</sup>Vix, ac ne vix quidem, umquam in praxi possibile erit, hanc functionem a priori assignare.



Si trattava dunque di valutare l'errore medio di  $U$  a partire dagli errori medi commessi sulle variabili da cui  $U$  dipende, cioè un problema di propagazione degli errori. Ricordiamo che l'errore medio di una distribuzione è per Gauss ciò che oggi è noto come scarto quadratico medio.

L'idea della dimostrazione è la seguente. Siano  $e, e', e''$  gli errori commessi nelle misure delle incognite da cui dipende  $U$ . L'errore commesso su  $U$  si ottiene calcolando

$$E = U(V + e, V' + e', V'' + e'', \dots) - U(V, V', V'', \dots) = \lambda e + \lambda' e' + \lambda'' e'' + \dots,$$

dove si trascurano termini contenenti potenze degli errori  $e$  superiori alla prima e le quantità  $\lambda = \frac{\partial U}{\partial V} |_{(V, V', V'', \dots)}$  sono *costanti* rispetto agli errori  $e$ . Dall'ipotesi di indipendenza delle osservazioni Gauss ricava che, se gli errori  $e$  sono a media nulla, altrettanto deve essere per  $E$  e che l'errore medio  $M$  da cui è affetto  $E$  è

$$M = \sqrt{\lambda^2 m^2 + \lambda'^2 m'^2 + \lambda''^2 m''^2 + \dots}. \tag{8.27}$$

Se risolviamo la (8.24) rispetto ad  $x_1$  ed osserviamo che  $k$  non contiene errori, ma solo costanti del modello e valori misurati, possiamo concludere che, quando gli errori  $v_1, v_2, \dots, v_n$  hanno tutti uno stesso errore medio  $\mu$ , in virtù della (8.27) l'errore medio  $\mu_x$  da cui è affetto  $x_1$  sarà tale che

$$\mu_x^2 = \mu^2 (\kappa_1^2 + \kappa_2^2 + \dots + \kappa_n^2) :$$

la condizione (8.24) è interpretabile come una condizione di minima varianza per la distribuzione degli errori da cui è affetta  $x_1$ .

Veniamo alla realizzazione tecnica di queste idee. Come primo passo, Gauss formò le  $p$  quantità

$$\begin{aligned} \xi &= a_1 v_1 + a_2 v_2 + \dots + a_n v_n \\ \eta &= b_1 v_1 + b_2 v_2 + \dots + b_n v_n \\ \zeta &= c_1 v_1 + c_2 v_2 + \dots + c_n v_n \\ &\dots\dots\dots \\ \omega &= p_1 v_1 + p_2 v_2 + \dots + p_n v_n \end{aligned} \tag{8.28}$$

che si possono esprimere in termini delle variabili  $x_1, \dots, x_p$  come

$$\begin{aligned} \xi &= \left(\sum a_i^2\right) x_1 + \left(\sum a_i b_i\right) x_2 + \left(\sum a_i c_i\right) x_3 + \dots + \left(\sum a_i \ell_i\right) \\ \eta &= \left(\sum a_i b_i\right) x_1 + \left(\sum b_i^2\right) x_2 + \left(\sum b_i c_i\right) x_3 + \dots + \left(\sum b_i \ell_i\right) \\ \zeta &= \left(\sum a_i c_i\right) x_1 + \left(\sum b_i c_i\right) x_2 + \left(\sum c_i^2\right) x_3 + \dots + \left(\sum c_i \ell_i\right) \\ &\dots\dots\dots \end{aligned} \tag{8.29}$$

in cui le somme coprono tutti i valori di  $i$  da 1 an  $n$ . Il nuovo sistema contiene  $p$  equazioni e  $p$  incognite e Gauss scrisse la soluzione nella forma

$$x_1 = A + [\alpha^2]\xi + [\alpha\beta]\eta + [\alpha\gamma]\zeta + \dots. \tag{8.30}$$

dove la notazione dei coefficienti delle  $\xi, \eta, \zeta$ , sarà chiarita in seguito, ed espressioni analoghe valgono per le altre incognite  $x_2, \dots, x_p$ . Se ora si sostituiscono al

posto delle  $\xi$ ,  $\eta$ ,  $\zeta$  le loro espressioni (8.28) e si pone

$$\begin{aligned}\alpha_1 &:= a_1[\alpha^2] + b_1[\alpha\beta] + c_1[\alpha\gamma] + \dots \\ \alpha_2 &:= a_2[\alpha^2] + b_2[\alpha\beta] + c_2[\alpha\gamma] + \dots \\ \alpha_3 &:= a_3[\alpha^2] + b_3[\alpha\beta] + c_3[\alpha\gamma] + \dots \\ &\dots\dots\dots\end{aligned}\tag{8.31}$$

l'equazione (8.30) si riscrive come

$$x_1 - A = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \dots + \alpha_n v_n\tag{8.32}$$

per cui, essendo  $A$  indipendente dalle variabili  $x_i$ , i coefficienti  $\alpha_i$  soddisfano la condizione (8.24). Sottraendo (8.24) da (8.32) si ottiene

$$(\kappa_1 - \alpha_1)v_1 + (\kappa_2 - \alpha_2)v_2 + (\kappa_3 - \alpha_3)v_3 + \dots = A - k\tag{8.33}$$

che, essendo  $A - k$  indipendente dalle  $x_i$ , equivale a richiedere la validità delle equazioni

$$\begin{cases} (\kappa_1 - \alpha_1)a_1 + (\kappa_2 - \alpha_2)a_2 + (\kappa_3 - \alpha_3)a_3 + \dots = 0 \\ (\kappa_1 - \alpha_1)b_1 + (\kappa_2 - \alpha_2)b_2 + (\kappa_3 - \alpha_3)b_3 + \dots = 0 \\ (\kappa_1 - \alpha_1)c_1 + (\kappa_2 - \alpha_2)c_2 + (\kappa_3 - \alpha_3)c_3 + \dots = 0 \\ \dots\dots\dots : \end{cases}\tag{8.34}$$

moltiplicando le equazioni ordinatamente per  $[\alpha^2]$ ,  $[\alpha\beta]$ ,  $[\alpha\gamma]$  e tenendo conto delle equazioni (8.31), si giunge a

$$(\kappa_1 - \alpha_1)\alpha_1 + (\kappa_2 - \alpha_2)\alpha_2 + (\kappa_3 - \alpha_3)\alpha_3 + \dots = 0.\tag{8.35}$$

Siccome

$$\begin{aligned}\sum_{i=1}^n (\kappa_i - \alpha_i)^2 &= \sum_{i=1}^n \kappa_i^2 + \sum_{i=1}^n \alpha_i^2 - 2 \sum_{i=1}^n \alpha_i \kappa_i = \\ &= \sum_{i=1}^n \kappa_i^2 - \sum_{i=1}^n \alpha_i^2 + 2 \sum_{i=1}^n \alpha_i (\alpha_i - \kappa_i),\end{aligned}$$

la condizione (8.35) impone

$$\sum_{i=1}^n (\kappa_i - \alpha_i)^2 = \sum_{i=1}^n \kappa_i^2 - \sum_{i=1}^n \alpha_i^2 \geq 0$$

che mostra la proprietà di minimo dei coefficienti  $\alpha_i$ . Inoltre, è possibile verificare che

$$[\alpha^2] = \sum_{i=1}^n \alpha_i^2 \quad [\alpha\beta] = \sum_{i=1}^n \alpha_i \beta_i \quad \text{ecc.}$$

Ritornando alla teoria degli errori il sistema (8.23), Gauss osserva che, essendo ragionevole supporre  $v_1, v_2, \dots, v_n \simeq 0$ , il valore plausibile per  $x$  è dato da  $A$ ,

per la (8.32), il che equivale a richiedere, per la (8.30),  $\xi = \eta = \zeta = \dots = 0$ . Ora la funzione

$$\Omega := v_1^2 + v_2^2 + \dots + v_n^2$$

è minima quando

$$\frac{\partial \Omega}{\partial x_1} = \frac{\partial \Omega}{\partial x_2} = \dots = \frac{\partial \Omega}{\partial x_p} = 0.$$

Abbiamo

$$\frac{\partial \Omega}{\partial x_1} = 2[v_1 a_1 + v_2 a_2 + v_3 a_3 + \dots + v_n a_n] = 2\xi,$$

$$\frac{\partial \Omega}{\partial x_2} = 2\eta, \quad \frac{\partial \Omega}{\partial x_3} = 2\zeta$$

e così via sino ad esaurire le variabili. Dunque, con il processo di risoluzione del sistema lineare, Gauss ha trovato che il valore plausibile per  $x$ , cioè  $A$ , e i valori che si ottengono similmente per le altre variabili, sono anche quelli che permettono di minimizzare la somma dei quadrati degli errori  $v_i$ , fornendo in questo modo la giustificazione cercata del metodo dei minimi quadrati.

Come applicazione semplice ma di uso frequente, Gauss illustrò ([13], pp. 29-30) la tecnica di soluzione nel caso in cui  $p = 1$  che corrisponde al caso in cui vi sia una sola variabile  $x_1$  che viene misurata  $n$  volte, fornendo le equazioni

$$v_i = x_i + \ell_i.$$

In questo caso, in cui le  $a_i$  sono tutte uguali ad 1 mentre le altre costanti  $b_i = c_i = \dots = p_i$  sono nulle, si ha

$$\xi = nx_1 + \sum \ell_i$$

da cui si ottiene, imponendo  $\xi = 0$ , come valore plausibile per  $x_1$  la media aritmetica degli  $n$  valori misurati:

$$x_1 = -\frac{\sum \ell_i}{n} = \frac{\sum L_i}{n}.$$

L'interazione a distanza tra Gauss e Laplace ha condotto da un lato a spiegare la centralità della distribuzione normale in virtù del fatto che essa rappresenta un limite verso cui tende la somma e quindi la media aritmetica di una serie di osservazioni. D'altro lato ha spinto Gauss a mettere da parte il principio della media aritmetica come fondamento della distribuzione normale e a cercare una legittimazione alternativa del metodo dei minimi quadrati in una condizione di minimo scarto quadratico medio, evitando la richiesta che le osservazioni sino in numero molto grande, come necessario supporre per applicare il teorema centrale del limite, nella formulazione di Laplace.

## 8.9 Appendice

In questa appendice vediamo come Augustin-Louis Cauchy ([5], pp. 104-106) risolse nel 1821 l'equazione funzionale

$$\varphi(x + y) = \varphi(x) + \varphi(y) \quad (8.36)$$

nella classe delle funzioni  $\varphi(x)$  *continue* sulla retta reale. Cauchy osservò anzitutto che, applicando ripetutamente la (8.36), si ha

$$\varphi(x + y + z + u + \dots) = \varphi(x) + \varphi(y) + \varphi(z) + \dots$$

per cui, se gli argomenti  $x, y, z, u, \dots$  sono in numero qualsiasi  $m$  e sono anche tutti uguali ad uno stesso numero reale  $\alpha$ , deve valere

$$\varphi(m\alpha) = m\varphi(\alpha).$$

Si consideri ora  $\beta := \frac{m}{n}\alpha$ , con  $n > 0$  un altro numero intero. Essendo allora  $n\beta = m\alpha$ , per quanto visto sopra si ha

$$m\varphi(\alpha) = \varphi(m\alpha) = \varphi(n\beta) = n\varphi(\beta)$$

per cui

$$\varphi\left(\frac{m}{n}\alpha\right) = \frac{m}{n}\varphi(\alpha)$$

per ogni razionale positivo. Se ora, preso un numero reale  $\mu > 0$  qualsiasi, si applica la relazione trovata a tutti i termini di una successione di numeri razionali che converge a  $\mu$ , si ottiene, grazie alla continuità di  $\varphi(x)$ ,

$$\varphi(\mu\alpha) = \mu\varphi(\alpha).$$

In particolare, quando  $\alpha = 1$  si ricava

$$\varphi(\mu) = \mu\varphi(1) \quad \forall \mu \in \mathbb{R}_+ \quad (8.37)$$

e, facendo tendere a zero  $\mu$ , si ottiene

$$\varphi(0) = 0.$$

Se poi nell'equazione (8.36) poniamo  $x = \mu > 0$  ed  $y = -\mu$  si giunge a

$$\varphi(-\mu) = \varphi(0) - \varphi(\mu) = -\mu\varphi(1)$$

che, grazie a (8.37), mostra come  $\forall x \in \mathbb{R}$  si debba avere

$$\varphi(x) = x\varphi(1) = kx$$

dove  $k = \varphi(1)$  è una costante.

# Bibliografia

- [1] R. Adrain: Research concerning the probabilities of the errors which happen in making observations, &c. *The Analyst*, **1**, (1808), 93-109.
- [2] D. Bernoulli: Diudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formanda. *Acta Academiae Scientiarum Petropolitanae*, **7**, 2-23, (1778).
- [3] I.-J. Bienaymé: Sur la probabilité des erreurs d'après la méthode des moindres carrés. *Journal de mathématiques pures et appliquées*, **17** (S. I), 33-78, (1852).
- [4] R. J. Boscovich, C. Maire: *Voyage astronomique et géographique dans l'État de l'Eglise*. Tilliard, Paris, (1770), pp. 501-512.
- [5] L.-A. Cauchy: *Cours d'Analyse de l'École Royale Polytechnique*. I Partie: *Analyse algébrique*. Debure, Paris, (1821).
- [6] E. Czuber: *Theorie der Beobachtungsfehler*. Teubner, Leipzig, (1891).
- [7] A.W.F. Edwards: *Likelihood*. The Johns Hopkins University Press, Baltimore and London, (1992).
- [8] R.L. Ellis: On the Method of Least Squares. *Transactions of the Cambridge Philosophical Society*, **8**, 204-219, (1844).
- [9] R.L. Ellis: Remarks on an alleged proof of the "Method of Least Squares", contained in a late Number of the Edinburgh Review. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, **37** (S. III), 321-328, (1850).
- [10] R.W. Farebrother: *Fitting Linear Relationships. A History of the Calculus of Observations 1750-1900*. Springer, New York, 1999.
- [11] G. Galilei: *Dialogo sopra i due massimi sistemi del mondo*. A cura di L. Sosio. Einaudi, Torino, (1970).
- [12] C.F. Gauss: *Theoria Motus corporum coelestium in sectionibus conicis solem ambientium*. Perthes, Amburgo, (1809).

- [13] C.F. Gauss: *Theoria combinationis observationum erroribus minime obnoxiae*. Dieterich, Göttingen, (1823).
- [14] J.-H. Lambert: *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Ved. Klett, Augusta, (1760).
- [15] J.-L. Lagrange: Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations. *Miscellanea Tauriniensa*, **5**, (1770-1773), In *Œuvres de Lagrange*. Vol. II, Gauthier-Villars, Paris, (1868), 173-234.
- [16] P. S. de Laplace: Sur la probabilité des causes par les événements. *Mémoires de l'Académie royale des Sciences de Paris*, **6**, 621-656, (1774). In *Œuvres de Laplace*. Vol. VIII, Gauthier-Villars, Paris, (1891), 27-65.
- [17] P. S. de Laplace: Sur quelques points du système du monde. *Mémoires de l'Académie royale des Sciences de Paris*, **6**, 1-87 (1789). In *Œuvres de Laplace*. Vol. XI, Gauthier-Villars, Paris, (1895), 477-558.
- [18] P. S. de Laplace: Mémoire sur les approximations des formules qui sont fonctions de très grand nombres et sur leur application aux probabilités. *Mémoires de l'Académie royale des Sciences de Paris*, **10**, 209-291, (1809-10). In *Œuvres de Laplace*. Vol. XII, Gauthier-Villars, Paris, (1898), 301-345.
- [19] P. S. de Laplace: Supplément au mémoire sur les approximations des formules qui sont fonctions de très grand nombres. *Mémoires de l'Académie royale des Sciences de Paris*, **10**, (1809-10). In *Œuvres de Laplace*. Vol. XII, Gauthier-Villars, Paris, (1898), 349-353.
- [20] A.-M. Legendre: *Nouvelles méthodes pour la détermination des orbites des comètes*. Didot, Paris, (1805).
- [21] K. Pearson: *The History of Statistics in the 17th & 18th Centuries against the changing background of intellectual, scientific and religious thought*. E.S. Pearson, editor. Griffin & Co., London (1978).
- [22] K. Pearson: Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London*, **185**, 71-110, (1894).
- [23] P. Pizzetti: I fondamenti matematici per la critica dei risultati sperimentali. In *Quarto centenario colombiano. Atti della Regia Università di Genova*. Tip. R. Istituto sordo-muti, Genova, (1892), pp. 113-333.
- [24] R.L. Plackett. Studies in the History of Probability and Statistics. XXIX The discovery of the method of least squares. *Biometrika*, **59**, 239-251, (1972).

- [25] S.-D. Poisson: *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris, Bachelier, (1837).
- [26] G. Pólya: Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Mathematische Zeitschrift*, **8**, 171-181, (1920).
- [27] E. Proverbio: R.G. Boscovich's Determination of Instrumental Errors in Observation. *Archive for History of Exact Sciences*, **38**, 135-152.
- [28] H.L. Seal: Studies in the History of Probability and Statistics. XV: The Historical Development of the Gauss Linear Model. *Biometrika*, **54**, 1-24 (1967).
- [29] O.B. Sheynin: J.H. Lambert's Work on Probability. *Archive for History of Exact Sciences*, **7**, 244-256, (1971).
- [30] O.B. Sheynin: R.J. Boscovich's Work on Probability. *Archive for History of Exact Sciences*, **9**, 306-324, (1973).
- [31] O.B. Sheynin: C.F. Gauss and the Theory of Errors. *Archive for History of Exact Sciences*, **20**, 21-72, (1979).
- [32] E. Shoemith: Thomas Simpson and the Arithmetic Mean. *Historia Mathematica*, **12**, 352-355 (1985).
- [33] T. Simpson: A Letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the Advantage of Taking the Mean of a Number of Observations, in Practical Astronomy. *Philosophical Transactions of the Royal Society of London*, **49**, (1755-1756), 82-93.
- [34] T. Simpson: *Miscellaneous Tracts on some curious, very interesting subjects*. Nourse, London, (1757).
- [35] S.M. Stigler: Laplace's 1774 memoir on inverse probability. *Statistical Science*, **1**, (1986), 359-363.
- [36] S.M. Stigler: *History of Statistics. The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, Massachusetts, U.S.A., (1986).
- [37] S.A. Wepster: *Between Theory and Observations. Tobias Mayer's Explorations of Lunar Motion, 1751-1755*. Springer, New York, (2010).



## Capitolo 9

# Sviluppi tecnici e crisi concettuali nel XIX secolo

### 9.1 La legge dei grandi numeri

Uno dei risultati più importanti presenti nelle *Recherches* pubblicate da Poisson nel 1837 è la “legge dei grandi numeri” (*lois des grands nombres* [25], p. 7), termine che fu introdotto proprio da Poisson per indicare una legge universale, cui sono sottomesse le cose di ogni tipo (*les choses de toutes natures*):

Essa consiste in questo: se si osserva un numero molto considerevole di eventi della stessa natura che dipendono da cause costanti come da cause che variano in modo irregolare, in un senso o nell'altro, cioè a dire che non variano in modo sistematico in un verso prestabilito, si troveranno, in questo numero di osservazioni, dei rapporti pressoché costanti. Per ciascun caso, questi rapporti avranno un valore speciale verso cui si avvicineranno sempre più, nella misura in cui la serie di eventi osservati aumenterà, e che raggiungerebbero rigorosamente se fosse possibile prolungare questa serie di osservazioni all'infinito.<sup>1</sup> ([25], p. 7)

Poisson fornì anche alcuni esempi per chiarire cosa intendesse enunciando la legge dei grandi numeri ([19], p. 272):

- sono date  $\nu$  urne, ciascuna delle quali contiene palline bianche e nere, in proporzione diversa. Se ne sceglie una a caso e si estrae una pallina, rimettendola nell'urna e si ripete l'estrazione un certo numero  $\mu \gg \nu$  di

---

<sup>1</sup>Elle consiste en ce que, si l'on observe des nombres très considérables d'événements d'une même nature, dépendants de causes constantes et des causes qui varient irrégulièrement, tantôt dans un sens, tantôt dans l'autre, c'est à dire sans que leur variation soit progressive dans aucune sens déterminé, on trouvera, dans ces nombres, des rapports à très peu près constants. Pour chaque nature de choses, ces rapports auront une valeur spéciale dont il s'écarteront de moins en moins, à mesure que la série des événements observés augmentera davantage, et qu'ils atteindraient rigoureusement s'il était possible de prolonger cette série à l'infini.

volte. Se in  $m$  estrazioni è uscita una pallina bianca, il rapporto  $\frac{m}{\mu}$  non cambia molto quando si ripete l'intero procedimento ([16], pp. 145-146);

- Si lanciano  $m$  volte delle monete da cinque franchi e si prende nota del numero di volte in cui sulla moneta  $i$ -esima è uscita testa; si ripete lo stesso esperimento lanciando  $m'$  volte le monete e registrando il numero  $a'_i$  di volte in cui la  $i$ -esima moneta ha dato testa per esito. I rapporti  $\frac{a_i}{m}$  ed  $\frac{a'_i}{m'}$  non varieranno molto, a patto che  $m$  ed  $m'$  siano sufficientemente grandi ([16], pp. 147-148);
- la nascita di un figlio maschio piuttosto che femmina all'interno di un matrimonio. Considerando tutti i nati in Francia, la preponderanza dei maschi è confermata in diversi anni e la frazione di maschi sul totale dei neonati si mantiene pressoché costante ([16], pp. 154-155):

Cosa che fornisce una notevole verifica della *legge dei grandi numeri*.<sup>2</sup>

La legge dei grandi numeri non fu accettata in modo unanime dagli studiosi nel XIX secolo ed anzi la sua validità, almeno nella forma vaga vista sopra, fu contestata da alcuni come Bienaymé.

Dal punto di vista matematico, per Poisson la legge dei grandi numeri è un'estensione del teorema di Jakob Bernoulli al caso in cui un evento abbia una probabilità che cambia di prova in prova, pur rimanendo le prove tra loro indipendenti. Supponiamo di eseguire una serie di  $\mu$  prove e che la probabilità che si presenti l'evento  $E$  sia  $p_1$  alla prima prova,  $p_2$  alla seconda, ...,  $p_\mu$  alla  $\mu$ -esima prova. Introdotte le probabilità medie (*chances moyennes*)

$$p' = \frac{p_1 + p_2 + \cdots + p_\mu}{\mu} \quad q' = \frac{q_1 + q_2 + \cdots + q_\mu}{\mu},$$

dove  $q_i = 1 - p_i$ , se  $m$  ed  $n$  sono il numero di prove in cui  $E$  o il suo complementare si sono presentati, allora i rapporti  $\frac{m}{\mu}$  e  $\frac{n}{\mu}$  tendono rispettivamente a  $p'$  e  $q'$  quando il numero di prove tende all'infinito. Quando l'evento  $E$  può verificarsi per delle cause  $C_j$  ( $j = 1, \dots, \nu$ , con  $\nu > 1$ ) mutuamente esclusive la cui distribuzione a priori è uniforme ( $P(C_j) = \frac{1}{\nu}$ ) e si pone

$$P(E|C_j) = c_j,$$

allora  $\frac{m}{\mu}$  tenderà, quando  $\mu \gg 1$ , al valore

$$\frac{1}{\nu} \sum_{j=1}^{\nu} c_j.$$

Per Poisson questo risultato è il tramite per concludere che, se in un'altra serie di  $\mu' \gg 1$  prove l'evento  $E$  si presenta  $m'$  volte, allora con buona approssimazione si avrà

$$\frac{m}{\mu} = \frac{m'}{\mu'},$$

<sup>2</sup>Ce qui présente une vérification remarquable de la *lois des grands nombres*

risultato che viene esteso al caso in cui le cause non siano distribuite uniformemente a priori oppure che siano in numero infinito.

È questa uguaglianza che per Poisson costituisce uno degli aspetti salienti, tradotto in termini matematici, della legge dei grandi numeri. Le dimostrazioni di Poisson, presentate nel Cap. IV di [25], utilizzano le tecniche analitiche di Laplace e sono basate sul calcolo integrale. In particolare, la dimostrazione di Poisson dipende dalla validità del teorema centrale del limite. Ricordiamo infine che a Poisson si deve anche l'introduzione del concetto di variabile aleatoria; riferendosi ad una prova in cui si possono realizzare  $\lambda \geq 2$  eventi mutuamente esclusivi, egli osservò ([16], p. 141):

Questo caso è quello in cui si considera una cosa  $A$  di natura qualsiasi, suscettibile di assumere un numero  $\lambda$  di valori e tra i quali uno solo si realizzerà in ciascuna prova.<sup>3</sup>

## 9.2 Čebišev e la legge dei grandi numeri

Con questa sezione cambiamo radicalmente la *geografia* del nostro percorso storico, passando a considerare alcuni risultati ottenuti dalla scuola russa di probabilità e, più in generale di matematica, sviluppatasi nel XIX secolo e fiorita soprattutto a S. Pietroburgo a partire da Pafnuti Lvovich Čebišev (1821-1894) che ebbe tra i suoi allievi Andrej Andreevič Markov (1856-1922). L'attenzione nell'Impero russo alla probabilità partì all'inizio del XIX secolo nell'università di Vilnius per poi diffondersi in altre sedi. In un rapporto diretto al Ministero dell'Istruzione scritto negli anni '20 del XIX secolo si osservava l'opportunità di familiarizzare gli studenti di matematica e fisica con il calcolo delle probabilità, viste le applicazioni nel mondo delle assicurazioni ed alla geodesia. A rafforzare queste proposte contribuì anche la posizione del matematico Mikhail Vasilevich Ostrogradskii (1801-1862) che, conseguita la laurea a Kharkov, aveva trascorso un periodo di perfezionamento a Parigi dove aveva seguito lezioni di Cauchy, Fourier e Laplace negli anni tra il 1822 ed il 1828 e dunque era in grado di apprezzare l'importanza della disciplina. Fu però Viktor Yakovlevitch Buniakovsky (1804-1889) [22] a dare un contributo decisivo alla promozione ed alla diffusione della probabilità nell'Impero Russo. È interessante osservare come, volendo scrivere un testo di calcolo delle probabilità, Buniakovsky lamentasse le difficoltà nel trovare i termini appropriati in lingua russa, nella quale non vi era una letteratura settoriale che avesse permesso di codificare una terminologia accettata ([11], p. 173). Buniakovsky scrisse un volume sul calcolo delle probabilità, pubblicato nel 1846, dove si prefisse lo scopo di semplificare l'esposizione della *Théorie analytique* di Laplace e nel quale egli sviluppò diverse tematiche. In ulteriori lavori sull'argomento, Buniakovsky si dedicò ad applicazioni della probabilità alla teoria dei numeri, alla demografia, alla statistica e lavorò alle prime applicazioni alla linguistica. La diffusione della probabilità operata a

<sup>3</sup>Ce cas est celui où l'on considère une chose  $A$  d'une nature quelconque, susceptible d'un nombre  $\lambda$  de valeurs, connues ou inconnues, que je représenterai par  $a_1, a_2, a_3, \dots, a_\lambda$ , et parmi lesquelles une seule devra avoir lieu à chaque épreuve.

vario titolo da Ostrogradsky e Buniakovsky portò frutti importanti con la generazione successiva di matematici, in particolare con Čebišev che, con i suoi metodi analitici, è una figura chiave nel passaggio dalla probabilità classica alla probabilità moderna come disciplina matematica, a dispetto del fatto che egli pubblicò solo quattro lavori riguardanti il calcolo delle probabilità. Non va dimenticato poi che, con la teoria dei numeri, il calcolo delle probabilità fu uno dei temi preferiti nelle sue lezioni universitarie ([11], p. 191; [23]).

L'interesse scientifico di Čebišev si rivolse verso teoremi asintotici come il teorema di Bernoulli, oggetto della sua tesi di dottorato del 1846, in cui già si coglie la cifra stilistica del resto della sua produzione: la predilezione per strumenti tecnici di natura elementare. Oltre a riottenere il teorema di Bernoulli egli trovò delle limitazioni sull'errore commesso ([11], p. 195). Sempre nel 1846 egli propose una dimostrazione alternativa della legge dei grandi numeri, o teorema di Poisson, ritenendo che quella originale mancasse di un accettabile livello di rigore [4]. Čebišev ritornerà una ventina di anni più tardi sul teorema di Poisson, dimostrando un ulteriore risultato, in grado di generalizzare i teoremi di Poisson e Bernoulli. Nella dimostrazione gioca un ruolo importante la disuguaglianza che porta il suo nome e che in effetti era già stata dimostrata da Bienaymé nel 1853. In questa sezione studiamo la dimostrazione di Čebišev nel 1867, che viene ottenuta per variabili aleatorie *discrete*. Il testo originale del teorema è reso da Čebišev in questi termini ([5], p. 177):<sup>4</sup>

Indicate con  $a, b, c, \dots$ , le speranze matematiche delle quantità  $X, Y, Z, \dots$ , e con  $a_1, b_1, c_1, \dots$ , le speranze matematiche dei loro quadrati  $X^2, Y^2, Z^2, \dots$ , la probabilità che la somma  $X + Y + Z + \dots$  sia compresa tra i limiti

$$a + b + c + \dots + \alpha \sqrt{a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots},$$

e

$$a + b + c + \dots - \alpha \sqrt{a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots},$$

sarà sempre maggiore di  $1 - \frac{1}{\alpha^2}$ , qualunque sia  $\alpha$ .<sup>5</sup>

Notiamo come Čebišev parli genericamente di quantità, invece del termine moderno di variabili aleatorie e come egli non menzioni mai la condizione di indipendenza che deve sussistere tra di esse, condizione di cui come vedremo, farà uso. Inoltre, nell'enunciato non è specificato se il numero delle variabili aleatorie debba essere finito o possa essere infinito. Infine, Čebišev assume tacitamente che tutte le quantità  $a, b, c, \dots$  e  $a_1, b_1, c_1, \dots$  siano finite. Se indichiamo con

<sup>4</sup>Ho inserito una sola variante tipografica all'originale di Čebišev, indicando le variabili aleatorie con lettere maiuscole:  $X, Y$  e  $Z$ , anziché minuscole:  $x, y, z$ .

<sup>5</sup>Si l'on designe par  $a, b, c, \dots$ , les espérances mathématiques des quantités  $X, Y, Z, \dots$ , et par  $a_1, b_1, c_1, \dots$ , les espérances mathématiques de leurs carrés  $X^2, Y^2, Z^2, \dots$ , la probabilité que la somme  $X + Y + Z + \dots$  est renfermée entre les limites

$$a + b + c + \dots + \alpha \sqrt{a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots},$$

et

$$a + b + c + \dots - \alpha \sqrt{a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots},$$

sera toujours plus grande que  $1 - \frac{1}{\alpha^2}$ , quel que soit  $\alpha$ .

$E(X)$ ,  $E(Y)$ ,  $E(Z)$ ... le speranze matematiche delle variabili aleatorie coinvolte nel teorema e con  $E(X^2)$ ,  $E(Y^2)$ ,  $E(Z^2)$ ... quelle dei loro quadrati, la disuguaglianza oggetto del teorema di Čebišev diventa<sup>6</sup>

$$P \left\{ \left| \sum X - \sum E(X) \right| \leq \alpha \sqrt{\sum \sigma^2(x)} \right\} > 1 - \frac{1}{\alpha^2} \quad (9.1)$$

se indichiamo con  $\sigma^2(X) := E(X^2) - [E(X)]^2$  la varianza di  $X$ . I valori che le variabili  $X$ ,  $Y$ ,  $Z$ ,... possono assumere sono indicati come

$$X = \{x_1, x_2, \dots, x_\ell\} \quad Y = \{y_1, y_2, \dots, y_m\} \quad Z = \{z_1, z_2, \dots, z_n\}$$

con probabilità date, rispettivamente, da

$$\{p_1, p_2, \dots, p_\ell\} \quad \{q_1, q_2, \dots, q_m\} \quad \{r_1, r_2, \dots, r_n\}$$

cosicché possiamo scrivere:

$$a = \sum_{\lambda=1}^{\ell} p_\lambda x_\lambda \quad b = \sum_{\mu=1}^m q_\mu y_\mu \quad c = \sum_{\nu=1}^n r_\nu z_\nu, \dots, \quad (9.2)$$

$$a_1 = \sum_{\lambda=1}^{\ell} p_\lambda x_\lambda^2 \quad b_1 = \sum_{\mu=1}^m q_\mu y_\mu^2 \quad c_1 = \sum_{\nu=1}^n r_\nu z_\nu^2, \dots, \quad (9.3)$$

e

$$\sum_{\lambda=1}^{\ell} p_\lambda = \sum_{\mu=1}^m q_\mu = \sum_{\nu=1}^n r_\nu = \dots = 1. \quad (9.4)$$

Čebišev considerò la somma

$$S = \sum_{\lambda\mu\nu\dots} [x_\lambda + y_\mu + z_\nu + \dots - (a + b + c + \dots)]^2 p_\lambda q_\mu r_\nu \dots = \sum_{\lambda\mu\nu\dots} C_{\lambda\mu\nu\dots} p_\lambda q_\mu r_\nu \dots$$

sviluppando il generico termine  $C_{\lambda\mu\nu\dots}$ , possiamo distinguere vari tipi di addendi in  $S$ :

$$p_\lambda q_\mu r_\nu \dots (x_\lambda^2 + y_\mu^2 + z_\nu^2 + \dots) \quad (9.5)$$

$$2p_\lambda q_\mu r_\nu \dots (x_\lambda y_\mu + x_\lambda z_\nu + y_\mu z_\nu + \dots) \quad (9.6)$$

$$-2(a + b + c + \dots) p_\lambda q_\mu r_\nu \dots (x_\lambda + y_\mu + z_\nu + \dots) \quad (9.7)$$

e

$$(a + b + c + \dots)^2 p_\lambda q_\mu r_\nu \dots. \quad (9.8)$$

Si inizia a sommare sui valori di  $\lambda$ , facendo uso delle (9.2)-(9.4) per riscrivere i termini precedenti nelle forme

$$\sum_{\lambda} p_\lambda q_\mu r_\nu \dots (x_\lambda^2 + y_\mu^2 + z_\nu^2 + \dots) = a_1 q_\mu r_\nu \dots + q_\mu r_\nu \dots (y_\mu^2 + z_\nu^2 + \dots), \quad (9.9)$$

<sup>6</sup>Se vi è un'unica variabile aleatoria  $X$ , la disuguaglianza (9.1) assume la nota forma  $P \{|X - E(x)| \leq \alpha \sigma\} > 1 - \frac{1}{\alpha^2}$ .

$$2 \sum_{\lambda} p_{\lambda} q_{\mu} r_{\nu} \cdots (x_{\lambda} y_{\mu} + x_{\lambda} z_{\nu} + y_{\mu} z_{\nu} + \cdots) = 2 q_{\mu} r_{\nu} \cdots [a(y_{\mu} + z_{\nu} + \cdots) + y_{\mu} z_{\nu} + \cdots] \quad (9.10)$$

$$-2 \sum_{\lambda} (a + b + c + \cdots) p_{\lambda} q_{\mu} r_{\nu} \cdots (x_{\lambda} + y_{\mu} + z_{\nu} + \cdots) = \quad (9.11)$$

$$-2(a + b + c + \cdots) q_{\mu} r_{\nu} \cdots (a + y_{\mu} + z_{\nu} + \cdots)$$

e

$$(a + b + c + \cdots)^2 \sum_{\lambda} p_{\lambda} q_{\mu} r_{\nu} \cdots = (a + b + c + \cdots)^2 q_{\mu} r_{\nu} \cdots \quad (9.12)$$

Alla stessa maniera si possono svolgere le altre somme ottenendo alla fine, in corrispondenza agli addendi (9.9)-(9.12) i contributi

$$a_1 + b_1 + c_1 + \cdots$$

$$2(ab + ac + bc + \cdots)$$

$$-2(a + b + c + \cdots)(a + b + c + \cdots) = -2(a + b + c + \cdots)^2$$

e

$$(a + b + c + \cdots)^2$$

che, sommati, forniscono il valore

$$S = a_1 + b_1 + c_1 + \cdots + 2(ab + ac + bc + \cdots) - (a + b + c + \cdots)^2 =$$

$$a_1 + b_1 + c_1 + \cdots - (a^2 + b^2 + c^2 + \cdots)$$

da cui Čebišev dedusse

$$\frac{S}{\alpha^2 [a_1 + b_1 + c_1 + \cdots - (a^2 + b^2 + c^2 + \cdots)]} = \frac{1}{\alpha^2}$$

ed osservò che il membro di sinistra diminuisce se in  $S$  si sostituiscono con 0 gli addendi per i quali

$$\frac{C_{\lambda\mu\nu\dots}}{\alpha^2 [a_1 + b_1 + c_1 + \cdots - (a^2 + b^2 + c^2 + \cdots)]} = \quad (9.13)$$

$$\frac{[x_{\lambda} + y_{\mu} + z_{\nu} + \cdots - (a + b + c + \cdots)]^2}{\alpha^2 [a_1 + b_1 + c_1 + \cdots - (a^2 + b^2 + c^2 + \cdots)]} < 1$$

e con 1 gli addendi per i quali

$$\frac{C_{\lambda\mu\nu\dots}}{\alpha^2 [a_1 + b_1 + c_1 + \cdots - (a^2 + b^2 + c^2 + \cdots)]} > 1 \quad (9.14)$$

Pertanto

$$\frac{1}{\alpha^2} > \sum_{\lambda\mu\nu\dots}^* p_{\lambda} q_{\mu} r_{\nu} \cdots$$

dove il simbolo  $\sum^*$  indica che la somma è estesa ai valori di  $\lambda, \mu, \nu, \dots$  per i quali vale la (9.14). Tale somma rappresenta la probabilità  $1 - P$  che la variabile

aleatoria  $X + Y + Z + \dots$  abbia valori che *non* verificano la (9.1). Pertanto la probabilità  $P$  che valgano le limitazioni (9.1) soddisfa

$$P > 1 - \frac{1}{\alpha^2}$$

che completa la dimostrazione. Čebišev applicò subito il teorema al caso di  $N$  variabili aleatorie di cui considerò la media aritmetica

$$\Xi := \frac{1}{N}(X + Y + Z + \dots);$$

preso  $\alpha = \frac{\sqrt{N}}{t}$ , Čebišev può riformulare il teorema appena trovato dicendo che

$$p \left( |\Xi - E(\Xi)| \leq \frac{1}{t} \sqrt{\frac{a_1 + b_1 + c_1 + \dots}{N} - \frac{a^2 + b^2 + c^2 + \dots}{N}} \right) > 1 - \frac{t^2}{N}.$$

quale che sia il valore di  $t$  finito. È a questo punto che Čebišev introdusse l'ipotesi che tutti i valori medi  $a, b, c$ , e le varianze  $a_1, b_1, c_1$  fossero uniformemente limitati per concludere che, a patto di prendere  $t$  sufficientemente grande, ([5], p. 183)

Se le speranze matematiche delle quantità  $U_1, U_2, U_3$  e dei loro quadrati  $U_1^2, U_2^2, U_3^2, \dots$ , non superano un limite finito qualsiasi, la probabilità che la differenza tra la media aritmetica di un numero qualunque  $N$  di queste quantità e la media aritmetica delle loro speranze matematiche, sarà minore di una quantità assegnata, tende all'unità, quando  $N$  tende ad infinito.<sup>7</sup>

In particolare, quando  $U_i$  vale 1 o 0, a seconda che si verifichi o meno un evento  $E$  alla  $i$ -esima prova di un esperimento che consiste di  $N$  prove, la variabile

$$\Xi = \frac{U_1 + U_2 + \dots + U_N}{N}$$

rappresenta la *frequenza* con cui  $E$  si è presentato nelle  $N$  prove. Dette allora  $p_1, p_2, \dots, p_N$  le probabilità che l'evento  $E$  si realizzi alla prima, seconda, ...,  $N$ -esima prova, si ha

$$E(\Xi) = \frac{p_1 + p_2 + \dots + p_N}{N}$$

ed il teorema precedente si rilegge in questi termini ([5], p. 184):

Quando il numero di prove tende all'infinito, si ottiene una probabilità, vicina quanto si vuole all'unità, che la differenza tra la media aritmetica

<sup>7</sup>Si les espérances mathématiques des quantités  $U_1, U_2, U_3$  et de leurs carrés  $U_1^2, U_2^2, U_3^2, \dots$ , ne dépassent pas un limite finite quelconque, la probabilité que la difference entre la moyenne arithmétique d'un nombre  $N$  de ces quantités, et la moyenne arithmétique de leurs espérances mathématiques, sera moindre qu'une quantité donnée, se réduit à l'unité, quand  $N$  devient infini.

delle probabilità di questo evento nelle prove, ed il rapporto tra il numero totale delle ripetizioni dell'evento ed il numero totale delle prove, è minore di una quantità assegnata.<sup>8</sup>

In questo modo, Čebišev ha riottenuto la legge dei grandi numeri in modo da renderla indipendente dal teorema centrale del limite sul quale ritornerà più tardi in un lavoro pubblicato in Russia nel 1887 e sugli *Acta Mathematica* nel 1891.

### 9.3 Le critiche al calcolo delle probabilità

Il XIX secolo fu anche un periodo di analisi critica dei fondamenti del calcolo delle probabilità che prese le mosse dalle applicazioni incontrollate del teorema di Bayes. Il primo ad evidenziare la delicatezza del teorema fu Antoine-Augustin Cournot (1801-1877) che, oltre a matematico, fu un filosofo e si può ritenere il fondatore della teoria matematica dell'economia. Nel 1838, subito dopo la pubblicazione delle *Recherches* di Poisson, egli scrisse una lunga memoria dedicata alle applicazioni della probabilità alla statistica giudiziaria [6] e, alcuni anni più tardi, nel 1843, pubblicò una monografia sul calcolo delle probabilità, dove venne posta con forza la distinzione tra significato *soggettivo* ed *oggettivo* di probabilità, distinzione necessaria per evidenziare la fallacia delle conclusioni cui erano giunti Condorcet e Laplace applicando la teoria della probabilità inversa.

È dunque dal linguaggio dei metafisici che ho preso in prestito senza scrupolo i due epiteti di oggettivo e soggettivo che mi sono stati indispensabili per distinguere radicalmente le due accezioni del termine di probabilità a cui si applicano le combinazioni del calcolo.<sup>9</sup> ([7], p. v)

Per Cournot, il significato oggettivo o, meglio, l'esistenza oggettiva, della probabilità è ciò che misura le possibilità delle cose, del realizzarsi degli eventi. Al contrario le probabilità soggettive sono relative in parte alla conoscenza ed in parte all'ignoranza che abbiamo delle cose e sono

variabili da un essere intelligente ad un altro, a misura delle loro capacità e delle informazioni a loro disposizione.<sup>10</sup> ([7], p. 155)

Notiamo, a questo proposito, che anche Poisson aveva introdotto una distinzione analoga nelle *Recherches*, dove aveva scelto di usare *due* vocaboli diversi per il concetto di probabilità, a seconda che si volesse indicare la sua componente *soggettiva* o quella *oggettiva*. Per Poisson

<sup>8</sup>Lorsque le nombre des épreuves devient infini, on obtient une probabilité, aussi rapprochée que l'on veut de l'unité, que la différence entre la moyenne arithmétique des probabilités de cet événement, au nombre total des épreuves, est moindre que toute quantité donnée.

<sup>9</sup>C'est pourtant à la langue des métaphysiciens que j'ai emprunté sans scrupule les deux épithètes d'objective et de subjective, qui m'étaient nécessaires pour distinguer radicalement les deux acceptions du terme de probabilité, auxquelles s'appliquent les combinaisons du calcul.

<sup>10</sup>variables d'une intelligence à une autre, selon leurs capacités et les données qui leur sont fournies.

La *probabilità* (*probabilité*) di un evento è il motivo che noi abbiamo di credere che esso avrà o abbia avuto luogo.<sup>11</sup> ([22], p. 30)

La natura soggettiva della probabilità (*probabilité*) viene ribadita poco dopo in modo da non lasciar spazio ad equivoci:

Poiché la probabilità (*probabilité*) di un evento dipende dalle conoscenze che noi abbiamo di esso, essa può essere diversa, per uno stesso evento, per persone diverse.<sup>12</sup> ([22], p. 30)

Se, circa la composizione di un'urna, una persona sa solo che essa contiene palline bianche e nere, mentre un'altra persona sa anche che le palline bianche sono in numero maggiore delle nere, la seconda persona ha più motivi per ritenere che un'estrazione condurrà ad una pallina bianca rispetto alla prima persona. Ora, conformandosi al linguaggio corrente, Poisson afferma che userà le parole *chance* e *probabilité* come sinonimi se non quando sia necessario operare una distinzione, nel qual caso egli riserverà

la parola *chance* agli eventi in sé stessi ed indipendentemente dalla conoscenza che ne abbiamo, conservando alla parola probabilità (*probabilité*) la sua precedente definizione<sup>13</sup> ([22], p. 31)

Definita la probabilità, Poisson considera, alla De Moivre, come sua *misura* il rapporto tra casi favorevoli e casi possibili, nell'ipotesi che tutti abbiano la stessa *chance*. Notiamo come la distinzione semantica tra *probabilité* e *chance* permetta a Poisson di eludere, almeno formalmente, la circolarità presente nella definizione classica di probabilità.

Per comprendere i limiti del teorema di Bayes, Cournot ricorse ad un esempio già discusso da Laplace e Poisson. Vi sono tre tipologie di urne contenenti tre palline: ciascuna urna della prima tipologia contiene tre palline bianche; quelle della seconda tipologia ne contengono due bianche ed una nera, quelle della terza ne contengono una bianca e due nere. Ogni tipologia è formata dallo stesso numero di urne. Scelta un'urna  $U_i$  a caso ( $P(U_i) = \frac{1}{3}$ ) si estrae una pallina. Se questa pallina è bianca, quali sono le probabilità che essa provenga da un'urna appartenente alla prima, piuttosto che alla seconda o alla terza categoria? Se  $P(B)$  indica la probabilità di estrarre una pallina bianca, allora

$$P(B|U_1) = 1 \quad P(B|U_2) = \frac{2}{3} \quad P(B|U_3) = \frac{1}{3}$$

e

$$P(B) = \sum_{i=1}^3 P(B|U_i)P(U_i) = \sum_{i=1}^3 P(B \cap U_i) = \frac{1}{3} + \frac{2}{9} + \frac{1}{9} = \frac{2}{3}.$$

<sup>11</sup>La *probabilité* d'un événement est la raison que nous avons de croire qu'il aura ou il a eu lieu.

<sup>12</sup>La probabilité dépendant des connaissances que nous avons sur un événement, elle peut être inégale pour un même événement et pour diverses personnes.

<sup>13</sup>Il mot chance aux événements en eux-mêmes et indépendamment de la connaissance que nous en avons, et l'on conservera au mot probabilité sa définition précédente.

Se il processo di scelta casuale di un'urna e successiva estrazione di una pallina viene ripetuto un numero  $m$  di volte molto grande ci si può attendere, grazie al teorema di Bernoulli, che  $\frac{m}{3}$  volte la pallina bianca estratta provenga da un'urna di tipo  $U_1$ ,  $\frac{2m}{9}$  volte di tipo  $U_2$  e  $\frac{m}{9}$  volte di tipo  $U_3$ . Se tre giocatori scommetteranno, rispettivamente sul realizzarsi di  $B \cap U_1$ ,  $B \cap U_2$  e  $B \cap U_3$ , convenendo di considerare nulla la scommessa ogni volta che viene estratta una pallina nera, allora le cifre da scommettere dovranno stare tra loro come 3:2:1, cioè secondo le possibilità che hanno di successo, possibilità verificabili ripetendo molte volte il gioco. Ora, una applicazione diretta del teorema di Bayes mostra che

$$P(U_1|B) = \frac{1}{2} \quad P(U_2|B) = \frac{1}{3} \quad P(U_3|B) = \frac{1}{6}.$$

Se, estratta una pallina bianca, i tre giocatori scommettessero su quale sia la categoria di urne da cui essa proviene, le cifre da mettere in palio dovrebbero stare ancora nel rapporto 3:2:1, visto che

$$P(U_1|B) = 3P(U_3|B) \quad P(U_2|B) = 2P(U_3|B).$$

Chiosa Cournot:

La parola probabilità è qui intesa in senso oggettivo, come sinonimo di possibilità, di modo che, se si ripetesse la stessa scommessa un grande numero di volte, a parità di circostanze, i numeri di partite vinte da ciascuno dei tre giocatori starebbero tra loro come i numeri 3, 2 ed 1.

È in questo senso che bisogna intendere la regola attribuita a Bayes (...). Nel nostro esempio la causa, ovvero l'avvenimento che si verifica per primo, è la scelta dell'urna da una delle tre categorie: l'evento successivo ed osservato, cioè l'estrazione di una pallina bianca, dovrà avere *a priori* probabilità diverse, a seconda che l'estrazione sia effettuata da un'urna che appartiene ad una piuttosto che all'altra delle categorie. Intesa in questo modo, la regola di Bayes è un teorema che non dà luogo ad equivoci e di cui non si può contestare la correttezza.<sup>14</sup> ([7], pp. 157-158)

In questa formulazione del problema, tutti i dati sono *ben* determinati ma Cournot ora considera una situazione radicalmente diversa, anche se in apparenza simile a quella appena trattata. Si supponga di sapere che le urne contengono ciascuna tre palline—bianche o nere—ma di *ignorare* sia la composizione ( $B, N$ ) in termini del numero di palline bianche e nere, sia la consistenza numerica delle

<sup>14</sup>le mot de probabilité est pris ici dans le sens objectif, comme équivalent de possibilité en sorte que, si les mêmes paris étaient répétés un grand nombre de fois, dans les mêmes circonstances, les nombres de parties gagnées respectivement par chacun des trois joueurs seraient sensiblement dans les rapports des nombres 3, 2 et 1.

C'est en ce sens qu'il faut entendre la règle attribuée à Bayes (...). Dans notre exemple, la cause ou l'événement antécédent est le triage de l'urne dans l'une des trois catégories: l'événement subséquent et observé, c'est-à-dire l'extraction d'une boule blanche, devant avoir *à priori* une probabilité différente, selon que l'urne de tirage appartient à l'une ou à l'autre des catégories. Ainsi entendue, la règle de Bayes est un théorème qui ne donne lieu à aucune équivoque, et dont on ne peut contester la justesse.

urne di diverse categorie che ora, peraltro, sono 4, essendovi anche la possibilità di urne  $U_4$  di composizione  $(0, 3)$ . Si *ignori* inoltre quale sia stato il meccanismo di scelta dell'urna, se a caso oppure se dettato da qualche causa esterna. Quando viene estratta una pallina bianca, si chieda ai tre giocatori di scommettere sulla tipologia dell'urna da cui la pallina è stata estratta. Ora, prima di scegliere l'urna, si possono formulare quattro ipotesi: che sia di tipo  $U_1$ , cioè  $(3, 0)$ ;  $U_2$ , cioè  $(2, 1)$ ;  $U_3$ , cioè  $(1, 2)$ ;  $U_4$ , cioè  $(0, 3)$ , senza che si abbia alcun motivo di preferire una all'altra. Ad ogni ipotesi, nota Cournot, corrispondono 3 "sotto-ipotesi" sulla pallina che potrà essere estratta e, anche qui, non abbiamo alcun motivo di preferire una all'altra. Ora, sapendo che è stata estratta una pallina bianca, l'ipotesi  $U_4$  è chiaramente falsa e, di conseguenza, sono venute meno le tre relative sotto-ipotesi, così come le tre altre sottoipotesi legate all'estrazione di una pallina nera da urne di composizione  $(2, 1)$  o  $(1, 2)$ . Si resta allora con 6 casi tra cui scegliere: 3 relativi alla prima ipotesi (urna di tipo  $(3, 0)$ ), 2 alla seconda ipotesi (urna di tipo  $(2, 1)$ ) ed 1 relativo alla terza ipotesi (urna di tipo  $(1, 2)$ ): le cifre da scommettere dovranno essere ancora nel rapporto 3:2:1 ma Cournot nota una cesura fondamentale rispetto alla situazione precedente:

Se però si dicesse che anche le probabilità di guadagno sono nei rapporti dei numeri 3, 2, 1 e si intendesse in questo senso la regola di Bayes, il termine probabilità verrebbe preso nella sua accezione puramente soggettiva, variabile da persona a persona, relativo in parte alle sue conoscenze, in parte alla sua ignoranza. Ciò non significherebbe affatto che, in un numero molto grande di scommesse, le partite vinte da ciascun giocatore saranno nella proporzione di 3 a 2 a 1.<sup>15</sup> ([7], pp. 159-160)

Il giocatore che scommette sulla provenienza dell'urna da quelle di tipo  $(3, 0)$  potrebbe perdere *sempre* se l'urna è stata estratta da un campione che *non* conteneva questa tipologia o se chi effettua l'estrazione ha deliberatamente escluso urne di quel tipo: si tratta di circostanze che, allo stato delle conoscenze dei giocatori, sono altrettanto plausibili quanto una scelta puramente casuale:

la regola di Bayes, così applicata alla determinazione di probabilità soggettive, non ha altra utilità che quella di fissare una scommessa, sotto una certa ipotesi sulle cose che l'arbitro conosce e su quelle che ignora. Essa condurrà a fissare una scommessa iniqua se l'arbitro conosce più cose di quante si ipotizzi circa le condizioni reali in cui si svolge l'esperimento aleatorio.<sup>16</sup> ([7], p. 160)

---

<sup>15</sup>Mais si l'on dit que les probabilités de gain des trois joueurs sont aussi dans les rapports des nombres 3, 2, 1, et si l'on entend en ce sens la règle de Bayes, le terme de probabilité sera pris dans son acception purement subjective, variable d'un individu à un autre, relative en partie à ses connaissances, en partie à son ignorance. Cela ne signifiera point que, dans un grand nombre de paris semblables, les nombres de paris gagnés respectivement par les joueurs (...) seront, à très peu près, dans les rapports des nombres 3, 2, 1.

<sup>16</sup>La règle de Bayes, ainsi appliquée à la détermination de probabilités subjectives, n'a donc d'autre utilité que celle de conduire à une fixation de paris, dans une certaine hypothèse sur les choses que connaît et sur celles qu'ignore l'arbitre. Elle conduirait à une fixation inique si l'arbitre en savait plus qu'on ne le suppose, sur les conditions réelles de l'épreuve aléatoire.

Cournot effettua una rilettura simile di un'altra applicazione della prima ora della formula di Bayes. Ritorniamo nelle condizioni del primo esempio e supponiamo che, estratta una pallina bianca, la si riponga nell'urna: qual è la probabilità di estrarre una pallina bianca in una seconda estrazione? La soluzione diretta è semplice. Dopo la prima estrazione, le probabilità delle tre ipotesi sono state rimodulate ed ora valgono

$$P(U_1) = \frac{1}{2} = \frac{3}{6} \quad P(U_2) = \frac{1}{3} = \frac{2}{6} \quad P(U_3) = \frac{1}{6}$$

da cui si ottiene la probabilità di estrarre una seconda pallina bianca

$$P(B) = \sum_{i=1}^3 P(B|U_i)P(U_i) = \sum_{i=1}^3 P(B \cap U_i) = \frac{1}{2} + \frac{2}{9} + \frac{1}{18} = \frac{7}{9} :$$

ancora una volta il termine probabilità è inteso in chiave oggettiva:

Se si sceglie a caso un gran numero di volte l'urna eseguendo ogni volta due estrazioni successive dall'urna scelta casualmente; allora, contando il numero  $m$  di prove in cui per prima è stata estratta una pallina bianca e, tra queste, il numero  $n$  di prove nelle quali la seconda estrazione ha condotto ancora ad una pallina bianca il rapporto  $\frac{n}{m}$  differirà poco da  $\frac{7}{9}$ .<sup>17</sup> ([7], p. 161)

Al contrario, se si ignora il meccanismo di scelta dell'urna, le somme da scommettere saranno ancora nel rapporto di 7 a 2

solo in virtù dell'ipotesi fatta circa ciò che è noto e ciò che è incognito della prova aleatoria.<sup>18</sup> ([7], p. 161)

Ciò non può condurre a ritenere che, ripetendo la scommessa un numero molto grande di volte, il rapporto tra il numero di vittorie dei giocatori stia nel rapporto di 7 contro 2.

Ora, osserva Cournot, nelle applicazioni più comuni della regola di Bayes, non si conosce *nulla* circa la composizione dell'urna per cui

si assume che le probabilità (*chances*) possano variare in modo continuo o, in altri termini, che l'urna contenga un'infinità di palline e che il rapporto tra il numero di palline bianche ed il numero totale di palline possa assumere tutti gli infiniti valori compresi tra 0 ed 1.<sup>19</sup> ([7], p. 162)

<sup>17</sup>Si l'on triait un fort grand nombre de fois l'urne au hasard, et qu'on fit chaque fois deux tirages consécutifs avec l'urne choisie fortuitement; si l'on comptait le nombre  $m$  des épreuves où le premier tirage a donné une boule blanche, et parmi celles-ci le nombre  $n$  des épreuves pour lesquelles le second tirage a encore amené une boule blanche, le rapport  $\frac{n}{m}$  différerait peu de  $\frac{7}{9}$ .

<sup>18</sup>mais en raison seulement de l'hypothèse qu'on a faite sur les choses connues et inconnues dans les conditions de l'épreuve aléatoire.

<sup>19</sup>On admet que les chances sont susceptibles de varier d'une manière continue, ou en d'autres termes, que l'urne renferme une infinité de boules, et que le rapport du nombre des boules blanches au nombre total des boules peut prendre toutes les valeurs, en nombre infini, comprises entre zéro et l'unité.

Per mostrare come, in generale, non si debba fare sempre affidamento sui valori numerici ottenuti applicando la regola di Bayes, Cournot considerò il lancio di una moneta, estratta da un mucchio di monete appena fabbricate. Si suppone che sia uscita croce al primo lancio: nessuno però scommetterebbe 2 contro 1 per l'uscita di croce al secondo lancio perché

la probabilità di ottenere *croce* per ogni moneta del mucchio, benché variabile un poco da una moneta all'altra a motivo delle irregolarità nella struttura fisica, non può discostarsi molto da  $\frac{1}{2}$  per ciascuna moneta, in eccesso o in difetto. Sarebbe dunque contrario alle informazioni che abbiamo in questo caso, sulle condizioni del gioco, di ritenere tutti i valori tra 0 ed 1 come attribuibili indifferentemente, *a priori* alla possibilità di ottenere croce.<sup>20</sup> ([7], p. 164)

Le obiezioni di Cournot mettono a nudo il punto debole delle applicazioni della formula di Bayes-Laplace, cioè il significato ed il ruolo del principio di ragione insufficiente che, ricordiamo, è il cardine attorno a cui ruota la visione laplaciana dell'intero edificio del calcolo delle probabilità e che è il motivo che spinse sistematicamente Laplace a supporre le cause, o le ipotesi, a priori uniformi. Cournot mise in discussione il significato di questa ipotesi ed i rischi che si correivano nel considerare l'indifferenza di fronte a più scelte come sinonimo dell'*ignoranza* completa delle condizioni sotto cui un evento può realizzarsi o meno.

## 9.4 I Paradossi di Bertrand

Il matematico francese Joseph Bertrand (1822-1900) inserì nel primo capitolo del suo *Cours de Probabilité*, pubblicato nel<sup>21</sup> 1889, alcuni paradossi che dovevano evidenziare come, estendendo il concetto di probabilità a situazioni in cui i casi possibili sono *infiniti*, si andasse incontro ad una perdita di precisione che generava risoluzioni ambigue. Il paradosso più famoso è un esempio di *probabilità geometrica*, in cui la definizione di probabilità viene *estesa* al continuo. Abbiamo già visto all'opera questa estensione nel lavoro di Bayes quando la probabilità di un evento era stata misurata da un rapporto di aree. Poco più tardi Buffon considerò nell'*Arithmétique Morale* del 1777 [3] il *jeu du franc-carreau*, in cui si supponeva di lanciare una moneta circolare su un piano, suddiviso in modo regolare con piastrelle quadrate, triangolari od esagonali. Il gioco consisteva nello scommettere sul fatto che la moneta, cadendo, *non* avrebbe tagliato linee delle divisioni del piano e la probabilità che questo evento si realizzasse era vista come

<sup>20</sup>la probabilité d'amener *croix*, pour chacune des pièces du tas, quoique variant certainement un peu d'une pièce à l'autre, à cause des irrégularités de structure physique, ne peut pas différer beaucoup pour chacune d'elles de la fraction  $\frac{1}{2}$  en plus ou en moins. Il serait donc contraire aux notions que nous avons dans ce cas, sur les conditions du hasard, de regarder toutes les valeurs comprises entre zéro et l'unité comme pouvant être attribuées indifféremment, *a priori*, à la chance d'amener croix.

<sup>21</sup>La data di pubblicazione è un "giallo" perché, secondo Sheynin, vi sono indizi che esso sia stato pubblicato nel 1888 e che alcune copie vennero datate nel 1889 ([24], p. 157).

un rapporto di aree.<sup>22</sup> Per trovare le probabilità di vittoria dei due giocatori, Buffon tracciò un poligono interno ad uno di quelli che formano la divisione, i cui lati hanno distanza da quelli omologhi del poligono originario pari al raggio della moneta. Il rapporto tra l'area dei due poligoni dà la probabilità di vittoria di chi ha scommesso che la moneta non toccherà alcuna linea di divisione, cioè giacerà su una piastrella (*carreau*) libera (*franc*) da suddivisioni.

Nelle *Recherches*, dopo aver definito la misura della probabilità di un evento come rapporto tra casi favorevoli e casi possibili, Poisson aggiunse un inciso sulle proprietà aritmetiche della probabilità:

Da questa misura di probabilità, sembra risultare che questa frazione debba essere sempre una quantità commensurabile; se però il numero dei casi possibili e quello dei casi favorevoli ad un evento sono infiniti, la probabilità, cioè il rapporto del secondo numero col primo, potrà essere una quantità incommensurabile.<sup>23</sup> ([16], p. 33)

Poisson ha in mente proprio la *probabilità geometrica*: si consideri una superficie piana di area  $s$  ed una sua porzione di area  $\sigma$  e si lanci un oggetto circolare il cui centro  $C$  ha uguali *chance* di ricadere su un punto di  $s$ : è evidente—afferma Poisson—che la probabilità che  $C$  cada su un punto della seconda porzione è

$$\frac{\sigma}{s}.$$

Quali sono allora le insidie di questo passaggio al continuo? Bertrand propose questo problema:

Si tracci *a caso* una corda all'interno di un cerchio. Qual è la probabilità che essa [non] sia più corta del lato del triangolo equilatero inscritto?<sup>24</sup> ([1], p. 4)

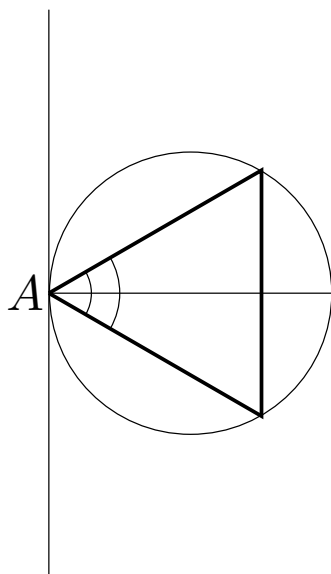
Il punto cruciale per Bertrand è che la locuzione: *a caso* (*au hasard*), può intendersi in modi differenti.

1. Si supponga di fissare un estremo  $A$  della corda. La direzione della corda va scelta a caso. Si tracci il triangolo equilatero inscritto al cerchio, avente un vertice in  $A$  e sia  $\vartheta \in [0, \pi]$  l'angolo che una corda passante per  $A$  forma con la retta tangente alla circonferenza passante per  $A$ . Ora, affinché la corda sia più *lunga* del lato del triangolo equilatero inscritto al cerchio, occorre prendere  $\theta \in [\frac{\pi}{3}, \frac{2\pi}{3}]$ . Supponendo *uniforme* la distribuzione di  $\vartheta$ , la probabilità cercata vale  $\frac{1}{3}$ .

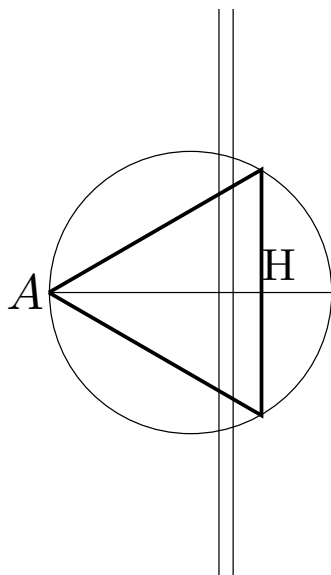
<sup>22</sup>Il secondo giocatore scommette sul fatto che la moneta tagli una delle linee di suddivisione. La versione originale di Buffon prevede quattro giocatori: gli altri due scommettono sul fatto che la moneta coprirà due lati delle divisioni oppure su tre o più lati, rispettivamente

<sup>23</sup>De cette mesure de la probabilité, il semble résulter que cette fraction doit toujours être une quantité commensurable; mais si le nombre de tous les cas possibles et celui des cas favorable à un événement, sont infinis, la probabilité ou le rapport du second nombre au premier, pourra être une quantité incommensurable.

<sup>24</sup>On trace *au hasard* une corde dans un cercle. Quelle est la probabilité pour qu'elle soit plus petite que le côté du triangle équilatéral inscrit?

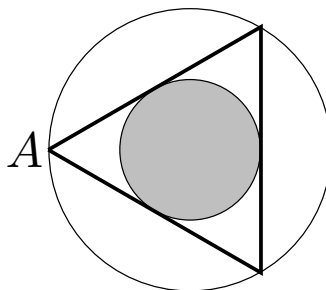


2. Si assegni la direzione della corda e si tracci il diametro ortogonale a questa direzione. Affinché una corda parallela alla direzione assegnata sia più lunga del lato del triangolo equilatero inscritto al cerchio occorre che il suo punto medio  $H$  abbia distanza dal centro  $d \leq \frac{R}{2}$ . Poiché tutti i valori di  $d \in [0, R]$  sono equivalenti, la probabilità cercata è  $\frac{1}{2}$ .



3. Si scelga a caso il punto medio  $H$  della corda non più su un diametro ma liberamente, all'interno del cerchio. Affinché la corda sia più lunga

del lato del triangolo equilatero inscritto al cerchio occorre che  $H$  giaccia all'interno di un cerchio concentrico a quello dato, di raggio  $\frac{R}{2}$ . Nell'ipotesi che tutti i punti siano equivalenti, la probabilità richiesta è il rapporto tra l'area di questo cerchio e di quello originario e vale dunque  $\frac{1}{4}$ :



Qual è la risposta corretta tra queste tre? Nessuna delle tre è falsa, nessuna è vera, è la questione ad essere mal posta.<sup>25</sup> ([1], p. 5)

La presenza di infinite scelte rende dunque il principio di ragione insufficiente incapace di dare una risposta univoca ed è a partire da questa osservazione che l'esempio di Bertrand viene riletto da altri autori [29]. Osserviamo che Bertrand non fu il primo ad osservare i rischi che si corrono quando le scelte che si possono fare *a caso* sono infinite. Per esempio, Charles Darwin condusse nel 1881 un esperimento in cui erano coinvolti tre diversi modi di intendere una scelta aleatoria nel quale le valutazioni di probabilità poggiavano su considerazioni geometriche ([21], p. 350).

John Maynard Keynes (1883-1946) propose nel suo trattato di calcolo delle probabilità [9] un altro esempio, tratto in effetti da un'opera di Johannes von Kries (1853-1928) [10], in cui non vi sono considerazioni geometriche. Sia data una miscela di due liquidi che, seguendo l'esposizione di Gillies [29], supporremo essere acqua e vino. Di questa miscela si sa solo che uno dei due componenti non può essere più del triplo dell'altro. Si vuole trovare la probabilità che

$$\frac{1}{3} \leq \frac{\text{acqua}}{\text{vino}} \leq 2.$$

Poiché il rapporto tra le quantità di acqua e vino può assumere tutti i valori nell'intervallo  $[\frac{1}{3}, 3]$  e non c'è nessun motivo per preferirne uno all'altro, si assuma per questo rapporto una distribuzione uniforme con densità costante  $\frac{1}{3-\frac{1}{3}} = \frac{3}{8}$ . Dunque

$$P\left(\frac{1}{3} \leq \frac{\text{acqua}}{\text{vino}} \leq 2\right) = \frac{3}{8} \times \left(2 - \frac{1}{3}\right) = \frac{5}{8}$$

Calcoliamo ora  $P\left(\frac{1}{2} \leq \frac{\text{vino}}{\text{acqua}} \leq 3\right)$  osservando che anche la frazione  $\frac{\text{vino}}{\text{acqua}}$  deve avere densità di probabilità costante e sempre pari  $\frac{3}{8}$ , in ossequio al principio di

<sup>25</sup>Entre ces trois réponses, quelle est la véritable? Aucune des trois n'est fausse, aucune n'est exacte, la question est mal posée.

ragione insufficiente. Abbiamo ora

$$P\left(\frac{1}{2} \leq \frac{\text{vino}}{\text{acqua}} \leq 3\right) = \frac{3}{8} \left(3 - \frac{1}{2}\right) = \frac{15}{16} :$$

eppure i due “eventi” sono *lo stesso* evento, formulato in due modi equivalenti. Il messaggio che questi esempi vogliono convogliare è che *ignorare* le condizioni in cui si svolge un fenomeno *non* è sinonimo di indifferenza rispetto alle scelte possibili.

## 9.5 Le istanze della fisica e l'opera di Poincaré

GLi sviluppi più importanti nella fisica del XIX secolo riguardarono i fenomeni termici, l'elettricità ed il magnetismo, questi ultimi sintetizzati da Maxwell nell'elettromagnetismo. Per i nostri obiettivi è importante tracciare brevemente gli sviluppi che condurranno alla termodinamica, per i legami stringenti e concettualmente problematici di questa disciplina con la probabilità. Lo studio dei fenomeni termici non nacque nell'Ottocento ma fu a metà di questo secolo che si consolidò un cambiamento radicale di interpretazione: il calore non è più visto come una *sostanza* che fluisce all'interno dei corpi (calorico) ma è una forma di energia legata al movimento dei costituenti elementari dei corpi, le molecole, che si muovono in modo deterministico seguendo le leggi della meccanica newtoniana.<sup>26</sup> Occorre ricordare che lo studio dei fenomeni termici nell'Ottocento prese due vie diverse, con una propria evoluzione. Da una parte, lo studio della propagazione del calore per conduzione nei mezzi omogenei che venne affrontato soprattutto con gli strumenti messi a disposizione da Jean-Baptiste Joseph Fourier (1768-1830) nella *Théorie analytique de la Chaleur*, pubblicata nel 1822, in cui *non* si fa alcuna ipotesi sulla natura del calore, in un momento in cui la teoria del calorico doveva già affrontare critiche serie. Lo studio dell'equazione del calore

$$\frac{\partial u}{\partial t} = \kappa \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right)$$

poggia soprattutto sugli sviluppi delle soluzioni di problemi al contorno ricorrendo alle serie trigonometriche il cui studio, giova ricordarlo, fornì abbondanza di controesempi all'analisi matematica, fu alla base delle ricerche di Georg Cantor (1845-1918) sulla teoria degli insiemi. La teoria della propagazione termica rappresenta un esempio di *annessione* di un ramo della fisica alla matematica. Osserviamo come il ruolo dell'equazione del calore in probabilità non sia stato percepito che verso la fine del XIX secolo.

A fianco ed indipendentemente dalla teoria analitica del calore si sviluppò la teoria meccanica del calore, nome che mette in risalto forse più di quello attualmente adoperato di termodinamica, l'oggetto di studio: il calore in quanto

<sup>26</sup>Per sgombrare il campo da equivoci, è bene ricordare che le molecole dell'Ottocento sono immaginate come le porzioni elementari dei corpi ma non hanno un corrispettivo sperimentalmente verificabile.

forma di movimento. Ora, se la prima legge della termodinamica si può formulare come legge di conservazione dell'energia, estesa ad abbracciare il calore oltre alle varie forme di energia meccanica, la seconda legge della termodinamica si fonda sul concetto di *entropia*, introdotto da Rudolph Clausius (1811-1888): l'entropia, a differenza dell'energia, in un sistema isolato come l'universo, non si mantiene costante ma *cresce* sempre. Ora, nello sviluppo parallelo della teoria cinetica dei gas, è possibile fornire una interpretazione *statistica* di grandezze macroscopiche come la pressione e la temperatura, legandole, rispettivamente, alla quantità di moto ed all'energia cinetica di traslazione di molecole concepite come punti materiali e riottenere in questo modo l'equazione di stato dei gas perfetti, nota come equazione di Boyle. Nella prima formulazione della teoria cinetica, ad opera di August Krönig (1822-1879) e Clausius, pressione e temperatura sono valori medi del momento e dell'energia cinetica associate al complesso moto delle molecole ma non vi è alcuno studio di una eventuale *legge di distribuzione* delle velocità, in condizioni di equilibrio macroscopico. Il primo a porsi il problema di determinare tale legge fu il fisico scozzese James Clerk Maxwell (1831-1879) che dedusse la distribuzione delle velocità una prima volta seguendo sostanzialmente gli argomenti addotti da Herschel per ricavare la teoria degli errori, salvo poi proporre una nuova deduzione che teneva in debito conto gli urti molecolari. Ora, l'introduzione dell'entropia è subordinata all'accettazione del postulato di Clausius secondo cui è impossibile realizzare una trasformazione termodinamica che abbia come risultato finale *solo* il passaggio di calore da un corpo ad un altro a temperatura *maggiore*. Attorno al 1870 diversi studiosi, come Clausius e Ludwig Boltzmann (1844-1906), proposero delle complesse deduzioni del secondo principio della termodinamica, basate solo sulla meccanica, ottenendo una funzione che aveva le proprietà dell'entropia e che era ottenuta da operazioni di media su quantità legate ai moti molecolari. A parte il fatto che queste deduzioni erano tecnicamente molto onerose, il numero e la natura delle ipotesi imposte ai moti molecolari lasciavano molto insoddisfatti. L'intera proposta di giustificare su basi puramente meccaniche il secondo principio della termodinamica fu anche attaccata alla radice dalle obiezioni sull'irreversibilità. L'entropia cresce in trasformazioni cicliche *irreversibili* mentre le leggi della meccanica classica, invarianti rispetto alla sostituzione della variabile temporale  $t$  in  $-t$ , non possono rendere conto di questa circostanza. Si poneva dunque una alternativa: o la meccanica classica era inadeguata o il secondo principio non era corretto. Una possibile via di uscita venne proposta da Maxwell attraverso il celebre esempio del *demone*. Si supponga di avere due gas, a temperature inizialmente diverse, a contatto termico attraverso un setto. Supponiamo sia praticabile una piccola apertura nel setto che possa essere aperta o chiusa da un "demone" che agisce in questo modo: ogni volta in cui una molecola "veloce" del gas a temperatura più bassa transita davanti alla fenditura, questa viene aperta per permetterle il passaggio nella regione occupata dal gas a temperatura maggiore e poi subito richiusa; se una molecola "lenta" nel contenitore del gas a temperatura maggiore passa davanti alla fenditura, il demone la apre per consentirle il passaggio nella regione occupata dal gas a temperatura minore e poi essa viene subito richiusa. Poiché la temperatura è associata all'energia cinetica

dei moti molecolari, l'azione del demone ha come solo effetto finale quello di far passare calore dal gas a temperatura minore a quello a temperatura maggiore, violando il secondo principio che non è dunque una legge *universale* di natura: moti molecolari che consentano questo sbilanciamento non sono impossibili ma *estremamente* improbabili e dunque il principio di Clausius e la seconda legge della termodinamica diventano leggi *statistiche*. Con la teoria cinetica dei gas la probabilità entra dunque nella formulazione del modello e non solo nell'analisi dei dati sperimentali per un controllo *a posteriori* dell'accordo tra teoria ed esperimento, come avviene nella teoria degli errori.

La contraddizione tra una visione puramente meccanicista e la termodinamica venne posta in evidenza anche da Henri Poincaré (1854-1912) in un breve scritto pubblicato nel 1893:

Un teorema facile da dimostrare ci insegna che un mondo limitato, soggetto alle sole leggi della meccanica, ripasserà sempre in uno stato vicino a quello iniziale. Al contrario, secondo le leggi sperimentali ammesse (se attribuiamo loro un valore assoluto e se vogliamo spingerne le conseguenze fino in fondo), l'universo tende verso uno stato finale da cui non potrà più uscire. In questo stato finale, che sarà una sorta di morte, i corpi saranno in quiete ed alla stessa temperatura.<sup>27</sup> ([12], p. 536)

Poincaré, che teneva la cattedra di Calcolo delle probabilità e fisica matematica alla Sorbona di Parigi, si riferisce alla “morte entropica” come conseguenza ultima dell'irreversibilità dei fenomeni naturali mentre nella premessa si riferisce al teorema noto come “teorema del ritorno” da lui formulato nel monumentale lavoro con cui aveva vinto alcuni anni prima il Premio per la matematica istituito dal re di Svezia e Norvegia Oscar II e che ricomparirà nel terzo volume dei *Méthodes nouvelles de la Mécanique céleste*, pubblicato nel 1899 [14]: con questi lavori di straordinaria ricchezza si può dire nasca la teoria dei sistemi dinamici nella sua connotazione moderna. Il teorema del ritorno è interessante perché apre la strada per trovare un esempio di sistema deterministico cui compete una proprietà con probabilità 1. Poincaré deve discutere la stabilità delle soluzioni del problema dei tre corpi: tre punti materiali che si attraggono con forze che sono funzioni della loro distanza. Vi sono tre requisiti che appaiono naturali affinché vi possa essere una stabilità del sistema:

1. I tre corpi non possono allontanarsi oltre ogni limite nel corso del moto, cioè il loro moto deve avvenire in una regione *confinata*;
2. nessuna coppia di corpi può entrare in collisione;
3. il sistema ritornerà una infinità di volte in un intorno della configurazione iniziale, requisito che è certamente rispettato nel caso dei moti periodici.

---

<sup>27</sup>Un théorème facile à établir nous apprend qu'un monde limité soumis aux seules lois de la mécanique, repassera toujours par un état très voisin de son état initial. Au contraire, d'après les lois expérimentales admises (si on leur attribue une valeur absolue et qu'on veuille en pousser les conséquences jusqu'au bout), l'univers tend vers un certain état final dont il ne pourra plus sortir. Dans cet état final, qui sera une sorte de mort, tous les corps seront en repos et à la même température.

Un sistema che obbedisce alla sola terza condizione è detto da Poincaré stabile alla Poisson. Poincaré intendeva mostrare che questa condizione era soddisfatta quando lo era la prima. Per illustrare il risultato, egli considerò un fluido incompressibile<sup>28</sup> che esegue un moto stazionario restando dentro una regione  $\mathcal{B}$  di volume  $V$  ed una porzione  $U_0 \in \mathcal{B}$  occupata da certe molecole di fluido all'istante  $t = 0$ . Questa regione evolve, all'istante  $\tau$ , in una regione  $U_1 \in \mathcal{B}$  che ha lo stesso volume di  $U_0$  e, ad istanti successivi  $2\tau, 3\tau, \dots, n\tau$ , le molecole inizialmente in  $U_0$  occuperanno delle regioni  $U_2, U_3, \dots, U_n$ , tutte con lo stesso volume. Ora, essendo  $V$  finito, non è possibile che tutte queste regioni siano disgiunte tra loro quando  $n$  è tale che

$$V < (n + 1)U_0,$$

dove  $U_0$  indica, con un abuso di notazione, anche il volume della regione omonima. Siano  $U_i$  ed  $U_k$ , con  $k > i$  due regioni che hanno parti in comune:  $U_i \cap U_k \neq \emptyset$ . Poiché il sistema è retto da equazioni autonome, cioè non dipendenti esplicitamente dal tempo, ciò significa che  $U_0 \cap U_{k-i} \neq \emptyset$ , ovvero che esiste un indice  $\alpha > 0$  tale che  $U_0 \cap U_\alpha \neq \emptyset$ . Sia ora  $U'_0 := U_0 \cap U_\alpha$  e, come appena fatto per  $U_0$ , seguiamo l'evoluzione delle molecole che, al tempo  $t = 0$  sono in  $U'_0$ . Esisterà un insieme  $U'_\beta$ , di ugual volume a  $U'_0$  perché sua immagine nel moto, tale che  $U'_0 \cap U'_\beta = U''_0$ . Iterando questo schema si viene a determinare una famiglia di insiemi

$$U_0 \supseteq U'_0 \supseteq U''_0 \supseteq \dots \supseteq U_0^{(p)} \supseteq \dots$$

tali che  $U_0^{(p)}$  non è mai vuoto, per quanto grande si possa prendere  $p$ . Detto  $E$  l'insieme limite agli insiemi  $U_0^{(p)}$ , quando  $p \rightarrow \infty$ , i punti di  $E$  sono tali che le molecole che vi si trovano all'istante  $t = 0$  vi ritorneranno una infinità di volte. Le considerazioni più interessanti per i nostri scopi sono altre. Poincaré osservò come, in generale, si possono concepire anche traiettorie che attraverseranno la frontiera della regione  $U_0$  solo un numero *finito*  $k$  di volte: restando nel quadro del problema dei tre corpi, queste traiettorie sono legate a moti che hanno una meta asintotica. Poincaré tuttavia affermò:

Mi propongo di dimostrare che queste ultime debbono essere considerate come eccezionali ovvero, per essere più precisi, che la *probabilità* che una molecola non attraversi  $U_0$  se non un numero finito di volte è infinitamente piccola *se si ammette* che questa molecola si trovi all'interno di  $U_0$  all'origine dei tempi.<sup>29</sup> ([14], p. 151)

Poincaré precisa subito in che senso bisogna intendere la parola probabilità: considerata una funzione  $\varphi(x, y, z) > 0$  delle coordinate cartesiane ortogonali

<sup>28</sup>Si tratta di una analogia: le equazioni di Hamilton per sistemi soggetti a vincoli scleronomi conservano il volume nello spazio delle fasi.

<sup>29</sup>Je me propose de montrer que ces dernières doivent être regardées comme exceptionnelles ou, pour préciser davantage, que la *probabilité* pour qu'une molécule ne traverse  $U_0$  qu'un nombre fini des fois est infiniment petite, *si l'on admet* que cette molécule est à l'intérieur de  $U_0$  à l'origine des temps.

$(x, y, z)$  di un punto qualsiasi di  $\mathcal{B}$ , egli definì la probabilità  $P$  che una molecola si trovi, all'istante iniziale, in  $U_0$  come

$$P := \frac{\int_{U_0} \varphi(x, y, z) dV}{\int_{\mathcal{B}} \varphi(x, y, z) dV} = \frac{J(U_0)}{J(\mathcal{B})}. \quad (9.15)$$

La funzione  $\varphi$ , su cui Poincaré fa solo l'ipotesi di continuità, è per il resto, *arbitraria* ma inizialmente Poincaré esamina dapprima il caso particolare in cui  $\varphi \equiv 1$ , cosicché  $P$  è la probabilità geometrica di Buffon e Poisson. Poincaré considera una regione  $\sigma_0 \subset U_0$  tale che, se una molecola è al suo interno all'istante  $t = 0$ , allora non potrà attraversare la frontiera  $U_0$  più di  $k$  volte. La probabilità  $P$  da determinare è allora

$$P = \frac{\sigma_0}{U_0},$$

dove  $\sigma_0$  ed  $U_0$  sono i volumi delle regioni omonime. Le immagini di  $\sigma_0$  agli istanti di tempo  $\tau, 2\tau, \dots, n\tau$  siano

$$\sigma_1, \quad \sigma_2 \cdots \quad \cdots \sigma_n$$

che, come prima, avranno tutte lo stesso volume di  $\sigma_0$ . Ora, per la definizione di  $\sigma_0$  non ci potranno essere più di  $k$  regioni, tra le  $(n+1)$  regioni  $\sigma_i$  che abbiano una porzione condivisa. In caso contrario, ripetendo il ragionamento precedente, si potrebbero trovare *più* di  $k$  regioni che hanno intersezione non vuota con  $\sigma_0$ , contrariamente all'ipotesi fatta. Dunque deve aversi

$$(n+1)\sigma_0 < kV, \quad (9.16)$$

essendo  $V$  il volume di  $\mathcal{B}$ . Quindi

$$P = \frac{\sigma_0}{U_0} < \frac{kV}{(n+1)U_0} :$$

per quanto piccolo si prenda il volume finito  $U_0$  e grande l'intero  $k$ , è sempre possibile far crescere  $n$  in modo che  $P$  scenda al di sotto di ogni soglia prefissata  $\varepsilon$ : al limite per  $n \rightarrow \infty$ ,  $P$  tende a 0. L'argomento può essere ripetuto per tutti i valori finiti di  $k$ . Ancora più importante è la variazione che presenta Poincaré, lasciando completamente arbitraria la funzione continua  $\varphi(x, y, z)$ . In quel caso si ha

$$P = \frac{J(\sigma_0)}{J(U_0)}$$

e, detto  $\mu$  il massimo assoluto di  $\varphi$  in  $\mathcal{B}$ , deve essere

$$J(\sigma_0) < \mu\sigma_0 :$$

per il resto la disuguaglianza (9.16) vale ancora e pertanto ora

$$P < \frac{\mu kV}{(n+1)J(U_0)}$$

cosicché

riassumendo, le molecole che non attraversano  $U_0$  se non un numero finito di volte sono eccezionali tanto quanto lo sono i numeri commensurabili che altro non sono se non un'eccezione nella serie dei numeri, mentre i numeri incommensurabili sono la regola.<sup>30</sup> ([14], p. 154)

Sottolineamo due punti nell'esposizione Poincaré: la probabilità va definita in base a certe *convenzioni* che possono mutare caso per caso. La funzione di densità  $\varphi(x, y, z)$  è scelta con grandi margini di discrezionalità, purché essa sia continua. Poincaré ha dunque ottenuto un risultato che vale *indipendentemente* dalla scelta delle “probabilità iniziali”, cioè dalla scelta di una particolare  $\varphi(x, y, z)$ , al contrario di quanto succedeva nel paradosso di Bertrand.

In uno scritto di natura filosofica, *La Science et l'Hypothèse* [15] pubblicato nel 1902, Poincaré dedicò al calcolo delle probabilità un intero capitolo all'interno della sezione rivolta ai metodi di indagine della fisica perché

il fisico si trova nelle stesse condizioni di un giocatore che soppesi le proprie possibilità. Ogni volta che ragiona per induzione, [il fisico] fa uso, in modo più o meno consapevole, del calcolo delle probabilità.<sup>31</sup>

Provocatoriamente, Poincaré si chiese quale grado di fiducia meritasse questo calcolo. Già nelle lezioni tenute alla Sorbona nel 1892-93 sul calcolo delle probabilità e pubblicate nel 1896, egli aveva esordito in un modo che sembrava lasciare poche speranze:

Non si può quasi dare una definizione soddisfacente della *Probabilità*.<sup>32</sup>  
([13], p. 1)

Eppure l'indagine scientifica non può fare a meno della probabilità: condannare il calcolo delle probabilità comporterebbe anche la condanna della scienza. Anche Poincaré distinse due tipi di probabilità, quella *soggettiva* e quella *oggettiva*:

un giocatore vuole scommettere e mi chiede consiglio; se accetto di darglielo, mi ispirerò al calcolo delle probabilità, ma non gli garantirò il successo. Ecco quella che chiamerò la *probabilità soggettiva*. In questo caso, ci si potrà accontentare della spiegazione che ho appena accennato. Ma supponiamo che un osservatore assista al gioco, che prenda nota di tutti i risultati e che il gioco si protragga a lungo; quando egli farà il bilancio dei dati raccolti, constaterà che gli eventi si sono distribuiti in modo conforme

<sup>30</sup>En résumé, les molécules qui ne traversent  $U_0$  qu'un nombre fini de fois sont exceptionnelles au même titre que les nombres commensurables qui ne sont qu'une exception dans la série des nombres, pendant que les nombres incommensurables sont la règle.

<sup>31</sup>Le physicien se trouve dans la même position que le joueur qui suppose ses chances. Toutes les fois qu'il raisonne par induction, il fait plus ou moins consciemment usage du calcul de probabilité.

<sup>32</sup>L'on ne peut guère donner une définition satisfaisante de la *Probabilité*.

alle leggi del calcolo delle probabilità. Ecco ciò che chiamerò la *probabilità oggettiva* ed è questo fenomeno che si dovrebbe spiegare.<sup>33</sup> ([15], p. 214)

La probabilità soggettiva è un rimedio alla nostra ignoranza, da usare nelle situazioni in cui occorre agire e non si ha tempo o possibilità di procurarsi tutte le informazioni necessarie per decidere con sicurezza. È un insieme di regole puramente indicative cui non si deve far troppo affidamento. Solo con la probabilità oggettiva, basata sull'osservazione delle frequenze e sulla legge dei grandi numeri, è possibile costruire qualcosa di positivo. Poincaré sottolineò il ruolo centrale svolto nelle indagini scientifiche dalla probabilità inversa. Per esempio, quando si voglia trovare una legge sperimentale basandosi su un numero finito di osservazioni, i cui risultati sono rappresentati come punti in un piano,

Dopo aver ottenuto questi vari punti, vi faccio passare una curva sforzandomi di scostarmi da essa il meno possibile e, tuttavia, di conservare alla mia curva una forma regolare, priva di punti angolosi, senza inflessioni troppo accentuate, senza brusche variazioni del raggio di curvatura. Questa curva rappresenterà la legge probabile ed io suppongo non solo che essa mi farà conoscere i valori intermedi della funzione tra quelli osservati ma che mi farà conoscere gli stessi punti osservati in modo più preciso dell'osservazione diretta. (...)

Ecco un problema di probabilità delle cause. Gli effetti sono le misure registrate; essi dipendono dalla combinazione di due cause: la vera legge che regola i fenomeni e gli errori di osservazione. Si tratta, noti gli effetti, di cercare la probabilità affinché il fenomeno obbedisca a quella legge e le osservazioni siano affette da quell'errore. La legge probabile corrisponde allora alla curva tracciata e l'errore più probabile di una osservazione è rappresentato dalla distanza del punto corrispondente a questa curva.

Il problema non avrebbe però alcun senso se, prima di tutte le osservazioni, non mi facessi un'idea *a priori* della probabilità che compete a questa piuttosto che a quella legge e delle possibilità di errore cui sono esposto.<sup>34</sup> ([15], pp. 230-231)

<sup>33</sup>un joueur veut tenter un coup; il me demande conseil. Si je lui le donne, je m'inspirerai du calcul des probabilités, mais je ne lui garantirai pas les succès. C'est là ce que j'appellerai la *probabilité subjective*. Dans ce cas, on pourrait se contenter de l'explication que je viens d'esquisser. Mais je suppose qu'un observateur assiste au jeu, qu'il en note tous les coups et que le jeu se prolonge longtemps; quand il fera le relevé de son carnet, il constatera que les événements se sont répartis conformément aux lois du calcul des probabilités. C'est là que j'appellerai la *probabilité objective*, et c'est ce phénomène qu'il faudrait expliquer.

<sup>34</sup>Quand j'ai obtenu ces différents points, je fais passer une courbe entre ces points en m'efforçant de m'en écarter le moins possible et, cependant, de conserver à ma courbe une forme régulière, sans point anguleux, sans inflexions trop accentuées, sans variation brusque du rayon de courbure. Cette courbe représentera la loi probable, et j'admets, non seulement qu'elle me fait connaître les valeurs de la fonction intermédiaires entre celles qui ont été observées, mais encore qu'elle me fait connaître les valeurs observées elles-mêmes plus exactement que l'observation directe. (...)

C'est là un problème de probabilité des causes. Les effets, ce sont les mesures que j'ai enregistrées; ils dépendent de la combinaison de deux causes: la loi véritable du phénomène et les erreurs d'observations. Il s'agit, connaissant les effets, de chercher la probabilité pour que le phénomène obéisse à telle loi, et pour que les observations aient été affectées de telle erreur. La loi plus probable correspond alors à la courbe tracée, et l'erreur la plus probable

La scelta di tracciare una curva regolare come curva più plausibile per rappresentare la legge sperimentale in discussione è dovuta al fatto che si consideri a priori più probabile una legge di tipo continuo:

Senza questa fiducia, il problema di cui parliamo non avrebbe alcun senso; l'interpolazione sarebbe impossibile; non si potrebbe dedurre una legge a partire da un numero finito di osservazioni; la scienza non esisterebbe.<sup>35</sup>  
([15]. p. 231)

Si percepisce in modo palpabile dalle parole di Poincaré il bisogno che la scienza ha di un calcolo delle probabilità rigoroso, attendibile, libero dalle incertezze sui fondamenti e dai paradossi che ne avevano punteggiato fin a quel momento lo sviluppo.

---

d'une observation est représentée par la distance du point correspondant à cette courbe.

Mais le problème n'aurait aucun sens si, avant toute observation, je ne me faisais une idée *a priori* de la probabilité de telle ou telle loi, et des chances d'erreur auxquelles je suis exposé.

<sup>35</sup>Sans cette croyance, le problème dont nous parlons n'aurait aucun sens; l'interpolation serait impossible; on ne pourrait déduire une loi d'un nombre fini d'observations; la science n'existerait pas.

# Bibliografia

- [1] J. Bertrand: *Calcul des probabilités*. Gauthier-Villars, Paris, (1889).
- [2] I.-J. Bienaymé: Sur la probabilité des erreurs d'après la méthode des moindres carrés. *Journal de mathématiques pures et appliquées*, **17** (S. I), 33-78, (1852).
- [3] G.-L. Leclerc de Buffon. *Essai d'arithmétique morale*. In *Œuvres Complètes de Buffon*, Vol. XV, Didot, Paris, (1829), pp. 338-447.
- [4] P.L. Čebišev: Démonstration élémentaire d'une proposition générale de la théorie des probabilités. *Journal für die reine und angewandte Mathematik*, **33**, 269-267, (1846).
- [5] P.L. Čebišev: Des valeurs moyennes. *Journal de mathématiques pures et appliquées*, **12** (S. II), 177-184, (1867).
- [6] A.A. Cournot: Mémoire sur les applications du calcul des chances à la Statistique judiciaire. *Journal de mathématiques pures et appliquées*, **3** (S. I), 257-334, (1838).
- [7] A.A. Cournot: *Exposition de la théorie des chances et de probabilités*. Hachette, Paris, (1843).
- [8] D.A. Gillies: *Philosophical Theories of Probability*. Routledge, London and New York, (2000).
- [9] J.M. Keynes: *A treatise on probability*. Mac Millan & co., London, (1921).
- [10] J. von Kries: *Die Principien der Wahrscheinlichkeits-Rechnung. Eine logische Untersuchung*. Mohr, Freiburg i. B., (1886).
- [11] L.E. Maistrov: *Probability Theory. A Historical Sketch*. Translated and Edited by S. Kotz. Academic Press, New York, (1974).
- [12] H. Poincaré: Le mécanisme et l'expérience. *Revue de Métaphysique et de Morale*, **1**, 534-537, (1893).
- [13] H. Poincaré: *Calcul des probabilités*. Gauthier-Villars, Paris, (1896).

- [14] H. Poincaré: *Les Méthodes Nouvelles de la Mécanique Céleste*. Tome III. Gauthier-Villars, Paris, (1899).
- [15] H. Poincaré: *La Science et l'Hypothèse*. Flammarion, Paris, (1902).
- [16] S.-D. Poisson: *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris, Bachelier, (1837).
- [17] O.B. Sheynin: J.H. Lambert's Work on Probability. *Archive for History of Exact Sciences*, **7**, 244-256, (1971).
- [18] O.B. Sheynin: R.J. Boscovich's Work on Probability. *Archive for History of Exact Sciences*, **9**, 306-324, (1973).
- [19] O.B. Sheynin: S.D. Poisson's Work on Probability. *Archive for History of Exact Sciences*, **18**, 245-300, (1978).
- [20] O.B. Sheynin: C.F. Gauss and the Theory of Errors. *Archive for History of Exact Sciences*, **20**, 21-72, (1979).
- [21] O.B. Sheynin: On the History of the Statistical Method in Biology. *Archive for History of Exact Sciences*, **22**, 323-371, (1980).
- [22] O.B. Sheynin: On V. Ya. Buniakovsky's Work on Probability. *Archive for History of Exact Sciences*, **43**, 199-223, (1991).
- [23] O. Sheynin: Chebyshev's Lectures on the Theory of Probability. *Archive for History of Exact Sciences*, **46**, 321-340, (1994).
- [24] O. Sheynin: Bertrand's Work on Probability. *Archive for History of Exact Sciences*, **48**, 155-199, (1994).
- [25] S.M. Stigler: *History of Statistics. The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, Massachusetts, U.S.A., (1986).

## Capitolo 10

# La critica dei fondamenti della probabilità

### 10.1 Primi tentativi di assiomatizzazione

Benché il matematico e logico inglese George Boole (1815-1864) avesse già auspicato un trattamento assiomatico della probabilità fin dal 1854, né lui né altri vi posero mano fino all'inizio del XX secolo.

Sul finire del XIX secolo avvertiamo un cambiamento nell'ambiente, leggendo per esempio le parole che il matematico russo Nikolaj Bugaev (1837-1903), esponente della scuola di Mosca, pronunciò in occasione del I Congresso Internazionale dei Matematici tenutosi a Zurigo nel 1897 ([14], pp. 112-113)

Una certa dose di casualità che compare nelle nostre azioni introduce un elemento di contingenza nella natura stessa. La contingenza entra in scena dunque come una proprietà essenziale di certi fenomeni della realtà. Nella realtà non vi è solo il regno della certezza. C'è anche l'impero della probabilità.

La dottrina dei fenomeni aleatori o teoria delle probabilità appare come una scienza matematica essenziale nel quadro generale delle nostre conoscenze. La filosofia deve fare i conti sia con la probabilità che con la certezza.

Laddove non è possibile far ricorso all'analisi o all'aritmetologia, dove cioè si ignora la legge dei fenomeni, ecco che la teoria delle probabilità deve fornire delle risposte.

In generale la probabilità si manifesta nell'ambito di eventi molto complessi. Senza dubbio, bisogna ricondurvi molti fenomeni sociali. La teoria delle probabilità si può applicare a molti fenomeni sociali. La legge dei grandi numeri mostra che l'influenza delle circostanze fortuite, che distruggono l'andamento regolare dei fenomeni, può essere indebolita con un grande numero di osservazioni.

Fondandoci su questa legge, le nostre conclusioni riferite ai fenomeni aleatori possono acquisire una certa autorità.<sup>1</sup>

Di lì a poco, nel 1900, sarebbe comparso il primo lavoro contenente una presentazione assiomatica della probabilità, in un articolo sulla matematica attuariale—la matematica delle assicurazioni—redatto per la *Encyklopädie der mathematischen Wissenschaften* curata da Felix Klein (1849-1925) [37]. L'articolo fu scritto da Georg Bohlmann (1869-1928) che aveva ottenuto nel 1894 una abilitazione all'insegnamento universitario a Göttingen, dove si trovava Klein. Fu Klein a fondare a Göttingen, nel 1895, un *Seminar für Versicherungswissenschaft*, cioè un seminario per le scienze attuariali, diretto dallo statistico Wilhelm Lexis (1837-1914). Nel giro di pochi anni Bohlmann ottenne reputazione di esperto nel campo della matematica attuariale e nel 1903 accettò un impiego presso la filiale di Berlino della Mutual Life Insurance Company, mantenendo comunque contatti con gli esponenti di spicco della statistica matematica europea. Mentre scriveva l'articolo per la *Encyklopädie*, David Hilbert (1862-1943), che era appena giunto a Göttingen da Königsberg, tenne un ciclo di lezioni dedicato ai fondamenti della geometria dove veniva sottolineato il ruolo delle relazioni che legano gli oggetti uno all'altro, un approccio che segnava una cesura con l'impostazione classica per cui gli assiomi erano affermazioni evidenti da cui trarre altre conseguenze. L'8 agosto 1900, in occasione del II Congresso Internazionale dei Matematici tenutosi a Parigi, Hilbert propose una lista di 23 problemi ritenuti fondamentali per lo sviluppo della matematica del XX secolo. Il VI problema—*Mathematische Behandlung der Axiome der Physik*—riguardava l'assiomatizzazione di quelle discipline *fisiche* (*physikalischen Disziplinen*) in cui

già fin d'ora la matematica gioca un ruolo preponderante: in prima linea vi sono il calcolo delle probabilità e la meccanica.<sup>2</sup> ([32], p. 306)

Tra i problemi da affrontare, Hilbert sottolineava quello di fornire un approccio rigoroso e soddisfacente ai metodi di calcolo dei valori medi utilizzati soprattutto nella teoria cinetica dei gas, tema molto caro ad Hilbert, su cui tenne anche dei corsi. L'assiomatizzazione di Bohlmann si esaurisce in tre assiomi e tre definizioni, anche perché il punto focale del lavoro di Bohlmann era di fondare

<sup>1</sup>Une certaine part de hasard, qui apparaît dans nos actions, introduit un élément d'éventualité dans la nature même. L'éventualité entre ainsi en scène, comme une propriété essentielle de certains phénomènes du monde. Dans le monde il n'y a pas que le règne de la certitude seule. La probabilité y a aussi son empire. La doctrine des phénomènes fortuits ou théorie des probabilités apparaît comme une science mathématique essentielle dans le système général de nos connaissances. Le philosophe doit compter avec la probabilité autant qu'avec la certitude. La théorie des probabilités doit donner des réponses là, où l'on ne peut recourir à l'analyse et à l'arithmologie, là, où l'on ignore la loi des phénomènes. En général la probabilité se manifeste dans la sphère des événements très compliqués. On doit y rattacher sans conteste plusieurs phénomènes sociaux. La théorie des probabilités peut s'appliquer à plusieurs phénomènes sociaux. La loi des grands nombres démontre que l'influence des causes fortuites, qui détruisent la marche régulière des phénomènes, peut être affaiblie par un grand nombre d'observations. En se fondant sur cette loi, nos conclusions par rapport aux phénomènes fortuits peuvent avoir une certaine autorité.

<sup>2</sup>schon heute die Mathematik eine hervorragende Rolle spielt: dies sind in erster Linie die Wahrscheinlichkeitsrechnung und die Mechanik.

sul calcolo delle probabilità la matematica attuariale, impostazione che ancora veniva messa in dubbio in quei tempi. ([6], p. 859)

*Definizione I.* Pertanto, la probabilità che un evento  $E$  abbia luogo, è una frazione propria positiva  $p$  che è associata ad  $E$ .

*Assioma I.* Se  $E$  è certo, allora  $p = 1$ . Se  $E$  è impossibile, allora  $p = 0$ .

*Definizione II.* Due eventi si escludono mutuamente quando la probabilità che sia  $E_1$  che  $E_2$  si verifichino è uguale a 0.

*Assioma II.* Sia  $p_1$  la probabilità che  $E_1$  abbia luogo,  $p_2$  quella che abbia luogo  $E_2$  e  $p$  la probabilità che abbia luogo uno dei due eventi  $E_1$  o  $E_2$ . Allora:

$$p = p_1 + p_2.$$

*Assioma III.* Sia  $p_1$  la probabilità che  $E_1$  abbia luogo,  $p'_2$  quella che abbia luogo  $E_2$  quando si sa che  $E_1$  è avvenuto,  $p$  la probabilità che sia  $E_1$  che  $E_2$  avvengano. Allora

$$p = p_1 p'_2.$$

*Definizione III.* Usando gli stessi simboli dell'Assioma III, sia  $p_2$  la probabilità che  $E_2$  abbia luogo. Si dice che  $E_1$  ed  $E_2$  sono indipendenti mutualmente quando<sup>3</sup>

$$p = p_1 p_2.$$

Si tratta di una formulazione molto scarna, in cui *non* viene definita la nozione di evento e la definizione di indipendenza viene limitata al caso di *due* eventi. Su quest'ultimo punto Bohlmann ritornò in un lavoro [7] presentato al IV Congresso Internazionale dei Matematici, tenutosi a Roma nel 1908, nel quale diede questa formulazione dell'indipendenza: siano dati  $m$  eventi  $E_1, E_2, \dots, E_m$  e sia  $E'_i$  il complementare di  $E_i$ ,  $p_i$  la probabilità che  $E_i$  si realizzi e  $q_i = 1 - p_i$ . Sia inoltre  $a_i$  una variabile assume i soli valori 1 o 0, a seconda che  $E_i$  si realizzi o meno. allora, indicata con  $\mathbf{a}$  la  $m$ -upla ordinata  $\mathbf{a} := (a_1, a_2, \dots, a_m)$ , gli eventi

<sup>3</sup>*Definition I.* Die Wahrscheinlichkeit dafür, dass ein Ereignis  $E$  eintritt, ist ein positiver echter Bruch  $p$ , der  $E$  zugeordnet ist.

*Axiom I.* Ist  $E$  gewiss, so ist  $p = 1$ . Ist  $E$  unmöglich, so ist  $p = 0$ .

*Definition II.* Zwei Ereignisse schliessen sich aus, wenn die Wahrscheinlichkeit dafür, dass sowohl  $E_1$  als  $E_2$  eintritt, gleich 0 ist.

*Axiom II.* Sei  $p_1$  die Wahrscheinlichkeit, dass  $E_1$  eintritt,  $p_2$  die, dass  $E_2$ ,  $p$  die, dass eines der beiden Ereignisse  $E_1$  oder  $E_2$  eintritt; alsdann ist:

$$p = p_1 + p_2.$$

*Axiom III.* Sei  $p_1$  die Wahrscheinlichkeit, dass  $E_1$  eintritt,  $p'_2$  die, dass  $E_2$  eintritt, wenn man weiss, dass  $E_1$  eingetreten ist,  $p$  die Wahrscheinlichkeit, dass sowohl  $E_1$  als  $E_2$  eintritt. Als dann ist:

$$p = p_1 p'_2.$$

*Definition III.* Seit unter sonst gleichen Bezeichnungen wie in Axiom III  $p_2$  die Wahrscheinlichkeit, dass  $E_2$  eintritt. Man sagt, dass  $E_1$  und  $E_2$  von einander unabhängig sind, wenn

$$p = p_1 p_2$$

ist.

$E_1, E_2, \dots, E_m$  sono indipendenti se

$$p(\mathbf{a}) = \prod_{i=1}^m p_i^{a_i} q_i^{1-a_i}.$$

Fu in questa sede che Bohlmann fornì per primo un esempio che dimostrava come l'indipendenza a coppie non fosse lo stesso dell'indipendenza in generale. L'esempio di Bohlmann è molto più complicato di quello del tetraedro discusso in precedenza ed attribuito a Sergej Bernštein.

A piccoli passi, la formulazione assiomatica progredì attraverso due tesi di dottorato: la prima, discussa a Zurigo da Rudolph Lämmel (1879-1962) nel 1904; la seconda, sotto la guida di Hilbert, fu pubblicata nel 1907 a Göttingen da un italiano, Ugo Broggi (1880-1965). In entrambe le tesi si utilizzano concetti elementari tratti dalla teoria degli insiemi e della misura. A margine, ricordiamo che nel 1904, Henri Lebesgue (1875-1941) aveva esposto una teoria dell'integrazione più generale [38] di quella di Riemann.

Per Lämmel ([40], pp. 359-366), l'introduzione della probabilità deve essere preceduta da un processo logico articolato in tre ipotesi:

1. si individua un insieme  $M \equiv \{A_1, A_2, \dots, A_n\}$  pertinente al problema in esame;
2. si determinano i sottoinsiemi di  $M$  che contengono tutti gli elementi  $A_i$  che verificano una condizione  $x$ ;
3. si associa agli elementi  $A_i$  un peso elementare o valenza.

L'insieme  $M$  rappresenta i casi possibili (*möglichen*) mentre il sottoinsieme (*Teilmenge*)  $T = \{A_1, A_2, \dots, A_\nu\}$  rappresenta i casi favorevoli. Infine la funzione di peso  $G_\alpha = f(\alpha) \geq 0$  è un numero associato al generico elemento  $A_\alpha$ . Se  $T$  è costituito dall'unico elemento  $A_x$ , la probabilità elementare associatagli è

$$w = \frac{f(x)}{\sum_{\alpha=1}^n f(\alpha)}$$

mentre la probabilità associata a  $T$  è

$$W = \sum_{i=1}^{\nu} \frac{f(x_i)}{\sum_{\alpha=1}^n f(\alpha)}.$$

Lämmel affrontò anche il passaggio ad insiemi infiniti considerando due tipi di problemi: quelli per i quali sia  $M$  che  $T$  sono numerabili e quelli per i quali essi hanno la potenza del continuo. In quest'ultimo caso la definizione di probabilità coincide con quella della probabilità geometrica. I due assiomi posti alla base del *sistema minimale* (*Minimalsystem*) della teoria delle probabilità, cioè del sistema ridotto al minimo numero possibile di definizioni ed assiomi a partire dai quali tutti i risultati della teoria possono essere dedotti, sono il teorema delle

probabilità totali per eventi incompatibili e quello della probabilità composta, per eventi indipendenti:

$$W\left(\sum_{\alpha=1}^n x_{\alpha}\right) = \sum_{\alpha=1}^n W(x_{\alpha}) \quad \text{e} \quad W\left(\prod_{\alpha=1}^n x_{\alpha}\right) = \prod_{\alpha=1}^n W(x_{\alpha}),$$

rispettivamente. Lämmel annoverò tra gli assiomi, nel caso infinito, anche la richiesta di additività numerabile ([46], p. 33)

Passando a Broggi, anche lui come Bohlmann, era particolarmente interessato alla matematica attuariale, su cui scrisse un manuale, pubblicato nel 1906 [12]. Come e forse più di Lämmel, Broggi risentì nella sua formulazione del fermento culturale che portò ad un rapido incremento degli strumenti in possesso dell'analisi matematica: il linguaggio della teoria degli insiemi e l'utilizzo del concetto di misura, nel senso di Borel e Lebesgue. Broggi dichiarò apertamente di prendere a modello i *Grundlagen der Geometrie* di Hilbert, in cui si prestava attenzione alla non-contraddittorietà (*Widerspruchslosigkeit*), completezza (*Vollständigkeit*) ed indipendenza (*Unabhängigkeit*) del sistema di assiomi proposto. Gli eventi sono definiti in termini di una coppia di insiemi. Broggi è debitore ad un assistente di Hilbert, Ernst Zermelo (1871-1953), quando afferma che

È possibile decidere in modo oggettivo se un elemento arbitrario  $x$  di un insieme  $M$  ( $x \in M$ ) possieda una determinata proprietà o meno.<sup>4</sup> ([13], p. 6)

A dispetto di ciò, il calcolo delle probabilità ha un senso perché

È possibile caratterizzare un qualunque  $x \in M$  in modo tale che la definizione data in precedenza, non consenta di affermare se  $x$  goda o meno della proprietà  $A$ .

Si tratta di una richiesta di natura soggettiva, la cui verifica dipende dallo stato della nostra conoscenza.<sup>5</sup> ([13], p. 6)

È pertanto ancora la nostra ignoranza a rendere possibile un discorso su ciò che è probabile. Broggi considerò un insieme di casi possibili  $M$  ed un sottoinsieme  $M_1 \subseteq M$  di casi favorevoli: è questa coppia, indicata con  $[M_1, M]$  a definire un *evento*. La probabilità di un evento  $E$  è allora una grandezza  $p(E) \geq 0$  (*eine nicht-negative Grösse*) tale che

1. Se un evento  $E$  è certo, la sua probabilità è 1.
2. Se due eventi  $E_1$  ed  $E_2$  sono incompatibili, allora, detto  $E = E_1 \cup E_2$ ,  $p(E) = p(E_1) + p(E_2)$ .

<sup>4</sup>Es ist möglich, objektiv zu entscheiden, ob einem beliebigen Elemente  $x$  einer Menge  $M$  ( $x \in M$ ) eine definite Eigenschaft  $A$  zukommt oder nicht.

<sup>5</sup>Es ist möglich, ein beliebiges  $x \in M$  sowohl charakterisieren (...) dass die gegebene Definition darüber nichts aussagt, ob dem betrachteten  $x$  die Eigenschaft  $A$  zukommt oder nicht.

Es handelt sich um eine Forderung subjektiver Natur, deren Erfülltsein vom Stande unserer Kenntnisse abhängig ist."

Da questi due assiomi Broggi dedusse subito che la probabilità di un evento impossibile è  $p(E) = 0$  e che, se  $E \supset E_1$  o  $E \supset E_2$ ,  $p(E) \geq p(E_1)$  e  $p(E) \geq p(E_2)$ , per poi porsi il problema dell'*esistenza* di una tale funzione. È qui che entra in scena la definizione di misura di un insieme  $M$ , inteso come sottoinsieme del segmento  $[0, 1]$  della retta reale. Broggi richiamò allora la definizione di misura di un insieme  $m(M) \geq 0$  data da Lebesgue, come numero che gode di queste proprietà:

1. Ad insiemi uguali corrispondono uguali misure;
2. la misura di un insieme che risulti essere somma<sup>6</sup> di un numero finito o numerabile di insiemi disgiunti è la somma della misura di questi ultimi;
3. la misura del continuo  $[0, 1]$  è 1.

Se è dato un evento  $E = [M, M_1]$ , allora la quantità

$$p(E) := \frac{m(M_1)}{m(M)}$$

soddisfa gli assiomi introdotti in precedenza. Poiché, quando  $M$  ed  $M_1$  sono formati da un numero finito o numerabile di punti,  $P(E)$  è della forma indeterminata  $\frac{0}{0}$ , la definizione di probabilità diventa, se i due insiemi sono finiti,

$$p(E) = \frac{m_1}{m}$$

dove ora  $m_1$  ed  $m$  sono il numero di punti che costituiscono l'insieme  $M_1$  dei casi favorevoli ed  $M$  dei casi possibili, rispettivamente. Quando questi insiemi sono numerabili, la definizione di probabilità viene ottenuta tramite un passaggio al limite. A testimonianza dello stato embrionale dell'approccio assiomatico alla probabilità, notiamo come Broggi ritenesse di poter dimostrare, con un passaggio al limite, che dall'additività finita si potesse dimostrare quella *numerabile*. Sarà il matematico polacco Stefan Banach (1892-1945) a mostrare, nel 1923, come sia possibile ottenere una funzione definita sui sottoinsiemi di un insieme assegnato, per la quale vale solo l'additività finita e non quella numerabile [1]: il lavoro di Banach è di teoria della misura, l'osservazione a proposito dell'errore di Broggi si trova in un lavoro di un altro matematico polacco, Hugo Steinhaus (1887-1972) [42]. Nella seconda ed ultima parte della sua tesi, Broggi distinse tra la probabilità matematica di un evento  $E$ , la probabilità che deve obbedire agli assiomi esposti in precedenza, dalla probabilità *statistica* di  $E$  definita come il limite, se esiste,

$$\lim_{s \rightarrow \infty} \frac{z}{s}$$

tra il numero  $z$  di prove (*Versuche*) in cui si è prodotto  $E$  ed il numero totale  $s$  di prove. Alla luce di questa doppia definizione, Broggi divide il classico teorema di Jakob Bernoulli in due parti:

<sup>6</sup>Per somma di insiemi si deve intendere la loro unione.

- Se esiste la probabilità matematica di un evento, esiste anche la sua probabilità statistica;
- le due probabilità coincidono.

Broggi si chiede se sia possibile invertire la prima parte del teorema di Bernoulli, che permetterebbe di dedurre, dall'esistenza di una probabilità statistica, anche quella di una probabilità matematica di un evento. Egli *postula* l'esistenza di una probabilità statistica e da ciò mostra come questa funzione debba accordarsi con la definizione di probabilità matematica. Riemerge così in questo punto l'interesse per le applicazioni del calcolo delle probabilità ai fenomeni collettivi, basate sull'esistenza di rapporti pressoché costanti in osservazioni ripetute molte volte.

Broggi non fu il primo ad utilizzare la teoria della misura come strumento del calcolo delle probabilità. Le prime esplorazioni in tal senso risalgono a due lavori del matematico svedese Anders Wiman (1865-1959), pubblicati tra il 1900 ed il 1901 su problemi aritmetici legati a sviluppi in frazioni continue che figuravano in questioni di meccanica celeste ([46], p. 31).

## 10.2 Borel e la probabilità

Una figura chiave nello sviluppo della probabilità moderna fu Émile Borel (1871-1956), che succedette a Poincaré sulla cattedra alla Sorbona. La sua azione si dispiegò a vari livelli; dal punto di vista organizzativo, con l'aiuto del matematico americano George Birkhoff (1884-1944) egli fondò nel 1928 a Parigi l'Institut Henri Poincaré che divenne un centro di diffusione delle nuove idee sul calcolo della probabilità; dal punto di vista dell'alta divulgazione, egli sovrintese alla stesura del *Traité du calcul des probabilités et ses applications*, un progetto editoriale di ampio respiro che si snodò dal 1921 al 1939 e che coagulò nella stesura di quattro tomi, articolati in più fascicoli che coprono sia aspetti teorici che applicazioni [14]; dal punto di vista della ricerca, egli contribuì direttamente o indirettamente alla crescita del calcolo delle probabilità come solida disciplina matematica approntando nuovi metodi e risultati.

Abbiamo sentito Poincaré sottolineare come l'estensione della probabilità al caso continuo richieda di fissare una convenzione preliminare. Questa idea venne ripresa vigorosamente da Borel che apre il suo primo lavoro [8] dedicato alla probabilità, pubblicato nel 1905, con queste parole:

È noto che le questioni di probabilità dove intervengono variabili continue non possono aver senso che in virtù di precise *convenzioni*.<sup>7</sup> ([8], p. 123)

Questa è sostanzialmente la lezione che si apprende dai paradossi di Bertrand e la convenzione più immediata è quella di considerare la probabilità di un evento come proporzionale alla sua estensione: lunghezza, area o volume, a seconda

<sup>7</sup>On sait que les questions de probabilité où interviennent des variables continues ne peuvent acquérir de sens qu'en vertu de *conventions* précises.

della natura dell'insieme degli eventi. Occorre però precisare lo strumento tecnico per quantificare l'estensione. L'integrale di Riemann non basta perché la probabilità che un numero scelto a caso nell'insieme  $[0, 1]$  sia razionale è, intuitivamente, nulla ma, utilizzando una rappresentazione integrale, essa sarebbe rappresentata da

$$\int_0^1 \chi_{\mathbb{Q}}(x) dx$$

dove  $\chi_{\mathbb{Q}}$  è la funzione caratteristica dei razionali: come è noto, questo integrale non ha senso, come integrale di Riemann. Poco prima però, nel 1904, Henri Lebesgue (1875-1941) aveva esposto una teoria dell'integrazione più generale [38] e Borel fu tra i primi a sottolinearne l'importanza:

I metodi del sig. Lebesgue permettono dunque di studiare questioni di calcolo delle probabilità che sembravano inaccessibili ai processi classici di integrazione.<sup>8</sup> ([8], p. 126)

Sorprendentemente, Borel *non* applicò i metodi di Lebesgue in quello che si può considerare uno dei suoi lavori più influenti nel calcolo delle probabilità, pubblicato nel 1909 [10]. Qui Borel introdusse le *probabilità numerabili* (*probabilités dénombrables*) come ponte tra le probabilità discontinue in cui gli esiti possibili in una prova ed il numero di prove sono *entrambi* finiti e quelle continue o geometriche dove i casi possibili sono infiniti ed hanno la potenza del continuo. Le probabilità numerabili corrispondono al caso in cui il numero di eventi possibili ed eventualmente anche il numero delle prove considerate sono infiniti ma *numerabili*. Per Borel non si tratta di un perfezionamento tecnico ma di una estensione dal profondo valore concettuale ([10], pp. 247-248):

la potenza degli insiemi numerabili è la sola a noi nota in un modo positivo, la sola che interviene effettivamente nei nostri ragionamenti. È chiaro infatti che l'insieme degli elementi analitici suscettibili ad essere veramente definiti e presi in considerazione non può essere altro che un insieme numerabile; credo che questo punto di vista si imporrà via via sempre più tra i matematici e che il continuo non sarà stato altro che uno strumento transitorio, la cui utilità al presente non è affatto trascurabile (...), ma che dovrà esser visto come un mezzo per studiare gli insiemi numerabili che costituiscono la sola realtà che possiamo raggiungere.<sup>9</sup>

Per il caso in cui il numero di esiti che una prova può dare è finito, uguale a due per semplicità, mentre le prove sono infinite e numerabili, Borel ([10], p. 248)

<sup>8</sup>Les méthodes de M. Lebesgue permettent donc d'étudier des questions de probabilités qui paraissent inaccessibles par les procédés d'intégration classiques.

<sup>9</sup>la puissance des ensembles dénombrables étant la seule qui nous soit connue d'une manière positive, la seule qui intervienne effectivement dans nos raisonnements. Il est clair, en effet, que l'ensemble des éléments analytiques susceptibles d'être réellement définis et considérés ne peut être qu'un ensemble dénombrable; je crois que ce point de vue s'imposera chaque jour davantage aux mathématiciens et que le continu n'aura été qu'un instrument transitoire, dont l'utilité actuelle n'est pas négligeable (...), mais qui devra être regardé seulement comme un moyen d'étudier les ensembles dénombrables, lesquels constituent la seule réalité que nous puissions atteindre.

considerò il problema di trovare la probabilità che un evento (caso favorevole), avente probabilità  $p_n \neq 1$  di realizzarsi alla  $n$ -esima prova, *non* si verifichi in alcuna prova. Supponendo le prove *indipendenti*, nel caso di un numero finito di prove, la probabilità cercata è

$$(1 - p_1)(1 - p_2) \cdots (1 - p_n).$$

Il punto è di stabilire la liceità di un passaggio al limite per  $n \rightarrow \infty$  per assumere

$$A_0 := \prod_{k=1}^{\infty} (1 - p_k) \quad (10.1)$$

come risposta al problema. Borel, ricorrendo alla teoria dei prodotti infiniti distinse due casi, a seconda che la serie a termini positivi

$$\sum_{k=1}^{\infty} p_k \quad (10.2)$$

fosse convergente o divergente. Infatti, il prodotto infinito  $A_0$  converge verso un numero diverso da zero e dall'unità se e solo se la serie (10.2) è convergente. In questa circostanza, il valore di  $A_0$  è ben definito ed è diverso da 0 e da 1. Quando la serie (10.2) diverge,  $A_0$  *diverge* a zero. Borel sottolineò una importante differenza tra le probabilità discontinue (*discontinues*) e le probabilità numerabili o continue: un evento che abbia probabilità 1 *non* si verificherà necessariamente, così come un evento con probabilità 0 *non* è impossibile. Con questa precisazione, egli riteneva accettabile associare il valore  $A_0 = 0$  come soluzione al problema nel caso della divergenza della serie (10.2). Si tratta in effetti di un punto molto delicato nello sviluppo della moderna teoria matematica della probabilità in quanto tocca direttamente il legame che i suoi risultati hanno con il mondo reale. Il problema era stato affrontato già da Cournot nel suo volume sul calcolo delle probabilità del 1843, dove si era espresso in questi termini:

*L'evento fisicamente impossibile è dunque quello la cui probabilità matematica è infinitamente piccola; è solo questa osservazione a fornire consistenza, valore oggettivo e fenomenico alla teoria della probabilità matematica.*<sup>10</sup>

Questo enunciato è noto come *principio di Cournot* e ad esso, in modo più o meno indiretto, si riferiranno matematici come Borel nel momento in cui occorre discutere la rilevanza di risultati matematici sulla probabilità per le applicazioni al mondo reale ([41], § 2.2).

Tornando all'analisi del lavoro di Borel [10], osserviamo, sulla scorta di [2], che Borel grazie all'indipendenza degli eventi  $B_i$  considerati, lavorò soltanto con

<sup>10</sup>*L'événement physiquement impossible est donc celui dont la probabilité mathématique est infiniment petite; et cette seule remarque donne une consistance, une valeur objective et phénoménale à la théorie de la probabilité mathématique*

il passaggio al limite di

$$p\left(\bigcap_{k=1}^n B_k\right) = \prod_{k=1}^n p(B_k)$$

e assunse come evidente che

$$\lim_{n \rightarrow \infty} p\left(\bigcap_{k=1}^n B_k\right) = p\left(\bigcap_{k=1}^{\infty} B_k\right)$$

che è una conseguenza della additività numerabile.

Procedendo, Borel si chiese quale fosse la probabilità  $A_1$  che l'evento si verifici esattamente in una delle infinite prove considerando

$$\omega_1 = p_1(1-p_2)(1-p_3)\cdots(1-p_n) = \frac{p_1}{1-p_1}(1-p_2)(1-p_3)\cdots(1-p_n)$$

che rappresenta la probabilità che l'evento considerato si presenti solo alla prima di  $n$  prove. In generale, la probabilità che l'evento si presenti alla prova  $k$ -esima è

$$\omega_k = (1-p_1)(1-p_2)\cdots(1-p_{k-1})p_k(1-p_{k+1})\cdots(1-p_n) = \frac{p_k}{1-p_k} \prod_{j=1}^n (1-p_j).$$

Quando (10.2) è convergente e sempre assumendo tacitamente la validità della additività numerabile, Borel ottenne

$$A_1 = \lim_{n \rightarrow \infty} \sum_{k=1}^n \omega_k = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{p_k}{1-p_k} \prod_{j=1}^n (1-p_j)$$

da cui ricavò

$$A_1 = A_0 \sum_{k=1}^{\infty} \frac{p_k}{1-p_k} = A_0 \sum_{k=1}^{\infty} u_k,$$

dove anche la sommatoria che moltiplica  $A_0$  ha valore finito, sempre in virtù della convergenza di (10.2). Nel caso in cui la (10.2) diverga, l'argomento che porta Borel a concludere che  $A_1 = 0$  manca di rigore, anche se la conclusione è corretta. Con un procedimento simile Borel considerò la probabilità  $A_2$  che l'evento si presenti solo due volte,

$$A_2 = A_0 \sum u_k u_h,$$

con  $k \neq h$  e, in generale, la probabilità che l'evento si presenti esattamente  $k$  volte,

$$A_k = A_0 \sum u_i u_j u_\ell \cdots,$$

dove la somma è fatta su tutti i valori distinti dei  $k$  indici. Questi calcoli preliminari permisero a Borel di affrontare il problema più importante:

Qual è la probabilità che il caso favorevole si verifichi un'infinità di volte?<sup>11</sup>  
([10], p. 230)

Detta

$$S = A_0 + A_1 + \cdots + A_k + \cdots = A_0(1 + u_1)(1 + u_2) \cdots (1 + u_k) \cdots$$

e osservando che

$$(1 + u_k) = 1 + \frac{p_k}{1 - p_k} = \frac{1}{1 - p_k},$$

dalla definizione (10.1) di  $A_0$  si ricava

$$S = A_0 \prod_{k=1}^{\infty} (1 + u_k) = A_0 \prod_{k=1}^{\infty} \frac{1}{1 - p_k} = A_0 \frac{1}{A_0} = 1$$

e quindi, invocando ancora tacitamente l'additività numerabile, la probabilità  $A_{\infty}$  oggetto del problema è uguale ad  $1 - S = 0$ . Al contrario, quando (10.2) diverge, Borel mostrò, anche qui senza il necessario rigore, che  $A_{\infty}$  tende ad 1 e quindi poté concludere che ([10], p. 252)

Nei casi in cui la serie (10.2) è convergente, le probabilità  $A_0, A_1, \dots$  hanno valori determinati non nulli e la probabilità  $A_{\infty}$  è nulla; nel caso della divergenza, le probabilità  $A_k$  sono tutte nulle ed  $A_{\infty}$  è uguale all'unità.<sup>12</sup>

Questo risultato è anche noto con il nome di lemma di Borel-Cantelli, dal nome del matematico italiano Francesco Paolo Cantelli (1875-1966) che si servì di questo risultato nella dimostrazione della legge forte dei grandi numeri [15].

A dispetto della sua produzione scientifica, centrata soprattutto sull'analisi matematica, Borel fu molto attento al dialogo con la fisica e, più in generale, con le applicazioni della matematica. Un esempio di questa attenzione è la sua critica alla prima deduzione di Maxwell della legge di distribuzione delle velocità molecolari, rivolta non tanto all'uso dell'ipotesi di indipendenza delle componenti della velocità di una molecola, ma alla mancanza di una stima sull'effetto che un piccolo errore sui dati iniziali del problema può avere sulla struttura dell'equazione funzionale da cui Maxwell dedusse la distribuzione delle velocità molecolari. Il caos molecolare che si viene a produrre non appena le molecole urtano elasticamente tra loro è tale che, dopo un intervallo di tempo molto piccolo, la distribuzione delle molecole all'interno della massa gassosa sarà tanto complicata da poter essere pensata come prodotta dal caso:

Siamo così ricondotti a studiare un problema puro della teoria delle probabilità. Rappresentiamo la velocità di ciascuna molecola grazie ad un vettore  $OM$  la cui origine è in un punto fisso  $O$  e che ha la stessa grandezza della velocità; occorre trovare la distribuzione più probabile dei punti

<sup>11</sup>Quelle est la probabilité pour que le cas favorable se produise une infinité de fois?

<sup>12</sup>Dans les cas où la série (10.2) est convergente, les probabilités  $A_0, A_1, \dots$  ont des valeurs déterminées non nulles et la probabilité  $A_{\infty}$  est nulle; dans le cas de la divergence, les probabilités  $A_k$  sont toutes nulles et  $A_{\infty}$  est égale à l'unité.

$M$  nello spazio, sapendo che la somma dei quadrati delle loro distanze dall'origine ha un valore dato.<sup>13</sup> ([11], p. 169)

Per comprendere questa affermazione, rivolgiamo l'attenzione alla dimostrazione della legge di distribuzione delle velocità molecolari di Maxwell data da Borel nel 1906 [9]. Qui Borel considerò un gas come formato da  $n$  molecole sferiche, di diametro  $a$ , racchiuse in un contenitore rigido e mobili di moto rettilineo uniforme salvo nel momento di un urto *elastico* con un'altra molecola o contro le pareti. In questo caso la traiettoria cambia secondo le leggi della riflessione e l'energia cinetica totale si conserva nel corso del moto. Borel sostituì al moto molecolare estremamente complesso quello di un punto  $P$  in uno spazio  $m = 3n$  dimensionale dimostrando che, anche per questo punto, gli urti contro le ipersuperficie astratte che descrivono le superficie di altre molecole o le pareti del contenitore, avvengono secondo le leggi della riflessione nello spazio  $m$ -dimensionale introdotto. Se  $(v_{3i-2}, v_{3i-1}, v_{3i})$  sono le componenti lungo assi ortogonali della velocità dell' $i$ -esimo punto materiale, poiché l'energia cinetica complessiva si deve conservare, il punto  $P$  appartiene sempre ad una ipersfera nello spazio  $m = 3n$ -dimensionale, di equazione

$$\sum_{i=1}^m v_i^2 = nk^2,$$

in cui si suppone che tutte le molecole abbiano la stessa massa e  $k^2$  rappresenta l'energia cinetica media del sistema. Borel si chiese quale fosse la probabilità che  $v_1$  abbia valori compresi tra  $u$  ed  $u + du$  e riformulò questo problema cercando il rapporto tra la superficie della regione di ipersfera limitata tra i due piani  $v_1 = u$  e  $v_1 = u + du$  e la superficie totale dell'ipersfera. Posto  $y_i = v_i$  e  $r^2 = nk^2$ , Borel considerò le coordinate polari sferiche in  $m$  dimensioni, definite da

$$\begin{aligned} y_1 &= r \cos \varphi_1 \\ y_2 &= r \sin \varphi_1 \cos \varphi_2 \\ y_3 &= r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3 \\ &\dots \\ y_{m-1} &= r \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{m-2} \cos \varphi_{m-1} \\ y_m &= r \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{m-2} \sin \varphi_{m-1} \end{aligned}$$

e calcolò lo jacobiano della trasformazione ottenendo per l'elemento di ipersuperficie  $dw$  l'espressione

$$dw = r^{m-1} \sin^{m-2} \varphi_1 \sin^{m-3} \varphi_2 \dots \sin \varphi_{m-1} d\varphi_1 d\varphi_2 \dots d\varphi_{m-1}.$$

Poiché nella regione di ipersuperficie di cui si intende stimare la grandezza è solo  $y_1$  a variare, cioè l'angolo  $\varphi_1$ , il rapporto cercato è

$$\frac{\sin^{m-2} \varphi d\varphi}{\int_0^\pi \sin^{m-2} \varphi d\varphi}. \quad (10.3)$$

<sup>13</sup>On est ainsi conduit à étudier un pur problème de théorie des probabilités. Représentons la vitesse de chaque molécule par un vecteur  $OM$  ayant son origine en un point fixe  $O$  et équipollent à cette vitesse; il s'agit de trouver la distribution la plus probable des points  $M$  dans l'espace, sachant que la somme des carrés de leurs distances l'origine a une valeur donnée.

Borel calcolò il numeratore introducendo  $u = r \cos \varphi$ , ricavando  $du = -r \sin \varphi d\varphi$  e quindi, omettendo il segno meno che figura anche a denominatore,

$$\sin^{m-2} \varphi d\varphi = \left(1 - \frac{u^2}{r^2}\right)^{\frac{m-3}{2}} du = e^{\frac{m-3}{2} \log\left(1 - \frac{u^2}{r^2}\right)} du,$$

che è proporzionale alla probabilità cercata. Se si considera che  $m = 3n \gg 1$  e si ricorda che  $r^2 = nk^2$ , sviluppando l'esponente e trascurando i termini in cui  $n$  figura a denominatore dell'esponente, si ottiene il valore

$$e^{-\frac{3u^2}{2k^2}} du.$$

Osservando poi che

$$\int_{-\infty}^{\infty} e^{-\frac{3u^2}{2k^2}} du = k\sqrt{\frac{2\pi}{3}}$$

Borel ottenne che la probabilità che  $v_1$  sia compreso tra  $u$  ed  $u + du$  è la legge di Maxwell:

$$\frac{\sqrt{3}}{k\sqrt{2\pi}} e^{-\frac{3u^2}{2k^2}} du :$$

Dunque questa legge ci appare solo come una legge di probabilità (...). Nulla ci permette di dire che la legge di Maxwell *diventi* più probabile quando il tempo passa; tutto quello che si può dire è che, moltiplicando gli esperimenti o prolungandoli, si permette alla legge dei grandi numeri di manifestarsi, malgrado le possibili deviazioni transitorie.<sup>14</sup> ([9], p. 31).

Borel non fu l'unico matematico ad interessarsi alla deduzione della legge di Maxwell. Nel primo trattato italiano completo a stampa di calcolo delle probabilità, pubblicato da Guido Castelnuovo nel 1919 [16], il capitolo conclusivo (pp. 266-298) è dedicato all'argomento, e vi si mette in luce il ruolo del calcolo delle probabilità nella deduzione.

## 10.3 I fondamenti della probabilità secondo von Mises

Matematico ed ingegnere, Richard von Mises (1883-1953) nacque a Lemberg (Leopoli), oggi città Ucraina ma all'epoca sotto il controllo dell'impero austro-ungarico. Von Mises si pose il problema di uno scrutinio accurato dei fondamenti del calcolo delle probabilità in un lavoro apparso nel 1919 in cui esordiva scrivendo:

<sup>14</sup>Cette loi nous apparaît donc ainsi comme étant uniquement une loi de probabilité (...). Mais rien n'autorise à dire que la loi de Maxwell *devient* plus probable lorsque le temps croît; tout ce que l'on peut dire, c'est qu'en multipliant les expériences, ou en les prolongeant, on permet à la loi des grands nombres de se manifester malgré les écarts passagers possibles.

Nessun matematico, che abbia tra le mani uno dei libri di testo esistenti, negherà la necessità di fondamenti precisi per il calcolo delle probabilità: di fatto non è possibile qualificare in altro modo lo stato attuale se non dicendo che il calcolo delle probabilità oggi *non è una disciplina matematica*.<sup>15</sup> ([43], p. 52)

Non solo la definizione classica di probabilità, ma la formulazione stessa dei problemi è avvolta da ombre fitte. Von Mises vedeva i tentativi di Bohlmann, Broggi e Borel come accumulati dall'attenzione verso aspetti formali mentre l'approccio che si accingeva a presentare partiva da una diversa concezione (*Auffassung*) della probabilità come una scienza naturale dello stesso tipo della geometria o della meccanica teorica con lo scopo di essere

non una fedele rappresentazione del mondo esterno, quanto una sua astrazione ed idealizzazione.<sup>16</sup> ([43], p. 53)

Il concetto cruciale su cui si fonda la visione di von Mises è quello di *Collettivo* (*Kollektiv*)  $K$  che, nella sua idea, deve rendere matematicamente preciso l'analogo concetto formulato qualche tempo prima da Gustav Theodor Fechner (1801-1887) e Heinrich Bruns. Si consideri una successione infinita  $(e) := \{e_1, e_2, \dots\}$  di elementi a ciascuno dei quali viene associato un insieme (ordinato) di  $k$  numeri reali, ciascuno dei quali è concepibile come un punto in uno spazio a  $k$  dimensioni detto *spazio delle caratteristiche* (*Merkmalraum*), che indicherò di seguito con  $\mathcal{M}$ . Per esempio, si può pensare ad  $(e)$  come alla successione dei lanci di un dado; l'elemento  $e_j$  è il  $j$ -esimo lancio del dato e la caratteristica o attributo corrispondente è il punteggio ottenuto. L'associazione degli elementi della successione alle caratteristiche dà luogo ad un collettivo  $K$  purché siano verificate due richieste. Anzitutto, l'esistenza del limite (*Existenz der Grenzwerte*):

1. Sia  $A \subset \mathcal{M}$  un sottinsieme a piacere ed  $N_A$  il numero di elementi, tra i primi  $N$  che formano  $(e)$ , la cui caratteristica appartiene ad  $A$ , affinché  $(e)$  sia un collettivo deve esistere il limite

$$\lim_{n \rightarrow \infty} \frac{N_A}{N} = W_A : \quad (10.4)$$

Da questa prima condizione discende che  $W_A \in [0, 1]$  e che, se  $A$  e  $B$  sono sottinsiemi di  $\mathcal{M}$ , con  $A \cap B = \emptyset$ , allora

$$W_A + W_B = W_{A \cup B};$$

l'insieme dei valori di  $W_A$  al variare di  $A$  in  $\mathcal{M}$  costituisce una *distribuzione* (*Verteilung*). Inoltre, se  $A \equiv \mathcal{M}$ ,  $W_A = 1$  mentre se  $A \equiv \emptyset$ ,  $W_A = 0$ . Come vedremo  $W_A$  misurerà la probabilità dell'evento associato ad  $A$ .

<sup>15</sup>Dem Bedürfnis nach einer exacten Grundlegung der Wahrscheinlichkeitsrechnung wird sich kein Mathematiker verschließen, der eines der bestehenden Lehrbücher zur Hand nimmt: in der Tat kann man den gegenwärtigen Zustand kaum anders als dahin kennzeichnen, daß die Wahrscheinlichkeitsrechnung heute *eine mathematische Disziplin nicht ist*.

<sup>16</sup>nicht als getreuer Abbild des Außenwelt, sondern als deren Abstraktion und Idealisierung.

L'esistenza del limite non basta però ad individuare un collettivo. Von Mises lo chiarisce con un semplice esempio che togliamo da un suo importante saggio [45] che, dopo la prima pubblicazione nel 1928, sarà ristampato con modifiche e tradotto in inglese. Si consideri una strada lungo la quale vi sono due tipi di pietre miliari. Le prime, più grandi, sono a distanza di un chilometro una dall'altra mentre le altre, più piccole, sono distanziate tra loro di 100 metri una dall'altra. Se facciamo corrispondere il numero 1 alle volte in cui incontriamo nel nostro cammino una delle pietre miliari più grandi e 0 alle volte in cui incontriamo una delle pietre miliari più piccole, le caratteristiche ( $c$ ) della successione ( $e$ ) che è costituita dalle osservazioni fatte ogni 100 metri lungo la strada, sono del tipo

$$(c) = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, \dots).$$

Lo spazio delle caratteristiche è in questo caso  $\mathcal{M} = \{0, 1\}$  e consideriamo il suo sottoinsieme  $A = \{1\}$ ; allora il limite (10.4) esiste e vale

$$\lim_{N \rightarrow \infty} \frac{N_A}{N} = \frac{1}{10}.$$

Ora, *estriamo* una sottosuccessione da ( $c$ ) secondo una regola prestabilita che però *non* dipenda dalle caratteristiche di ( $c$ ): ad esempio, estraiamo la sottosuccessione ( $c'$ ) formata dalle osservazioni di posto dispari. La corrispondente successione di caratteristiche ( $c'$ ) è

$$(c') = (1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, \dots)$$

ed ora

$$\lim_{N \rightarrow \infty} \frac{N_A}{N} = \frac{1}{5},$$

*diverso* dal valore precedente. Per von Mises, una successione di questo tipo *non* può essere utilizzata per simulare una successione di eventi aleatori perché, prescindendo dalle caratteristiche, abbiamo modificato  $W_A$  che è candidata a rappresentare la probabilità di  $A$ . Se ciò fosse possibile, osserva von Mises, i banchi da gioco di Monte Carlo non riuscirebbero a sopravvivere. Al contrario, l'impossibilità pratica a costruire una strategia di gioco che dia la certezza di vincere sempre è la miglior verifica sperimentale al fatto che le successioni di eventi aleatori debbono presentare una estrema irregolarità, nel senso che non è possibile modificare  $W_A$  con una scelta fissa. Per chiarire ulteriormente, von Mises propose un semplice esperimento: si estraiga con reimbussolamento un gettone da un'urna che ne contiene 90, numerati da 1 a 90: se si estrae un numero pari, si associa all'estrazione la caratteristica 0 mentre se l'estrazione ha dato un numero dispari, le si associa la caratteristica 1. Si eseguano 100

estrazioni e si dispongano le caratteristiche in una matrice:

1	1	0	0	0	1	1	1	0	1
0	0	1	1	0	0	1	1	1	1
0	1	0	1	0	0	1	0	0	0
0	1	0	0	1	0	0	1	1	1
0	0	1	1	0	0	0	0	1	1
0	1	1	1	1	0	1	0	1	0
1	0	1	1	1	1	0	0	1	1
0	0	1	1	0	1	1	1	0	1
0	0	1	1	0	0	1	1	0	1
0	1	1	0	0	0	1	0	0	0

La frequenza relativa di 1 è di  $\frac{51}{100}$ ; se consideriamo solamente gli esiti sulle colonne dispari, a partire da sinistra, questa frequenza è  $\frac{24}{50} = \frac{48}{100}$ ; usando solo le righe dispari, partendo dall'alto, si ottiene  $\frac{25}{50} = \frac{50}{100}$ ; se si considerano solo i 26 esiti corrispondenti alla prima estrazione e a tutte quelle che, procedendo riga per riga, sono in una posizione data da un numero primo, la frequenza degli 1 è  $\frac{13}{26} = \frac{50}{100}$ , ecc. A differenza del caso precedente, i valori ottenuti per la frequenza sono tutti vicini tra loro e prossimi ad  $\frac{1}{2}$ , valore di  $W_A$  che non è possibile cambiare operando una selezione in base ad una regola esterna: è questo il segno della irregolarità della corrispondenza tra evento e caratteristica, tipico per von Mises dei fenomeni aleatori. Questo esempio però evidenzia anche i punti deboli, a livello formale, della proposta di von Mises perché il limite (10.4) che serve a definire la frequenza *non* è il limite dell'analisi, tanto che in quest'ultimo esempio  $N$  non tende affatto all'infinito. Questa vaghezza è il prezzo da pagare per radicare sin dall'inizio la probabilità nel mondo *reale*; tentativi di modificare la nozione di collettivo in modo da renderla formalmente più solida saranno effettuati da Hans Reichenbach (1891-1953) ed Abraham Wald (1902-1950). Un'altra obiezione che von Mises dovette affrontare fu il ricorso a successioni *infinite* per definire la probabilità associata ad un sottoinsieme dello spazio delle caratteristiche. In occasione di una Conferenza di Epistemologia tenutasi a Praga nel 1929, Otto Neurath chiese a von Mises ([30], p. 277):

Sorge la domanda (...) se non sia possibile escludere il concetto di infinito, non già costruendo il calcolo matematico delle probabilità sul concetto di collettivo ma su quello di un insieme finito di elementi.<sup>17</sup>

Nella risposta, von Mises riconosceva la serietà del problema sollevato ed ammetteva di averci pensato molte volte e proponeva all'attenzione dell'interlocutore l'analogia con la geometria, dove si ragiona su superficie e curve matematiche che sono delle idealizzazioni delle loro controparti reali. Come aveva osservato Klein mezzo secolo prima, occorre distinguere tra una matematica di precisione (*Präzisionsmathematik*) ed una matematica di approssimazione (*Approximationsmathematik*): nel secondo caso le curve, ad esempio, sono delle strisce

<sup>17</sup>Es entsteht (...) die Frage, ob sich nicht der Begriff des Infiniten ausschalten ließe, indem schon das Gebäude der mathematischen Wahrscheinlichkeitsrechnung statt auf dem Begriff des Kollektivs, auf dem einer endlichen Menge von Elementen aufgebaut werden könnte.

sottili che hanno uno spessore piccolo ma non trascurabile. Ciò comporta tutta una serie di complicazioni legate, per esempio, a cosa si deve intendere per intersezione di due curve, intese in questa accezione estesa. von Mises aveva al proposito un atteggiamento pragmatico: ([30], p. 279)

Su basi pratiche si adopera la teoria infinita, si calcola, anche se si ha a disposizione materiale formato da parti finite, come se fossero infinite e ci si serve in tutte le operazioni di regole di calcolo che riposano sull'ipotesi di collettivi infiniti perché sono più brevi, più sicure, più chiare.<sup>18</sup>

Un tentativo di fondare il calcolo delle probabilità a partire da un concetto di collettivo finito, con successioni *finite* di caratteristiche sarà proposta qualche anno dopo da Hans Blume [4, 5].

Tornando al lavoro del 1919, von Mises vi formulò in questi termini la condizione di irregolarità della corrispondenza (*Regellosigkeit der Zuordnung*) tra elementi di una successione e caratteristiche:

**2.** Siano dati due sottoinsiemi disgiunti  $A$  e  $B$  di  $\mathcal{M}$  per i quali esiste il limite (10.4), con  $W_A$  e  $W_B$  entrambi non nulli. Si rimuovano dalle successioni ( $e$ ) tutti gli elementi le cui caratteristiche *non* appartengono ad  $A$  o a  $B$ . Dalla successione ottenuta si estragga una sottosuccessione infinita ( $e'$ ) senza ricorrere alle caratteristiche per selezionarne i termini; allora i limiti  $W'_A$  e  $W'_B$  associati ad ( $e'$ ) debbono esistere ed inoltre si deve avere

$$W'_A : W'_B = W_A : W_B \quad (10.5)$$

Una corrispondenza tra successioni infinite ed i punti dello spazio  $k$ -dimensionale  $\mathcal{M}$  che obbedisca alle condizioni 1 e 2 è detta *casuale* (*zufallsartige*) e

un collettivo è una successione infinita di elementi a ciascuno dei quali corrisponde in modo casuale una caratteristica.<sup>19</sup> ([43], p. 58)

Se  $A$  e  $B$ , oltre ad essere disgiunti, sono tali che  $A \cup B = \mathcal{M}$ , allora si dovrà avere

$$W_A + W_B = 1 \quad \text{e} \quad W'_A + W'_B = 1$$

e, applicando la proprietà del comporre all'equazione (10.5), si ottiene  $W'_A : (W'_A + W'_B) = W_A : (W_A + W_B)$  e  $(W'_A + W'_B) : W'_B = (W_A + W_B) : W_B$ , da cui segue che

$$W_A = W'_A \quad \text{e} \quad W_B = W'_B$$

che, per von Mises, è una richiesta di conformità all'imprevedibilità del caso: in opere successive a [43], von Mises si limiterà a insiemi  $A$  e  $B$  disgiunti e tali che  $A \cup B = \mathcal{M}$ . Come conseguenza, se dalla successione ( $e$ ) di un collettivo  $K$  si

<sup>18</sup>Aus praktischen Gründen bedient man sich der infiniten Theorie, man rechnet, obwohl das Versuchsmaterial endlichen Abschnitten entnommen ist, so, als ob es unendlich wäre und in allen Operationen bedient man sich der Rechenregeln, die auf der Annahme unendlicher Kollektivs beruhen, weil sie kürzer, einfacher, übersichtlicher sind.

<sup>19</sup>Ein Kollektiv ist eine unendliche Folge von Elementen mit zufallsartiger Zuordnung von Merkmalen an die einzelnen Elemente.

estrae una sottosuccessione ( $e'$ ), senza utilizzare le caratteristiche degli elementi selezionati, si ottiene un nuovo collettivo  $K'$ , con la stessa distribuzione di  $K$ . Per sviluppare da queste premesse il calcolo delle probabilità occorre collegare la distribuzione di un collettivo con quelle di altri collettivi dedotti dal primo con opportune operazioni. von Mises sottolinea che

la nostra definizione di probabilità non ha nulla in comune con quella proposta da Laplace ed oggi usata generalmente, che emerge dalla considerazione dei cosiddetti casi ugualmente possibili. Il caso di un dado *truccato* non è in generale accessibile con questa definizione. (...) Nel quadro dei nostri sviluppi, ricondursi ai casi ugualmente possibili, finora generalmente trattabile e corretto, è un esempio molto particolare di *calcolo* delle probabilità.<sup>20</sup> ([43], p. 62)

Le operazioni più semplici da eseguire sui collettivi per generarne altri furono chiamate da von Mises: selezione (*Auswahl*), mescolamento (*Mischung*), partizione (*Teilung*) e collegamento o combinazione (*Verbindung*). L'operazione di selezione consente di ottenere un collettivo selezionando, secondo una regola stabilita, una sottosuccessione appartenente al collettivo di partenza. In virtù della condizione **2**, questa operazione non altera la distribuzione di probabilità del collettivo di partenza. L'operazione di mescolamento consiste generare un collettivo in cui sono raggruppate più caratteristiche presenti nel collettivo di partenza. Per esempio, se il collettivo di partenza ha come caratteristiche i sei valori che si possono presentare nel lancio di un dado, è possibile con l'operazione di combinazione far confluire le tre caratteristiche 2, 4, 6 nell'unica nuova caratteristica: "numero pari"; analogamente le tre caratteristiche 1, 3, 5 si possono raggruppare nell'unica caratteristica: "numero dispari". Con questa operazione si intravede come giungere alla probabilità per eventi composti da a partire da quella per eventi semplici. Con l'operazione di partizione si giunge alla definizione di probabilità condizionata. Sempre riferendoci all'esempio dei dadi, supponiamo che lo spazio  $\mathcal{M}$  del primo collettivo sia formato dai sei esiti di un lancio e si supponga di volere sapere quale sia la probabilità di ottenere 4, *sapendo* che il lancio ha dato per esito un numero pari. Se la successione del collettivo iniziale ha associate le caratteristiche che si sono presentate in un esperimento ([17], pp. 38-39):

6 6 1 3 5 1 3 2 2 5 5...

e si eliminano gli esiti corrispondenti ai numeri dispari, si ottiene la successione

6 6 2 2 ... .

<sup>20</sup>Unsere Definition der Wahrscheinlichkeit hat nichts gemein mit der von Laplace herrührenden, heute allgemein verwendeten, die von der Betrachtung sog.[ennanten] gleichmöglicher Fälle ausgeht. (...) Der Fall eines *falschen* Würfels ist dieser Definition überhaupt nicht zugänglich. (...) In Rahmen unserer Entwicklungen erscheint die Zurückführung auf gleichmögliche Fälle, soweit sehr überhaupt durchführbar und sachgemäss ist, als ein ganz spezielles Beispiel einer *Berechnung* von Wahrscheinlichkeiten.

Se  $W_2$ ,  $W_4$  e  $W_6$  erano le probabilità di ottenere 2, 4 o 6 nel primo collettivo, nel nuovo collettivo esse diventano

$$W'_2 = \frac{W_2}{W_2 + W_4 + W_6} \quad W'_4 = \frac{W_4}{W_2 + W_4 + W_6} \quad W'_6 = \frac{W_6}{W_2 + W_4 + W_6}.$$

Le probabilità sono cambiate e ciò non deve sorprendere perché, a differenza dell'operazione di selezione, la scelta della sottosuccessione che forma il nuovo collettivo è condotta a partire dalle caratteristiche del collettivo di partenza. Infine, con l'operazione di collegamento, si forma un nuovo collettivo a partire da due collettivi,  $K'$  e  $K''$ . Questa operazione è alla base della nozione di indipendenza tra due eventi ed infatti von Mises parlò in prima battuta di *collegamento di collettivi indipendenti* (*Verbindung unabhängiger Kollektivs*). L'operazione di combinazione consiste nello stabilire una corrispondenza *biunivoca* tra le caratteristiche di  $K'$  e quelle di  $K''$ . Con un esempio [45], supponiamo che le caratteristiche di  $K'$  siano gli esiti dei lanci di un dado, le cui facce sono rosse mentre le caratteristiche di  $K''$  siano gli esiti dei lanci di un dado dalle facce bianche. Siamo interessati all'evento: si ottiene il punteggio 3 con il primo dado e il punteggio 5 con il secondo. Sia  $n_3$  il numero di lanci del primo dado in cui si è ottenuto il punteggio 3 e consideriamo i lanci corrispondenti del secondo dado. *Tra questi*, siano  $n'_5$  quelli nei quali il secondo dado ha dato 5 come punteggio. Il rapporto  $\frac{n'_5}{n_3}$  tende ad un limite ma quale ne è il significato? L'indipendenza dei collettivi significa postulare che, nel nuovo collettivo, questo rapporto *coincida* con la probabilità di ottenere 5 con il secondo dado. Allora, preso un numero  $n$  di prove, possiamo scrivere

$$\frac{n'_5}{n} = \frac{n'_5}{n_3} \frac{n_3}{n}$$

e, quando  $n \rightarrow \infty$  la frequenza  $\frac{n_3}{n}$  tende alla probabilità  $p_3$  che il 3 esca sul primo dado mentre  $\frac{n'_5}{n_3}$  tende, se gli eventi sono indipendenti, a  $p'_5$ , probabilità che il secondo dado fornisca 5 come punteggio, sapendo che sul primo dado è apparso il 3. D'altra parte, grazie alla corrispondenza *biunivoca* stabilita tra i collettivi  $K'$  e  $K''$ ,  $\frac{n'_5}{n}$  tende alla probabilità  $p_{3 \cap 5'}$  che esca 3 sul primo dado e 5 sul secondo. Si ottiene allora

$$p_{3 \cap 5'} = p_3 p'_5.$$

Sottolineiamo come, in definitiva, abbiamo fatto un *campionamento* di  $K''$  tramite il collettivo  $K'$  e non una scelta fissa degli elementi di  $K''$ . Per questo, la condizione che  $\frac{n'_5}{n_3}$  rappresenti *ancora* la probabilità che 5 sia il punteggio ottenuto nel secondo dado non è una conseguenza gratuita dell'aleatorietà del collettivo  $K'$  ma la sua validità deve essere verificata. Per quanto nebulosa, la nozione di limite utilizzata da von Mises vuole comunque mettere al centro della definizione di probabilità il confronto con le esperienze del mondo reale ma proprio per questo essa esercita pesanti limitazioni sul tipo di problemi che rientrano nel dominio del calcolo delle probabilità [17]. Una conseguenza di questo approccio è che la teoria delle probabilità non può occuparsi di eventi singolari,

per i quali non si possano effettuare prove ripetute, almeno virtualmente, all'infinito. In questo modo le applicazioni della probabilità sono sostanzialmente limitate ai fenomeni fisici ed alla statistica. Si tratta di un'osservazione ripetuta più volte da von Mises che accostava la nozione di collettivo a quella di *popolazione* utilizzata in statistica [44].

## 10.4 Misurare la probabilità soggettiva

Bruno de Finetti (1906-1985) fu una delle personalità di maggior spessore nella storia del calcolo delle probabilità del XX secolo. Il suo nome resta legato in modo indissolubile all'interpretazione *soggettivista* del calcolo delle probabilità da lui proposta a partire da due lavori pubblicati nel biennio 1930-31: il primo, di natura schiettamente matematica [21], il secondo [24] di indole filosofica. Le interpretazioni date alla probabilità sono, è noto, differenti ed i confini non sono sempre netti. All'interpretazione classica, fondata sul problematico concetto di casi egualmente possibili, si è affiancata l'interpretazione frequentista—che tenta di rendere più preciso un fondamento sull'osservazione dell'esistenza di rapporti costanti in prove ripetute a lungo—e l'impostazione soggettivista

che considera la probabilità come misura del grado di fiducia di un soggetto determinato nell'avverarsi di un evento. ([28], p.36)

Anche se in [28] de Finetti annoverò con riserva l'economista britannico John Maynard Keynes tra i soggettivisti, l'approccio di Keynes alla probabilità è inquadrabile nel logicismo.

Come accennato, nel 1931 De Finetti scrisse un saggio, intitolato *Probabilismo*, che rappresenta il suo manifesto filosofico e apparve sulla rivista *Logos. Biblioteca di Filosofia* fondata e diretta dal filosofo siciliano Antonio Aliotta. Questo saggio contiene alcune citazioni che consentono di comprendere la posizione di De Finetti rispetto ad altre impostazioni del calcolo delle probabilità. De Finetti parte dalla constatazione che, tramontata l'epoca del determinismo, l'ufficio della scienza deve necessariamente cambiare, non essendo più sostenibile l'esistenza di “leggi immutabili e necessarie” che reggono l'universo intero.

Nessuna scienza ci permetterà di dire: il tale fatto accadrà, andrà così e così, perché ciò è conseguenza di tale legge, e tale legge è una verità assoluta, ma tanto meno ci condurrà a concludere scetticamente: la verità assoluta non esiste, e quindi tale fatto può accadere e non accadere, può andare così e può andare in tutt'altro modo, nulla io ne so.

Quel che si potrà dire è questo: *io prevedo* che il tale fatto avverrà, e avverrà nel tal modo, perché l'esperienza del passato e l'elaborazione scientifica cui il pensiero dell'uomo l'ha sottoposta mi fanno sembrare ragionevole questa previsione. [24]

A questo passaggio dalla predizione alla previsione come compito della scienza, risponde anche un nuovo tipo di logica:

la logica è la scienza che dalla verità o falsità di certe premesse insegna a dedurre e concludere la verità o falsità di certe conseguenze; a seconda del senso che daremo al concetto di verità, avremo due modi diversi di concepire la logica. Se la verità si concepisce in senso obbiettivo, la logica appare come una proprietà di cui deve godere il mondo reale, come una specie di legge esteriore che regola la verità o falsità, in senso obbiettivo, di certe proposizioni. Se ci si limita invece all'aspetto soggettivo, la logica non riguarda che i processi mentali, e non insegna se non la coerenza del pensiero con sè [sic!] stesso. (...) Il calcolo delle probabilità è la logica del probabile. Come la logica formale insegna a dedurre la verità o falsità di certe conseguenze dalla verità o falsità di certe premesse, così il calcolo delle probabilità insegna a dedurre la maggiore o minore verosimiglianza o probabilità di certe conseguenze dalla maggiore o minore verosimiglianza o probabilità di certe premesse. Per chi attribuisca alla probabilità un significato obbiettivo, il calcolo delle probabilità dovrebbe avere un significato obbiettivo, i suoi teoremi esprimere delle proprietà che nel campo del reale risultano soddisfatte. Ma è inutile fare simili ipotesi. Basta limitarsi alla concezione soggettiva, considerare cioè la probabilità come il grado di fiducia sentito da un dato individuo nell'avverarsi di un dato evento, e si può dimostrare che i noti teoremi del calcolo delle probabilità sono condizioni necessarie e sufficienti perché le opinioni di un dato individuo non siano intrinsecamente contraddittorie e incoerenti ([21], pp. 1-2)

Questo lungo passaggio contiene la parola chiave dell'approccio di De Finetti: *coerenza*. È la coerenza il requisito in base al quale giudicare se siano sensate le valutazioni sulle probabilità di realizzazione di un evento date da un soggetto. La formalizzazione di questo requisito di "buon senso" è il cuore del successivo lavoro [25] in cui viene illustrato lo schema delle scommesse che servì da tramite per rendere operativa l'impostazione soggettivista. Qui De Finetti propose alcuni esempi contenenti, magari in forma implicita, ragionamenti soggettivi di indole probabilistica:

Uno degli esempi più suggestivi è quello delle indagini poliziesche o giudiziarie, dove si procede sempre per indizi e individui, dove non si lavora mai sul certo, ma sempre e soltanto sul probabile. Supponiamo ad esempio: uno sconosciuto ha commesso un delitto; alcune tracce [sic!] fanno ritenere tre individui come fortemente indiziati. Il Commissario che attribuisce un certo grado d'attendibilità a ciascuna di queste tre ipotesi, quale fiducia può avere che le indagini siano bene avviate, e cioè che una delle tre ipotesi sia la vera? Sostanzialmente, se non numericamente, egli applica senza dubbio il teorema delle probabilità totali. E allo stesso modo applicherà il teorema delle probabilità composte se vorrà valutare con quale probabilità, ossia se penserà con quale fiducia, egli possa attendere di identificare in uno dei tre indiziati il colpevole e trarlo in arresto. ([25], p. 300)

Occorre ora comprendere come

*misurare* la probabilità soggettiva ([25], p. 301)

e De Finetti si ricollega ad alcune considerazioni di Bertrand:

La probabilità di un evento, quale che sia la sua natura, è detta essere uguale ad una frazione assegnata  $p$  quando chi si aspetta l'evento potrà scambiare indifferentemente i timori o le aspettative, i vantaggi o gli inconvenienti legati al realizzarsi di questo evento confrontate con le conseguenze ritenute identiche dell'estrazione di una palla da un'urna la cui composizione dà luogo ad una probabilità uguale.<sup>21</sup> ([3], p. 27)

Bertrand chiarisce con un esempio, cui fa riferimento De Finetti [24]:

Dopo aver chiamato un medico, si stimi in  $\frac{9}{10}$  la probabilità che venga e in  $\frac{1}{3}$  quella che, essendosi presentato, il medico guarisca il malato; senza discutere questi valori, chi li ritenga corretti può aggiungere: la probabilità che il malato sia visitato e guarito dal medico è *per me*

$$\frac{9}{10} \times \frac{1}{3} = \frac{3}{10}.$$

In effetti, chi accetta le probabilità di  $\frac{9}{10}$  ed  $\frac{1}{3}$  è, per definizione, nella stessa situazione di incertezza di chi, avendo di fronte a sé dieci urne indistinguibili, di cui nove contengono una pallina bianca e una nera, cercasse la probabilità di estrarre una pallina bianca da una delle urne di composizione nota. L'identità dei due problemi viene ammessa e fa parte dell'enunciato.<sup>22</sup> ([3], p. 27)

È a questo punto che De Finetti introduce lo schema delle scommesse, prendendo le mosse dal concetto di speranza matematica: in qualche modo è come se avesse voluto tornare alle origini, ad Huygens e Pascal per esempio, che fecero della nozione di gioco equo il perno attorno a cui far ruotare l'edificio del nascente calcolo delle probabilità. Ricordiamo che anche Bayes definì la probabilità indirettamente, partendo dal concetto di speranza matematica. De Finetti prende come punto di vista quello di un individuo  $\mathcal{O}$  che debba accettare scommesse

<sup>21</sup>La probabilité d'un événement, quelle qu'en soit la nature, est dite égale à une fraction donnée  $p$ , lorsque celui qui attend l'événement pourrait échanger indifféremment les craintes ou les espérances, les avantages ou les inconvénients attachés à l'arrivée de cet événement contre les conséquences supposées identiques de la sortie d'une boule puisée dans une urne dont la composition fait naître une probabilité égale.

<sup>22</sup>Si, après avoir appelé un médecin, on évalue it la probabilité pour qu'il vienne et à  $\frac{1}{3}$  la probabilité pour qu'il procure, s'il vient, la guérison du malade; sans discuter ces chiffres, celui qui les admet peut ajouter: La probabilité pour que le malade soit visité et guéri par le médecin est, *pour moi*,

$$\frac{9}{10} \times \frac{1}{3} = \frac{3}{10}.$$

Celui qui accepte, en effet, les probabilités  $\frac{9}{10}$  et  $\frac{1}{3}$  est, par définition, dans le même doute que si, eu présence de dix urnes indiscernables dont neuf renferment une boule blanche et deux boules noires, il cherchait la probabilité pour mettre la main sur l'une des neuf urnes dont la composition est connue et en tirer une boule blanche. L'identité des deux problèmes est admise; elle fait partie de l'énoncé.

su un certo numero  $n$  finito di eventi  $E_1, \dots, E_n$  che possono essere la vittoria di un concorrente in una gara, su un lotto di  $n$  partecipanti. Le scommesse sono rette da queste regole:

È in facoltà del soggetto  $\mathcal{O}$  che tiene il banco di stabilire il prezzo  $p$  di un buono, o obbligazione, che dà diritto a riscuotere una lira nel caso che un dato evento  $E$  si verifichi; (...) Qualunque competitore si presenti al banco di  $\mathcal{O}$  e voglia scommettere per l'evento  $E$ , è cioè in facoltà di comprare al prezzo  $pS$  un'obbligazione che gli dà diritto, se vince la scommessa (e cioè se  $E$  si verifica), a esigere una somma generica  $S$ . (O inversamente, se vuole scommettere contro l'evento  $E$ , può impegnarsi a pagare la somma generica  $S$ , se perde la scommessa, e cioè se  $E$  si verifica, esigendo la riscossione della somma  $pS$ . Caso questo che rientra nel precedente se si considerano valori di  $S$  negativi). ([25], p. 304)

Il numero  $p$  è detto da De Finetti

probabilità dell'evento  $E$  secondo il soggetto  $\mathcal{O}$  ([25], p. 104)

perché, crescendo, testimonia l'accresciuta fiducia che  $\mathcal{O}$  ripone nel realizzarsi di  $E$ . De Finetti mette subito al riparo la sua definizione di probabilità dalla critica che il valore di  $p$  possa dipendere da fattori, come la propensione maggiore o minore al rischio di un soggetto, che sono estranei all'idea di probabilità come misura del grado di fiducia nel realizzarsi di un evento. Proprio per questo, nota De Finetti, il ruolo del soggetto  $\mathcal{O}$  è quello di chi tiene il banco delle scommesse per cui

egli è costretto allora a rispettare certe restrizioni, che sono i teoremi del calcolo delle probabilità. Altrimenti egli pecca di *coerenza*, e perde *sicuramente*, purché l'avversario sappia sfruttare il suo errore. ([25], p. 305)

Un soggetto è detto coerente quando le probabilità assegnate agli eventi che si possono presentare *non* permettono ad alcun competitore di vincere *con certezza*, cioè se non è possibile praticare contro  $\mathcal{O}$  un *Dutch Book argument*. Questa osservazione è cruciale e coglie il nervo della proposta operativa di De Finetti per rendere misurabile la probabilità soggettiva: se  $\mathcal{O}$  non vuole perdere sicuramente, deve conformarsi ai teoremi del calcolo delle probabilità. Prima di passare ad una vera e propria formalizzazione, De Finetti introdusse un esempio: una gara a cui prendono parte, tra gli altri, due concorrenti italiani  $A$  e  $B$ . Si vuole determinare la probabilità di una vittoria italiana (evento  $E$ ), che è quanto dire dimostrare il teorema delle probabilità totali. Se continuiamo ad indicare con  $A$  e  $B$  gli eventi che consistono nella vittoria di  $A$  e  $B$ , è chiaro che  $E = A \cup B$  e che  $A \cap B = \emptyset$ . Se  $\mathcal{O}$  valutasse la probabilità dell'evento  $A$  come  $p_A = 0,6$  e quella dell'evento  $B$  come  $p_B = 0,2$  ma decidesse che la probabilità di  $E$  è  $p = 0,75$  o, comunque, diversa da  $p_A + p_B = 0,8$  egli andrebbe incontro ad una perdita certa. Infatti, se  $\mathcal{O}$  valutasse  $p = 0,75$  perderebbe certamente

contro uno scommettitore che adottasse questa strategia: scommettere uguali cifre sul realizzarsi di  $E$  e sul *non* realizzarsi di  $A$  e di  $B$ . Infatti, fissato a 100 lire il valore di un buono, scommettendo a favore di  $E$  vengono versate ad  $\mathcal{O}$  75 lire mentre  $\mathcal{O}$  deve versare  $60 + 20 = 80$  lire allo scommettitore che gli renderà 100 lire se uno di questi eventi si verificherà. Lo scommettitore, prima ancora di conoscere l'esito della competizione, ha già guadagnato 5 lire e, quale che sia l'esito della competizione, guadagni e perdite ulteriori saranno in pareggio perché: 1) se si realizza  $E$  lo scommettitore incassa 100 lire ma, essendosi realizzato uno ed uno solo degli eventi  $A$  e  $B$ , verserà anche 100 lire ad  $\mathcal{O}$  per questo; al contrario, se  $E$  non si realizza, non incassa nulla dal primo buono, ma non deve restituire nulla per gli altri due buoni. Con questa strategia, lo scommettitore riesce a realizzare un "Dutch Book" contro  $\mathcal{O}$ . Passando alla formalizzazione, De Finetti considerò dapprima scommesse su un unico evento  $E$  ed introdusse il guadagno  $G(E)$  di uno scommettitore che scommetta *sul* realizzarsi dell'evento  $E$  quando la posta sia  $S$ , in termini della probabilità che  $\mathcal{O}$  assegna all'evento  $E$ :

$$G(E) = (1 - p)S \quad G(E^c) = -pS, \quad (10.6)$$

dove  $E^c$  è l'evento complementare di  $E$  e  $G(E^c)$  rappresenta il guadagno nel caso in cui, puntando sul realizzarsi di  $E$ , questo evento non si realizzi. Quando la scommessa è *contro* il realizzarsi di  $E$ , allora il guadagno diventa

$$G'(E) = (p - 1)S \quad G'(E^c) = pS \quad (10.7)$$

che differisce da (10.6) per un segno: si può considerare dunque solo la funzione (10.6), ammettendo per  $S$  valori tanto positivi che negativi. Il primo teorema dimostrato da De Finetti è il seguente:

**Teorema.** *Condizione necessaria e sufficiente per la coerenza di  $\mathcal{O}$  è che, accettando scommesse su un unico evento  $E$ , la probabilità attribuita a quell'evento sia un unico numero  $p(E) \in [0, 1]$ ; se l'evento è certo, allora  $p(E) = 1$  mentre se è impossibile,  $p(E) = 0$ .*

**Dim.** Supponiamo l'evento  $E$  certo. L'unico valore da considerare per  $G(E)$  è dato dalla prima equazione delle (10.6). Se fosse  $p \neq 1$ , si vede che  $\mathcal{O}$  non sarebbe coerente, in quanto, se  $p < 1$ , basta scommettere *su*  $E$  e (10.6)<sub>1</sub> mostra che  $\mathcal{O}$  perderebbe sempre, con  $S > 0$  mentre, se  $p > 1$ , basterebbe scommettere *contro*  $E$  e, per (10.7)<sub>1</sub>,  $\mathcal{O}$  perderebbe sempre, con  $S > 0$ . Similmente si ragiona sulle seconde equazioni (10.6) o (10.7) se  $E$  fosse impossibile e si ponesse  $p \neq 0$ . Nel caso in cui  $E$  non è né certo né impossibile, de Finetti mostra come la coerenza imponga ad  $\mathcal{O}$  di assegnare un unico valore per  $p(E)$ . Se si assegnassero due valori alla probabilità,  $p_1$  e  $p_2$ , e se  $\mathcal{O}$  accettasse scommesse con entrambi questi valori, un giocatore che scommettesse su  $E$  una somma  $S_1$ , quando la probabilità è valutata  $p_1$  ed una somma  $S_2$ , quando la probabilità è valutata  $p_2$ , guadagnerebbe, grazie alle equazioni (10.6)

$$\begin{cases} G(E) = (1 - p_1)S_1 + (1 - p_2)S_2 \\ G(E^c) = -p_1S_1 - p_2S_2, \end{cases} \quad (10.8)$$

a seconda che  $E$  si realizzi o meno. Il sistema (10.8), interpretato come sistema lineare nelle incognite  $S_1$  ed  $S_2$ , ha determinante della matrice dei coefficienti  $p_1 - p_2$ : se dunque  $p_1 \neq p_2$ , è possibile determinare  $S_1$  ed  $S_2$  in modo che i guadagni  $G(E)$  e  $G(E^c)$  siano entrambi positivi, realizzando un *Dutch book*, il che non può succedere, se  $\mathcal{O}$  è coerente.<sup>23</sup> Sempre la coerenza obbliga a scegliere il valore di  $p$  nell'intervallo  $[0, 1]$ . Che la condizione sia necessaria è evidente dalla (10.6) o dalla (10.7); la sufficienza segue dall'osservazione che

$$pG(E) + (1 - p)G(E^c) = 0$$

e quindi, se  $p \in [0, 1]$  è impossibile che  $G(E)$  e  $G(E^c)$  siano sempre positivi.

Per mostrare il teorema delle probabilità totali, de Finetti considerò una classe *finita* e completa di eventi  $E_1, \dots, E_n$  incompatibili, per cui  $\bigcup_{k=1}^n E_k$  è un evento certo e quindi

$$p\left(\bigcup_{k=1}^n E_k\right) = 1.$$

Questa condizione non lede la generalità perché, se la classe non fosse completa, esisterebbe un evento  $E_0$  che la completerebbe, ottenendo

$$p\left(\bigcup_{k=0}^n E_k\right) = 1.$$

Ora, assumendo per dimostrato il teorema delle probabilità totali per la classe completa  $\{E_0, E_1, \dots, E_n\}$ , si avrebbe

$$\sum_{k=0}^n p(E_k) = 1.$$

D'altronde  $E_0$  è pur sempre l'evento complementare di  $\bigcup_{k=1}^n E_k$  e, con questo evento, costituisce una classe completa, per cui

$$p(E_0) + p\left(\bigcup_{k=1}^n E_k\right) = 1$$

che, confrontata con l'equazione precedente mostra che

$$\sum_{k=1}^n p(E_k) = p\left(\bigcup_{k=1}^n E_k\right).$$

De Finetti riprodusse lo stesso argomento usato in (10.8): dette  $S_1, \dots, S_n$  le scommesse accettate da  $\mathcal{O}$  sugli eventi  $E_1, \dots, E_n$ , il guadagno ottenuto al realizzarsi di un certo evento  $E_j$  è

$$G(E_j) = S_j - \sum_{k=1}^n p_k S_k \quad \forall j = 1, \dots, n :$$

<sup>23</sup>Se la soluzione ottenuta avesse, per esempio  $S_1 < 0$ , sarebbe sempre possibile ottenere  $S_1 > 0$  scommettendo questa cifra *contro* il realizzarsi dell'evento  $E$ .

mettendo a sistema questi  $n$  guadagni, si verifica che il determinante della matrice dei coefficienti è  $1 - \sum_{k=1}^n p_k$  che dunque deve essere eguale a 0 per garantire la coerenza, dimostrando il teorema delle probabilità totali. Infatti, la matrice dei coefficienti è

$$\begin{pmatrix} 1-p_1 & -p_2 & -p_3 & \cdots & \cdots & -p_n \\ -p_1 & 1-p_2 & -p_3 & \cdots & \cdots & -p_n \\ -p_1 & -p_2 & 1-p_3 & \cdots & \cdots & -p_n \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -p_1 & -p_2 & -p_3 & \cdots & \cdots & 1-p_n \end{pmatrix}$$

ed il suo determinante non cambia se aggiungiamo alla prima colonna la somma di tutte le altre colonne, ottenendo

$$\begin{pmatrix} \Sigma & -p_2 & -p_3 & \cdots & \cdots & -p_n \\ \Sigma & 1-p_2 & -p_3 & \cdots & \cdots & -p_n \\ \Sigma & -p_2 & 1-p_3 & \cdots & \cdots & -p_n \\ \Sigma & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma & -p_2 & -p_3 & \cdots & \cdots & 1-p_n \end{pmatrix},$$

dove abbiamo posto

$$\Sigma = 1 - \sum_{k=1}^n p_k.$$

Sostituiamo alla seconda riga la sua differenza con la prima riga, cosicché neanche ora il determinante cambi; avremo

$$\begin{pmatrix} \Sigma & -p_2 & -p_3 & \cdots & \cdots & -p_n \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \Sigma & -p_2 & 1-p_3 & \cdots & \cdots & -p_n \\ \Sigma & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma & -p_2 & -p_3 & \cdots & \cdots & 1-p_n \end{pmatrix}$$

e ripetiamo la procedura per la terza riga, arrivando così a

$$\begin{pmatrix} \Sigma & -p_2 & -p_3 & \cdots & \cdots & -p_n \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ \Sigma & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma & -p_2 & -p_3 & \cdots & \cdots & 1-p_n \end{pmatrix}$$

finché non si giunge alla matrice

$$\begin{pmatrix} \Sigma & -p_2 & -p_3 & \cdots & \cdots & -p_n \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

il cui determinante è pari a  $\Sigma$ .

Similmente, la sufficienza dipende dal fatto che

$$\sum_{j=1}^n p_j G(E_j) = \left(1 - \sum_{k=1}^n p_k\right) \sum_{j=1}^n p_j S_j$$

e dunque, se  $\sum_{k=1}^n p_k = 1$ , non è mai possibile avere tutti i guadagni  $G(E_k)$  positivi. Come corollari, de Finetti mostrò che  $p(E^c) = 1 - p(E)$  e che, se l'evento  $E$  implica l'evento  $E_1$ , allora  $p(E) > p(E_1)$ . Osserviamo come la proprietà di additività numerabile della probabilità su eventi incompatibili sia esclusa:

per una classe infinita di eventi (sia pure numerabile) non è necessario che sussista la proprietà analoga della proprietà additiva. La probabilità della somma logica di infiniti eventi incompatibili è cioè non necessariamente uguale, ma uguale o *maggiore* della somma delle probabilità (limite superiore delle somme di termini in numero finito). ([25], p. 314)

Un altro punto che de Finetti non tocca, consapevolmente, in [25] è il concetto di probabilità condizionata, su cui ritornerà più avanti, nella fondamentale memoria [27], frutto di un ciclo di lezioni da lui tenute presso l'Istituto Poincaré di Parigi nel maggio del 1935, dove attinse da idee contenute in un altro lavoro del 1930 [22]. Si considerino due eventi  $E_1$  ed  $E_2$  e si vincoli la scommessa su  $E_1$  al realizzarsi di  $E_2$  in questo modo: se  $E_2$  non si verifica, allora la scommessa su  $E_1$  viene ritirata; se  $E_2$  si verifica, la scommessa ha luogo e lo scommettitore vince se si verifica  $E_1$  mentre perde in caso contrario. Se  $S$  è la posta in gioco, i guadagni corrispondenti saranno, rispettivamente, 0,  $(1-p)S$  e  $-pS$ , dove  $p$  è la probabilità che  $\mathcal{O}$  assegna al realizzarsi di  $E_1$ , sapendo che  $E_2$  si è realizzato. Dette ora  $p_1$  e  $p_2$  le probabilità di  $E_1 \cap E_2$  ed  $E_2$ , si scommettono anche le somme  $S_1$  ed  $S_2$  sul realizzarsi di questi eventi. Si avranno in definitiva questi guadagni:

- se  $E_2$  ed  $E_1$  si realizzano:  $G_1 = (1-p_1)S_1 + (1-p_2)S_2 + (1-p)S$
- se  $E_2$  si realizza ma non si realizza  $E_1$ :  $G_2 = -p_1S_1 + (1-p_2)S_2 - pS$
- se non si realizza  $E_2$ :  $G_3 = -p_1S_1 - p_2S_2$

dove, nell'ultimo caso, si è tenuto conto che solo la scommessa su  $E_1$  condizionato al realizzarsi di  $E_2$  viene annullata. Mettendo a sistema i tre guadagni ed interpretando le quote  $S$ ,  $S_1$  ed  $S_2$  come incognite, il determinante della matrice dei coefficienti è  $p_1 - pp_2$  per cui deve valere il teorema della probabilità condizionata, altrimenti verrebbe meno la coerenza. Il quadro è così completo perché le leggi fondamentali del calcolo delle probabilità sono state riottenute come conseguenza della condizione di coerenza. L'insieme di risultati che permettono di vedere, nei teoremi fondamentali del calcolo delle probabilità, una condizione necessaria e sufficiente alla coerenza di un individuo è noto come *teorema di Ramsey-De Finetti*, associando a De Finetti il nome di Frank Plumpton Ramsey (1903-1930) che aveva proposto un approccio simile in un lavoro [39] scritto nel

1926 ma pubblicato postumo nel 1931, di cui De Finetti venne a conoscenza solo dopo aver proposto il proprio schema. Ramsey non utilizzò somme di denaro nel suo schema di scommesse ma sviluppò una teoria dell'utilità nei termini della quale erano definite le scommesse. De Finetti si accostò ad un approccio più vicino a Ramsey in suoi lavori posteriori, giungendo a dichiarare che le scommesse non sono pertinenti alla probabilità ma alla teoria dei giochi. Osserviamo che mentre Ramsey giunse alla interpretazione soggettiva criticando le tesi di Keynes, De Finetti vi fu condotto dalla lettura del trattato di Emanuel Czuber dove le interpretazioni frequentista e soggettivista venivano presentate nel loro sviluppo storico.

## 10.5 L'idea di scambiabilità

Un concetto di fondamentale importanza per l'applicazione del punto di vista soggettivista alla statistica è quello di *scambiabilità* tra eventi, concetto grazie al quale si crea la possibilità di apprendere attraverso l'esperienza, gettando così un ponte verso l'approccio frequentistico. Il concetto di eventi scambiabili, legato al nome di De Finetti che ne parlò al Congresso Internazionale dei Matematici di Bologna nel 1928 e vi dedicò un importante lavoro pubblicato due anni più tardi [23], non fu però introdotto la prima volta da De Finetti ma, indipendentemente, dal logico e filosofo inglese William Ernest Johnson (1858-1931) e dal matematico francese Jules Haag (1882-1953). A De Finetti spetta il merito di aver dimostrato il teorema di rappresentazione che porta il suo nome [20], basato sull'idea di eventi scambiabili.

L'importanza di Johnson nel calcolo delle probabilità fu indiretta e diretta: indiretta perché ebbe tra i suoi studenti Keynes e Ramsey che si occuparono di probabilità e furono influenzati dalle sue idee; diretta, grazie alla prima formulazione del concetto di eventi scambiabili che fu enunciata dapprima in un'appendice (*On Education*) al terzo volume del suo trattato di Logica, pubblicato nel 1924 [33] e poi riformulata in un lavoro pubblicato postumo nel 1932 [34]. Johnson parte da una visione in cui il calcolo delle probabilità è indissolubilmente legato alla logica: ([33], p. 179)

La probabilità è una grandezza da legare ad ogni proposizione che può essere vera o falsa; ciò però non alla proposizione considerata in se stessa e per se stessa, ma in rapporto ad un'altra proposizione la verità della quale si suppone essere nota. Per esempio, il valore della probabilità della proposizione “Il prossimo lancio di una certa moneta darà testa” può essere attribuito in base alla conoscenza che essa “Darà testa o croce”. Il valore della probabilità determinato in questo modo non coincide necessariamente con quello determinato dalla conoscenza che “I lanci precedenti della moneta hanno presentato testa o croce con una certa frequenza”.<sup>24</sup>

---

<sup>24</sup>Probability is a magnitude to be attached to any possibly true or possibly false proposition; not, however, to the proposition in and for itself, but in reference to another proposition the truth of which is supposed to be known. For example, the probability of the proposition

Dunque ogni probabilità è *condizionata* dalle informazioni a disposizione di chi deve farne la valutazione. Inoltre ([33], p. 182)

gli assiomi della probabilità ci permettono di inferire una conclusione sulle probabilità solo a partire da premesse sulla probabilità. In altre parole, il calcolo delle probabilità non ci consente di trarre delle conclusioni sulle probabilità se non sono *assegnate* delle probabilità o relazioni di probabilità. Il matematico non può fornire questi dati.

Per esempio, le regole dell'aritmetica e gli assiomi del calcolo delle probabilità sono del tutto incapaci a determinare, a partire dall'ipotesi che una moneta, lanciata in aria, debba atterrare mostrando testa o croce, ma non entrambe le facce, qual è la probabilità che essa mostrerà testa ovvero che mostrerà croce.<sup>25</sup>

Per operare una scelta, vi sono due postulati possibili che Johnson considerò nel caso in cui vengano effettuate  $M$  prove, in ciascuna delle quali si possa presentare solo un evento, tra  $\alpha$  possibili,  $E_1, E_2, \dots, E_\alpha$ . Se  $m_i$  è il numero di prove in cui si è presentato  $E_i$ , si dovrà avere

$$M = m_1 + m_2 + \dots + m_\alpha \quad (10.9)$$

e Johnson chiama la *proporzione*  $m_1 : m_2 : \dots : m_\alpha$  una possibile soluzione di questa equazione. Il primo postulato, detto postulato di combinazione (*Combination-Postulate*), afferma che ogni possibile soluzione dell'equazione (10.9) ha la stessa probabilità di presentarsi, prima che si abbia qualunque informazione. Per chiarire le cose, consideriamo  $M = 4$  lanci di una moneta ed indichiamo con  $T$  e  $C$ , rispettivamente, la realizzazione di testa o croce in un lancio. In questo caso  $\alpha = 2$  e le coppie  $(m_1, m_2)$  che risolvono (10.9), dove la prima componente si riferisce al numero di teste ottenute nei quattro lanci, sono  $(4, 0)$ ,  $(3, 1)$ ,  $(2, 2)$ ,  $(1, 3)$  e  $(0, 4)$ . A ciascuna di esse viene associata, sulla base del postulato di combinazione di Johnson una probabilità

$$\frac{1}{\binom{5}{4}} = \frac{1}{5}$$

che dunque non tiene conto dei modi differenti in cui una particolare partizione può presentarsi.

---

that "The next throw of a certain coin will yield head" may have its value assigned by the knowledge that "It will yield either head or tail". The value of the probability as so determined is not necessarily the same as that determined by the knowledge that "The previous throws of the coin have presented heads and tails with a certain frequency."

<sup>25</sup>the axioms of probability enable us to infer any probability-conclusion only from probability-premisses. In other words, the calculus of probability does not enable us to infer any probability-value unless we have some probabilities or probability relations *given*. Such data cannot be supplied by the mathematician. E.g. the rules of arithmetic and the axioms of the probability-calculus are utterly impotent to determine, on the supposed knowledge that the throw of a coin must yield either head or tail and cannot yield both, the probability that it will yield head or that it will yield tail.

Come inciso, vediamo di calcolare in quanti modi si può risolvere l'equazione (10.9), ritenendo distinte soluzioni che corrispondono a distinti valori di  $(m_1, m_2, \dots, m_\alpha$  e supponendo per ora che tutti gli  $m_i > 0$ . Prendiamo ora  $M$  unità e disponiamole lungo un segmento. Per separarle in  $\alpha$  addendi dobbiamo introdurre  $\alpha - 1$  segni di addizione. È possibile collocare il primo segno  $+$  in  $M - 1$  modi diversi, il secondo in  $M - 2$  modi diversi e così via fino all'ultimo segno  $+$  che si può collocare in  $M - (\alpha - 1) = M - \alpha + 1$  modi diversi. L'ordine con cui sono stati introdotti i segni  $+$  è del tutto ininfluenza e quindi il numero di modi in cui esse possono essere poste a separare le  $M$  sfere è

$$\frac{(M-1)(M-2)\cdots(M-\alpha+1)}{(\alpha-1)!} = \frac{(M-1)(M-2)\cdots(M-\alpha+1)(M-\alpha)!}{(\alpha-1)!(M-\alpha)!}$$

ovvero

$$\binom{M-1}{\alpha-1}.$$

Per adattare la soluzione al nostro caso, in cui  $m_i \geq 0$  e non solo strettamente positivi, introduciamo i numeri  $y_i$  tali che

$$m_i = y_i - 1$$

cosicché tutte le variabili  $y_i$  sono positive. Sostituendo nell'equazione (10.9), otteniamo

$$y_1 - 1 + y_2 - 1 + \cdots + y_\alpha - 1 = M$$

cioè

$$y_1 + y_2 + \cdots + y_\alpha = M + \alpha.$$

e quindi il numero di soluzioni di (10.9) è

$$\binom{\alpha + M - 1}{M}$$

a cui compete la stessa probabilità. In un secondo momento, Johnson riformulò il postulato di combinazione e la nuova versione comparve in un lavoro postumo, pubblicato su una rivista filosofica nel 1932 [34]. In questa versione Johnson postulò che la probabilità che si presenti  $E_j$  in una prova successiva dipende solo dal numero di volte  $m_j$  in cui  $E_j$  si è presentato e dal numero  $M$  di prove effettuate ma non dipende dall'ordine con cui gli eventi si sono presentati nelle  $M$  prove né dai valori che assumono gli altri interi  $m_1, \dots, m_\alpha$  che compongono con  $m_j$  una soluzione dell'equazione (10.9). In questa forma, il postulato è noto come postulato di sufficienza.

Il secondo postulato, postulato di permutazione (*Permutation-Postulate*), afferma che la probabilità che compete ad una soluzione di (10.9) non dipende dall'*ordine* con cui i vari eventi si presentano nelle diverse prove.

In base al secondo postulato, la probabilità associata ad una permutazione qualsiasi che lasci inalterata la soluzione di (10.9), è

$$\frac{m_1!m_2!\cdots m_\alpha!}{M!}; \quad (10.10)$$

questa proprietà di simmetria definisce gli eventi *scambiabili*, anche se Johnson non gli attribuì un nome ma se ne servì per affrontare il problema dell'induzione, riottenendo la regola di successione di Laplace nel caso discreto.

Riprendendo l'esempio del lancio di una moneta, ripetuto 4 volte, le probabilità associate a *ciascuna* sequenza di quattro esiti in un ordine prestabilito, sapendo che si è realizzata una delle cinque tipologie di partizioni possibili, sono:

$$(4, 0) \equiv \{(T, T, T, T)\} \quad m_1 = 4 \quad m_2 = 0 \quad p = 1;$$

$$(0, 4) \equiv \{(C, C, C, C)\} \quad m_1 = 0 \quad m_2 = 4 \quad p = 1;$$

$$(3, 1) \equiv \{(T, T, T, C), (T, T, C, T), (T, C, T, T), (C, T, T, T)\}$$

$$m_1 = 3 \quad m_2 = 1 \quad p = \frac{3!1!}{4!} = \frac{1}{4};$$

$$(1, 3) \equiv \{(C, C, C, T), (C, C, T, C), (C, T, C, C), (T, C, C, C)\}$$

$$m_1 = 1 \quad m_2 = 3 \quad p = \frac{1!3!}{4!} = \frac{1}{4};$$

$$(2, 2) \equiv \{(C, C, T, T), (C, T, T, C), (C, T, C, T), (T, C, C, T), (T, T, C, C), (T, C, T, C)\}$$

$$m_1 = 2 \quad m_2 = 2 \quad p = \frac{2!2!}{4!} = \frac{1}{6};$$

Tornando a De Finetti, egli considerò eventi  $E_1, E_2, \dots, E_n$  che si riferiscono ad un esperimento o fenomeno che viene ripetuto in  $n$  prove, sotto le stesse condizioni. Se  $\omega_h^{(n)}$  è la probabilità assegnata al realizzarsi di  $h$  successi nelle  $n$  prove, indipendentemente dall'ordine in cui i successi si verificano, gli eventi  $E_1, E_2, \dots, E_n$  sono detti *equivalenti* [23] se la probabilità di ottenere una particolare sequenza di  $h$  successi in  $n$  prove è

$$\frac{\omega_h^{(n)}}{\binom{n}{h}} \quad (10.11)$$

per ciascuna delle  $\binom{n}{h}$  sequenze contenenti  $h$  successi. Successivamente De Finetti adottò la terminologia di eventi *scambiabili* [27], seguendo un suggerimento di Maurice Fréchet (1878-1973). Per De Finetti il concetto di eventi scambiabili non solo serve a difendere il valore dell'approccio soggettivista ma anche a proporre una soluzione del problema dell'induzione che, ai suoi occhi, costituiva la sostanza del valore filosofico della probabilità [26]. L'obiezione all'approccio soggettivista si può formulare su un esempio concreto, quello del lancio, ripetuto  $n$  volte, di una moneta. Supponendo che testa si sia presentata  $h$  volte, la frequenza di successi  $\frac{h}{n}$  che si può assumere come stima della probabilità che testa si presenti in un lancio, quando  $n \gg 1$ , è un fatto oggettivo: come è possibile farla rientrare nella descrizione soggettivista? La risposta di De Finetti è racchiusa nell'uso della probabilità condizionata. Si consideri come evento  $e$  quello che consiste nella realizzazione di  $h$  teste su  $n$  lanci:

$(e) = \{T_1, C_2, T_3 \cdots C_n\}$ : qual è la probabilità che il lancio  $n + 1$ -esimo mostri “testa” come esito? Detto  $T_{n+1}$  questo evento, dovrà essere

$$p(T_{n+1}|e) = \frac{p(T_{n+1} \cap e)}{p(e)}.$$

Indicando come prima con  $\omega_h^{(n)}$  la probabilità che un soggetto assegna al realizzarsi di una sequenza di  $h$  successi (teste) in una serie di  $n$  lanci, l’ipotesi di scambiabilità permette di concludere che è la probabilità dell’evento  $(e)$  è espressa dalla (10.11). D’altronde, l’evento  $T_{n+1} \cap e$  è una successione in cui sono comparse  $h + 1$  teste in  $n + 1$  tentativi e dunque gli compete una probabilità

$$p(T_{n+1} \cap e) = \frac{\omega_{h+1}^{(n+1)}}{\binom{n+1}{h+1}}.$$

Si ha così

$$p(T_{n+1}|e) = \frac{p(T_{n+1} \cap e)}{p(e)} = \frac{h+1}{n+1} \frac{\omega_{h+1}^{(n+1)}}{\omega_h^{(n)}} \quad (*)$$

che, nell’ipotesi in cui

$$\lim_{n \rightarrow \infty} \frac{\omega_{h+1}^{(n+1)}}{\omega_h^{(n)}} = 1, \quad (10.12)$$

ipotesi che sembra non molto restrittiva, visto il significato di  $\omega_h^{(n)}$ , diventa

$$\lim_{n \rightarrow \infty} p(T_{n+1}|e) = \frac{h+1}{n+1}$$

che mostra come, assegnando le probabilità a priori  $\omega_h^{(n)}$  in un modo coerente e tale da rispettare la condizione al limite (10.12), anche un soggettivista possa giungere alla stima “frequentista”  $\frac{h+1}{n+1}$ . Nel caso in cui si volesse applicare il principio di ragione insufficiente scegliendo

$$\omega_0^{(n)} = \omega_1^{(n)} = \cdots = \omega_n^{(n)} = \frac{1}{n+1}$$

e

$$\omega_0^{(n+1)} = \omega_1^{(n+1)} = \cdots = \omega_{n+1}^{(n+1)} = \frac{1}{n+2}$$

si otterrebbe

$$p(T_{n+1}|e) = \frac{h+1}{n+2}$$

che coincide con la regola di successione proposta da Laplace. A questo proposito, è stato osservato [GILLIES, 2000, p. 73] la correzione di de Finetti offre il fianco alla critica avanzata, in generale, alla regola di successione dall’epistemologo Karl Popper che, riallacciandosi al problema della probabilità che il Sole

sorga domani osservava che, limitandosi ai 5000 anni di storia per cui vi sono testimonianze, il Sole è sorto 1826250 volte in altrettanti giorni per cui, posto nella regola di successione questo valore sia per  $h$  che per  $n$  si otterrebbe la probabilità 0.99999994. Ora, osservò Popper in un esempio riportato in una versione preliminare di un lavoro pubblicato nel 1957, se ad un certo giorno il Sole non sorgesse, gettando nel panico tutti, ma non vi fossero altri fenomeni inconsueti, la formula dovrebbe applicarsi prendendo  $n = 1826250$  ed  $h = 1826249$ , dando luogo ad una probabilità che il Sole sorga il giorno successivo pari a 0.9999989 che, osservò Popper, è troppo alta, data l'eccezionalità dell'evento. In generale, in scritti successivi de Finetti difese le modifiche alla probabilità di un evento in base all'accumularsi di esperienze come dovute unicamente al condizionamento Bayesiano e non ad una modifica più radicale della probabilità di un evento dopo che il suo reiterarsi è stato smentito in modo evidente dall'esperienza.

## 10.6 I Grundbegriffe di Kolmogorov

Nel 1933, il matematico russo Andrei Nikolaevič Kolmogorov (1903-1987) pubblicò le sue ricerche sulla formulazione assiomatica del calcolo delle probabilità in un'opera—*Grundbegriffe der Wahrscheinlichkeitsrechnung* [36]—destinata a lasciare un segno indelebile negli sviluppi successivi della disciplina. Fin dalla premessa di quest'opera, Kolmogorov osserva che i principi della probabilità, giudicati come strani (*eigenartig*) anche in tempi recenti, dovessero essere inquadrati nel filone della matematica moderna. Accingersi a questa impresa senza la teoria dell'integrazione di Lebesgue avrebbe però reso disperato (*hoffnunglos*) questo tentativo. La fiducia riposta nell'uso della teoria di Lebesgue nasceva dalle *analogie* tra misura di un insieme e probabilità di un evento, tra integrale di una funzione ed aspettazione matematica di una variabile casuale e Kolmogorov intendeva portare avanti questa analogia, mostrandone appieno la portata. Per questo egli riteneva che la teoria proposta da Lebesgue andasse liberata dai suoi elementi geometrici, un processo completato in analisi da Fréchet. Inoltre, Kolmogorov sottolineava come l'utilizzo di strumenti matematici raffinati quali l'analisi in spazi funzionali infinito-dimensionali non fosse un lusso analitico ma sorgesse

necessariamente da chiare questioni inerenti concreti problemi fisici.<sup>26</sup>  
([36], p. iii)

Kolmogorov chiama *elementare* il calcolo delle probabilità che si occupa di un numero finito di eventi. Se si vuole perseguire l'obiettivo di fondare il calcolo delle probabilità alla stessa stregua della geometria o dell'algebra

i nomi degli oggetti da esaminare e le loro relazioni fondamentali, così come gli assiomi cui queste relazioni debbono obbedire, vanno assegnati in modo che ogni ulteriore rappresentazione sia basata esclusivamente su

<sup>26</sup>notwendigerweise aus einigen ganz konkreten physikalischen Fragestellung entstanden sind.

questi assiomi senza far riferimento ai rispettivi significati concreti che possono essere associati a questi oggetti e relazioni.<sup>27</sup> ([36], p. 1)

Kolmogorov qui fa semplicemente eco all'impostazione data da Hilbert al problema dei fondamenti della geometria e si sottolinea subito come una impostazione assiomatica lasci libero spazio ad un numero qualsivoglia di interpretazioni. Una assiomatizzazione della probabilità non è l'assiomatizzazione della probabilità e Kolmogorov evidenzia la differenza tra il proprio approccio, in cui vengono definiti assiomaticamente i concetti di evento casuale e probabilità, con l'approccio di von Mises, in cui i concetti fondamentali sono altri.

Kolmogorov considera un insieme  $E$  ed una famiglia  $\mathcal{F}$  di sottoinsiemi di  $E$  i cui elementi sono detti eventi casuali (*zufällige Ereignisse*). Gli assiomi cui  $\mathcal{F}$  deve obbedire sono i seguenti

1.  $\mathcal{F}$  è un *corpo di insiemi* (*Mengenkörper*); qui Kolmogorov utilizza il linguaggio di Felix Hausdorff (1868-1942) che nell'opera *Grundzüge der Mengenlehre* [31] aveva definito con questo nome un sistema di insiemi (*Mengensystem*) che contiene l'insieme vuoto e che è chiuso rispetto alle operazioni di intersezione e differenza.
2.  $E \in \mathcal{F}$ .
3. Ad ogni insieme  $A \in \mathcal{F}$  è associato un numero reale non negativo  $P(A)$ , detto la probabilità dell'evento  $A$ .
4.  $P(E) = 1$
5. Se  $A \cap B = \emptyset$  allora  $P(A \cup B) = P(A) + P(B)$ .

Una famiglia di insiemi che ubbidisca a questi requisiti è detta un *campo di probabilità* (*Wahrscheinlichkeitsfeld*). Kolmogorov costruì un esempio elementare di campo di probabilità prendendo come  $E$  l'insieme composto da un numero  $k$  finito di elementi

$$E := \{\xi_1, \xi_2, \dots, \xi_k\}$$

cui si associano  $k$  numeri  $p_1, p_2, \dots, p_k$  non negativi tali che

$$\sum_{i=1}^k p_i = 1$$

e  $\mathcal{F}$  è la famiglia di tutti i sottoinsiemi di  $\mathcal{F}$ . Gli eventi  $\xi_1, \xi_2, \dots, \xi_n$  sono eventi elementari e la probabilità di un evento composto è definita in questo modo

$$P(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_\lambda}) = P(\xi_{i_1}) + P(\xi_{i_2}) + \dots + P(\xi_{i_\lambda}) :$$

il campo di probabilità così ottenuto è il prototipo di campo di probabilità *finito*.

<sup>27</sup>die Namen der zu untersuchenden Gegenstände und ihrer Grundbeziehungen sowie die Axiome, denen diese Grundbeziehungen zu gehorchen haben, angegeben sind, die ganze weitere Darstellung sich ausschließlich auf diese Axiome gründen soll und keine Rücksicht auf die jeweilige konkrete Bedeutung dieser Gegenstände und Beziehungen nehmen darf.

Allontanandosi dallo sviluppo assiomatico della teoria, Kolmogorov si occupa del rapporto con le applicazioni, ovvero delle relazioni con il mondo dell'esperienza (*Erfahrungswelt*). Questa applicazione avviene secondo uno schema ternario:

- si considera un certo complesso  $\mathfrak{S}$  di condizioni che è possibile ripetere senza alcuna limitazione.
- Si individua una cerchia di eventi che possono aver luogo in seguito alla realizzazione delle condizioni  $\mathfrak{S}$ . Nei singoli casi la realizzazione delle condizioni  $\mathfrak{S}$  porta gli eventi a presentarsi in ordine diverso:  $E$  è l'insieme di tutte le varianti possibili, anche se queste possono non realizzarsi nella pratica.
- Quando la realizzazione delle condizioni  $\mathfrak{S}$  è tale che la variante considerata appartenga ad un insieme  $A$ , si dice che l'evento  $A$  si è realizzato.

Come esempio delle condizioni  $\mathfrak{S}$ , Kolmogorov considera il lancio di una moneta ripetuto due volte; gli eventi di cui al punto 2 sono il presentarsi, in ogni singolo lancio, di testa ( $T$ ) o croce ( $C$ ) ed i quattro eventi elementari sono:  $(TT)$ ,  $(TC)$ ,  $(CT)$  e  $(CC)$ . Come evento  $A$  si può considerare la ripetizione di uno stesso risultato nei due lanci:  $(TT)$  o  $(CC)$ .

È poi possibile associare ad  $A$  un numero reale  $P(A)$  con queste proprietà:

Si può essere praticamente sicuri che, quando il complesso di condizioni  $\mathfrak{S}$  viene ripetuto un grande numero  $n$  di volte e si nota che l'evento  $A$  si è presentato un numero  $m$  di volte, il rapporto  $m/n$  differisce di poco da  $P(A)$ .<sup>28</sup> ([36], p. 4)

Inoltre, si riaffaccia il principio di Cournot:

Se  $P(A)$  è molto piccolo, si può essere praticamente certi che in una singola realizzazione delle condizioni  $\mathfrak{S}$  l'evento  $A$  non avrà luogo.<sup>29</sup> ([36], p. 4)

Ciò non significa affatto che, ripetendo più volte la serie di  $n$  prove si ottenga *sempre* un risultato prossimo a  $P(A)$ . Accettate queste considerazioni, per Kolmogorov è possibile una deduzione empirica degli assiomi, supponendo che  $\mathcal{F}$  sia un corpo di insiemi contenenti  $E$ ; inoltre il rapporto  $\frac{m}{n} \in [0, 1]$  verifica l'assioma 3 e, chiaramente, per  $E$  si ha  $m = n$  e dunque  $P(E) = 1$ , cosicché anche l'assioma 4. è obbedito; se poi si considerano due insiemi disgiunti  $A$  e  $B$ , il primo dei quali si verifica  $m_1$  volte ed il secondo  $m_2$  volte, il numero di volte  $m$  in cui si verifica  $A \cup B$  sarà ovviamente  $m = m_1 + m_2$  e dunque

$$\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n} :$$

<sup>28</sup>Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen  $\mathfrak{S}$  eine große Anzahl von  $n$  Malen wiederholt und dabei durch  $m$  die Anzahl der Fälle bezeichnet, bei denen das Ereignis  $A$  stattgefunden hat, das Verhältnis  $m/n$  sich von  $P(A)$  nur wenig unterscheidet.

<sup>29</sup>Ist  $P(A)$  sehr klein, so kann man praktisch sicher sein, daß bei einer einmaligen Realisation der Bedingungen  $\mathfrak{S}$  das Ereignis  $A$  nicht stattfindet.

Sembra dunque appropriato porre  $P(A + B) = P(A) + P(B)$ .<sup>30</sup> ([36], p. 4)

Kolmogorov, riprendendo l'impostazione puramente matematica, deduce le principali proprietà della probabilità a partire dagli assiomi:

- Se  $\bar{A}$  denota il complementare di  $A$ , allora

$$P(\bar{A}) = 1 - P(A);$$

- $P(\emptyset) = 0$ ;
- Se  $A_1, A_2, \dots, A_N$  sono disgiunti allora vale il teorema di addizione

$$P\left(\bigcup_{i=1}^N A_i\right) = \sum_{i=1}^n P(A_i);$$

- Se  $P(A) > 0$  allora si può introdurre la probabilità condizionata  $P_A(B)$  da  $A$  di un evento  $B$  come

$$P_A(B) = [P(B|A)] := \frac{P(A \cap B)}{P(A)}.$$

- La probabilità condizionata soddisfa gli assiomi 3, 4 e 5 introdotti in precedenza e dunque l'insieme  $\mathcal{F}$  con la funzione di insieme (*Mengenfunktion*)  $P_A(B)$  è un campo di probabilità.

A queste considerazioni fanno seguito, tra le altre cose, la dimostrazione della formula di Bayes e la distinzione tra eventi indipendenti ed indipendenti a coppie.

Nel secondo capitolo di [36], Kolmogorov integra l'insieme di assiomi per coprire il caso in cui i campi di probabilità siano *infiniti*, introducendo l'assioma di continuità. Data una successione di insiemi  $\{A_i\} \in \mathcal{F}$  tali che

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots \quad (10.13)$$

e supponendo che

$$\bigcap_{n=1}^{\infty} A_n = \emptyset, \quad (10.14)$$

allora

$$\lim_{n \rightarrow \infty} P(A_n) = 0 : \quad (10.15)$$

un campo di probabilità che soddisfi anche questo assioma è detto campo di probabilità in senso esteso (*im erweiterten Sinne*). La necessità dell'estensione è dovuta alla volontà di considerare campi infiniti di probabilità, perché altrimenti il

<sup>30</sup>Es erscheint also als angebracht,  $P(A + B) = P(A) + P(B)$  zu setzen. Ho mantenuto qui la notazione originale di Kolmogorov in cui si indica con  $E + F$  la somma logica di eventi incompatibili.

nuovo assioma sarebbe conseguenza dei cinque posti in precedenza. Kolmogorov sottolinea come i campi infiniti forniscano uno schema idealizzato dei processi casuali reali. La conseguenza immediata dell'assioma di continuità è la formulazione estesa del teorema di addizione: Si suppongano  $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$ , disgiunti tra loro e sia

$$A := \bigcup_{n=1}^{\infty} A_n.$$

Allora

$$P(A) = \sum_{n=1}^{\infty} P(A_n).$$

Infatti, per l'insieme

$$R_n = \bigcup_{m=n+1}^{\infty} A_m,$$

sono soddisfatte le condizioni (10.13) e (10.14) e dunque, per l'assioma di continuità

$$\lim_{n \rightarrow \infty} P(R_n) = 0;$$

d'altra parte, grazie alla proprietà di additività *finita*, segue

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_n) + P(R_n)$$

e quindi, al limite, si ottiene la forma estesa del teorema di addizione. Tra i campi infiniti di probabilità, un ruolo privilegiato è svolto dai campi di probabilità di Borel, in cui il corpo di insiemi  $\mathcal{F}$  è un corpo boreliano, nel senso che è chiuso rispetto all'unione finita o numerabile di insiemi disgiunti appartenenti ad  $\mathcal{F}$ . L'importanza di questi campi viene così espressa da Kolmogorov

Solo nei campi di probabilità di BOREL il calcolo delle probabilità raggiunge quella libertà di azione che non è legata al pericolo di arrivare ad eventi per i quali non sia assegnabile una probabilità.<sup>31</sup> ([36], p. 15)

I *Grundbegriffe* si snodano in sei capitoli e qui abbiamo dato un cenno soltanto al contenuto dei primi due. In particolare l'approccio alla probabilità condizionata, esteso a campi di probabilità infiniti, presentato nel Capitolo 5 risulta originale per l'epoca, anche se non fu generalmente notato dai contemporanei [41]. In generale, le reazioni al lavoro di Kolmogorov furono molto varie, passando da un'adesione immediata ed incondizionata (Feller, Doob, Cramér), ad un certo distacco (Cantelli).

<sup>31</sup>Nur in BORELSchen Wahrscheinlichkeitsfeldern erhält die Wahrscheinlichkeitsrechnung eine vollständige Handlungsfreiheit, die mit keiner Gefahr verbunden ist, zu Ereignissen zu gelangen, welche keine Wahrscheinlichkeit besitzen.

## 10.7 La logica del probabile

Lo studio dei legami tra logica e probabilità si può dire risalga alle stesse origini del calcolo delle probabilità come disciplina scientifica, all'*Ars Cogitandi*, la logica di Port Royal, e a Leibniz (1646-1716) che aveva scritto, nei *Nouveau essais*, pubblicati postumi nel 1765,

Ho detto più volte che c'è bisogno di un nuovo tipo di logica che tratti dei gradi di probabilità.<sup>32</sup> (*cfr.* [28], p. 2)

Verso la fine del '700, Louis-Frédéric Ancillon (1740-1814), un filosofo e pastore protestante tedesco di origini francesi, pubblicò tra le memorie dell'Accademia delle Scienze di Berlino un lavoro in cui, tra l'altro, asseriva, dopo aver distinto tra probabilità oggettiva e soggettiva:

Dire che un fatto passato, presente o futuro è probabile, significa dire che una proposizione è probabile.<sup>33</sup> (*cfr.* [28], p. 4)

Per Ancillon, una proposizione è probabile quando tale risulta il legame tra soggetto e predicato. Queste affermazioni di Ancillon vanno inquadrare in un intervento critico sulla possibilità stessa di costruire una teoria matematica della probabilità. Mezzo secolo dopo, il matematico boemo Bernard Bolzano (1781-1848) parlava, nell'opera *Wissenschaftslehre*, pubblicata nel 1837 ma scritta nel decennio precedente, della probabilità come di un *grado di validità* di proposizioni in cui vengono formulate delle ipotesi, riferite ad altre proposizioni che esprimono un ventaglio di possibilità. A dispetto di ciò, Bolzano mantenne l'idea che la probabilità, come la verità da cui discende, fosse una nozione oggettiva. Per Bolzano, quando un termine (*idea*) di una proposizione è variabile, è possibile studiare la validità della proposizione in funzione di questi cambiamenti. Se si considera un certo insieme di proposizioni, la *validità (Gültigkeit)* di una proposizione è la relazione che passa tra il numero di proposizioni ottenibili al cambiare del termine selezionato ed il numero di proposizioni che verificano quella in esame. Il grado di validità, rispetto all'idea selezionata, di una proposizione è il rapporto tra questi due numeri.

Riflessioni più sistematiche sul rapporto tra logica e probabilità si svilupparono in Inghilterra a partire da alcuni lavori, apparsi attorno alla metà del XIX secolo, di Augustus De Morgan (1806-1871), George Boole (1815-1864) e, una ventina d'anni dopo, di John Venn e William Stanley Jevons (1835-1882), per arrivare al trattato di Keynes, pubblicato nel 1921, e a quello di Johnson del 1924. Infine una figura di spicco da ricordare è quella di Harold Jeffreys (1891-1989) che pubblicò nel 1939 la prima edizione di un trattato di probabilità di stampo bayesiano.

In questa sezione conclusiva esponiamo i tratti essenziali della formalizzazione della probabilità presentata nel 1946 dal fisico statunitense Richard Threlkeld

<sup>32</sup>J'ai dit plus d'une fois qu'il faudrait une nouvelle espèce de logique, qui traiteroit des degrés de Probabilité.

<sup>33</sup>Dire qu'un fait passé, présent ou à venir est probable, c'est dire qu'une proposition est probable.

Cox (1898-1991) in un lavoro [18] ripreso ed ampliato in una monografia pubblicata nel 1961 [19]. Questo approccio rappresenta una sintesi critica ed una semplificazione originale dei lavori di Keynes e Jeffreys.

Dopo aver evidenziato alcuni punti deboli dell'impostazione frequentista della probabilità, come lo spettro ridotto di problemi cui essa è applicabile, Cox riprese le idee di Keynes, secondo cui la probabilità è una relazione tra un'ipotesi ed una conclusione, entrambe espresse da una proposizione, che corrisponde al grado di assenso razionale offerto da un individuo. Il fatto che le proposizioni coinvolte non siano a priori vere o false richiede l'elaborazione di una logica estesa, poggiata su principi che completino il quadro formale della logica binaria. Il punto, che secondo Cox, Keynes non aveva risolto, era quello di giustificare le regole di inferenza necessarie per avviare la teoria, limitandosi ad una formulazione assiomatica in cui però, è sempre il giudizio di Cox, gli assiomi erano troppo arbitrari e sofisticati, una critica valida a suo dire anche per l'approccio di Jeffreys. L'obiettivo di Cox è quello di arricchire le regole della logica formale con due ulteriori nozioni primitive per dedurre i principi del calcolo delle probabilità. Indichiamo con  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ ,... delle proposizioni; con  $\neg\mathbf{a}$  la negazione della proposizione  $\mathbf{a}$ , con  $\mathbf{a} \vee \mathbf{b}$  la disgiunzione delle proposizioni  $\mathbf{a}$  e  $\mathbf{b}$ , che dunque è vera se almeno una di esse è vera, con  $\mathbf{a} \cdot \mathbf{b}$  il prodotto logico o congiunzione di  $\mathbf{a}$  e  $\mathbf{b}$  che è vero se entrambe le proposizioni  $\mathbf{a}$  e  $\mathbf{b}$  sono vere. Cox indica con  $\mathbf{b}|\mathbf{a}$  una misura (numerica)

della credibilità ragionevole della proposizione  $\mathbf{b}$  quando è noto che la proposizione  $\mathbf{a}$  è vera.<sup>34</sup> ([18], p. 6)

la funzione  $\mathbf{b}|\mathbf{a}$ , quando esiste, è indicata come verosimiglianza (*likelihood*) di  $\mathbf{b}$  a partire dall'ipotesi  $\mathbf{a}$  e Cox sottolinea come una qualunque funzione  $f(\mathbf{b}|\mathbf{a})$  possa servire da misura della credibilità di  $\mathbf{b}$  data  $\mathbf{a}$ . La prima ipotesi per formulare i principi dell'inferenza probabile riguarda la verosimiglianza  $\mathbf{c} \cdot \mathbf{b}|\mathbf{a}$  della congiunzione  $\mathbf{c} \cdot \mathbf{b}$ , sotto l'ipotesi  $\mathbf{a}$  e consiste nel supporre

$$\mathbf{c} \cdot \mathbf{b}|\mathbf{a} = F(\mathbf{c}|\mathbf{b} \cdot \mathbf{a}, \mathbf{b}|\mathbf{a}) \quad (10.16)$$

dove  $F$  è una funzione di due variabili. La ragionevolezza di questa ipotesi è chiarita con un esempio in cui  $\mathbf{a}$  è l'insieme delle informazioni su un percorso che un podista deve percorrere all'andata ed al ritorno, senza fermarsi;  $\mathbf{b}$  è la proposizione: "il podista riesce a percorrere il percorso all'andata" e  $\mathbf{c}$  è la proposizione: "il podista riesce a percorrere il percorso di ritorno, senza fermarsi":

nello stimare [che] la verosimiglianza  $\mathbf{c} \cdot \mathbf{b}|\mathbf{a}$  (...) sia una funzione delle altre due [verosimiglianze], noi stiamo facendo l'ipotesi meno restrittiva possibile.<sup>35</sup> ([18], p. 6)

<sup>34</sup>of the reasonable credibility of the proposition  $\mathbf{b}$  when the proposition  $\mathbf{a}$  is known to be true.

<sup>35</sup>in estimating [that] the likelihood,  $\mathbf{c} \cdot \mathbf{b}|\mathbf{a}$ , (...) is some function of the other two, we are making the least restrictive assumption.

Se da un lato la funzione  $F$  eredita dalla verosimiglianza un grado di arbitrarietà, essa non può essere assegnata ad arbitrio ma le regole dell'algebra proposizionale impongono alcune restrizioni che conducono ad una equazione funzionale per  $F$ . Poiché  $\mathbf{d} \cdot \mathbf{c} \cdot \mathbf{b} = (\mathbf{d} \cdot \mathbf{c}) \cdot \mathbf{b}$ , abbiamo, grazie a (10.16)

$$\mathbf{d} \cdot \mathbf{c} \cdot \mathbf{b}|\mathbf{a} = (\mathbf{d} \cdot \mathbf{c}) \cdot \mathbf{b}|\mathbf{a} = F(\mathbf{d} \cdot \mathbf{c}|\mathbf{b} \cdot \mathbf{a}, \mathbf{b}|\mathbf{a}) \quad (10.17)$$

e

$$\mathbf{d} \cdot \mathbf{c}|\mathbf{b} \cdot \mathbf{a} = F(\mathbf{d}|\mathbf{c} \cdot (\mathbf{b} \cdot \mathbf{a}), \mathbf{c}|\mathbf{b} \cdot \mathbf{a}) = F(\mathbf{d}|\mathbf{c} \cdot \mathbf{b} \cdot \mathbf{a}, \mathbf{c}|\mathbf{b} \cdot \mathbf{a})$$

che, sostituito nell'equazione (10.17), fornisce

$$\mathbf{d} \cdot \mathbf{c} \cdot \mathbf{b}|\mathbf{a} = F[F(\mathbf{d}|\mathbf{c} \cdot \mathbf{b} \cdot \mathbf{a}, \mathbf{c}|\mathbf{b} \cdot \mathbf{a}), \mathbf{b}|\mathbf{a}]. \quad (10.18)$$

Similmente,

$$\mathbf{d} \cdot \mathbf{c} \cdot \mathbf{b}|\mathbf{a} = \mathbf{d} \cdot (\mathbf{c} \cdot \mathbf{b})|\mathbf{a} = F(\mathbf{d}|\mathbf{c} \cdot \mathbf{b} \cdot \mathbf{a}, \mathbf{c} \cdot \mathbf{b}|\mathbf{a}) = F[\mathbf{d}|\mathbf{c} \cdot \mathbf{b} \cdot \mathbf{a}, F(\mathbf{c}|\mathbf{b} \cdot \mathbf{a}, \mathbf{b}|\mathbf{a})]. \quad (10.19)$$

Uguagliando le due espressioni (10.18) e (10.19) e ponendo per semplicità  $x := \mathbf{d}|\mathbf{c} \cdot \mathbf{b} \cdot \mathbf{a}$ ,  $y := \mathbf{c}|\mathbf{b} \cdot \mathbf{a}$  e  $z := \mathbf{b}|\mathbf{a}$ , otteniamo l'equazione funzionale cui deve obbedire la funzione  $F$ :

$$F[F(x, y), z] = F[x, F(y, z)]. \quad (10.20)$$

Come dimostrato nella appendice, la soluzione generale di questa equazione è della forma

$$Cg[F(p, q)] = g(p)g(q)$$

dove  $C$  è una costante,  $p$  e  $q$  sono due argomenti qualsiasi e  $g$  una arbitraria funzione di una variabile: ricordando l'equazione (10.16), abbiamo

$$Cg(\mathbf{c} \cdot \mathbf{b}|\mathbf{a}) = g(\mathbf{c}|\mathbf{b} \cdot \mathbf{a})g(\mathbf{b}|\mathbf{a}). \quad (10.21)$$

Poiché è possibile porre  $C = 1$  e scegliere  $g$  come la funzione identica, vediamo che le verosimiglianze soddisfano la relazione

$$(\mathbf{c} \cdot \mathbf{b}|\mathbf{a}) = (\mathbf{c}|\mathbf{b} \cdot \mathbf{a})(\mathbf{b}|\mathbf{a}),$$

che riproduce la struttura del teorema delle probabilità composte.

Per ritrovare il teorema delle probabilità totali, Cox ricorre ad una seconda ipotesi sui processi di inferenza a partire dall'osservazione che la negazione  $\neg \mathbf{b}$  della proposizione  $\mathbf{b}$  è individuata quando  $\mathbf{b}$  lo è. Appare ragionevole supporre che

$$\neg \mathbf{b}|\mathbf{a} = S(\mathbf{b}|\mathbf{a}) \quad (10.22)$$

per una opportuna funzione  $S$ . Utilizzando ancora le regole della logica booleana è possibile dimostrare che questa funzione soddisfa l'equazione

$$xS\left[\frac{S(y)}{x}\right] = yS\left[\frac{S(x)}{y}\right]$$

che ha per soluzione generale  $S(p) = (1-p)^m$ , dove  $m$  è una costante qualsiasi, cui si può attribuire il valore  $m = 1$ , da cui si ricava che

$$\neg \mathbf{b} | \mathbf{a} = 1 - \mathbf{b} | \mathbf{a}$$

e, usando ancora le regole della logica di Boole,

$$\mathbf{c} \vee \mathbf{b} | \mathbf{a} = \mathbf{c} | \mathbf{a} + \mathbf{b} | \mathbf{a} - \mathbf{c} \cdot \mathbf{b} | \mathbf{a}$$

che corrisponde al teorema delle probabilità totali, comprendendo il caso in cui  $\mathbf{b}$  e  $\mathbf{c}$  non siano incompatibili.

## 10.8 Appendice

Per completezza, mostriamo la soluzione dell'equazione funzionale (10.20), riportata qui sotto

$$F[F(x, y), z] = F[x, F(y, z)], \quad (10.23)$$

seguendo l'esposizione di [4], in cui si suppone che la funzione  $F$  sia almeno di classe  $C^2$  nei suoi argomenti. Ponendo

$$u := F(x, y) \quad v := F(y, z), \quad (10.24)$$

l'equazione (10.23) diventa

$$F(u, z) = F(x, v). \quad (10.25)$$

Per semplificare la notazione, poniamo

$$F_1(p, q) = \frac{\partial F(p, q)}{\partial p} \quad F_2(p, q) = \frac{\partial F(p, q)}{\partial q}.$$

Se deriviamo (10.25), rispetto alle variabili indipendenti  $x, y$  e  $z$ , rispettivamente, otteniamo

$$F_1(u, z) \frac{\partial u}{\partial x} = F_1(x, v), \quad (10.26)$$

$$F_1(u, z) \frac{\partial u}{\partial y} = F_2(x, v) \frac{\partial v}{\partial y} \quad (10.27)$$

e

$$F_2(u, z) = F_2(x, v) \frac{\partial v}{\partial z}. \quad (10.28)$$

Poniamo ora

$$F_{11}(p, q) = \frac{\partial^2 F(p, q)}{\partial p^2} \quad F_{12}(p, q) = \frac{\partial^2 F(p, q)}{\partial p \partial q} \quad F_{22}(p, q) = \frac{\partial^2 F(p, q)}{\partial q^2}$$

e deriviamo l'equazione (10.27) rispetto ad  $x, y$  e  $z$ , nell'ordine. Avremo così:

$$F_{11}(u, z) \frac{\partial u}{\partial x} \frac{\partial u}{\partial y} + F_1(u, z) \frac{\partial^2 u}{\partial x \partial y} = F_{12}(x, v) \frac{\partial v}{\partial y}, \quad (10.29)$$

$$F_{11}(u, z) \left( \frac{\partial u}{\partial y} \right)^2 + F_1(u, z) \frac{\partial^2 u}{\partial y^2} = F_{22}(x, v) \left( \frac{\partial v}{\partial y} \right)^2 + F_2(x, v) \frac{\partial^2 v}{\partial y^2} \quad (10.30)$$

e

$$F_{12}(u, z) \frac{\partial u}{\partial y} = F_{22}(x, v) \frac{\partial v}{\partial y} \frac{\partial v}{\partial z} + F_2(x, v) \frac{\partial^2 v}{\partial y \partial z}. \quad (10.31)$$

Procediamo derivando l'equazione (10.26) rispetto a  $z$ . Arriviamo così a

$$F_{12}(u, z) \frac{\partial u}{\partial x} = F_{12}(x, v) \frac{\partial v}{\partial z}. \quad (10.32)$$

Nei prossimi passaggi, elimineremo gradualmente traccia delle derivate seconde di  $F$ . Moltiplichiamo (10.29) per  $\frac{\partial v}{\partial z}$  e (10.32) per  $\frac{\partial v}{\partial y}$ , ottenendo

$$F_{11}(u, z) \frac{\partial u}{\partial x} \frac{\partial u}{\partial y} \frac{\partial v}{\partial z} + F_1(u, z) \frac{\partial^2 u}{\partial x \partial y} \frac{\partial v}{\partial z} = F_{12}(u, z) \frac{\partial u}{\partial x} \frac{\partial v}{\partial y}. \quad (10.33)$$

Moltiplichiamo ora (10.31) per  $\frac{\partial u}{\partial x} \frac{\partial v}{\partial y}$  e sottraiamo al risultato l'equazione (10.33), moltiplicata a sua volta per  $\frac{\partial v}{\partial y}$ , così da eliminare  $F_{12}(u, z)$ :

$$\begin{aligned} & \left[ F_{11}(u, z) \left( \frac{\partial u}{\partial y} \right)^2 - F_{22}(x, v) \left( \frac{\partial v}{\partial y} \right)^2 \right] \frac{\partial u}{\partial x} \frac{\partial v}{\partial z} = F_2(x, v) \left( \frac{\partial^2 v}{\partial y \partial z} \right) \frac{\partial v}{\partial y} \frac{\partial u}{\partial x} - \\ & - F_1(u, z) \left( \frac{\partial^2 u}{\partial x \partial y} \right) \frac{\partial u}{\partial y} \frac{\partial v}{\partial z}. \end{aligned} \quad (10.34)$$

Un confronto di questa equazione con la (10.30) permette di eliminare sia  $F_{11}(u, z)$  che  $F_{22}(x, v)$ , ottenendo

$$\begin{aligned} & F_1(u, z) \left( \frac{\partial v}{\partial z} \right) \left[ \left( \frac{\partial^2 u}{\partial y^2} \right) \left( \frac{\partial u}{\partial x} \right) - \left( \frac{\partial^2 u}{\partial x \partial y} \right) \left( \frac{\partial u}{\partial y} \right) \right] = \\ & = F_2(x, v) \left( \frac{\partial u}{\partial x} \right) \left[ \left( \frac{\partial^2 v}{\partial y^2} \right) \left( \frac{\partial v}{\partial z} \right) - \left( \frac{\partial^2 v}{\partial y \partial z} \right) \left( \frac{\partial v}{\partial y} \right) \right] \end{aligned} \quad (10.35)$$

che, combinata a sua volta con (10.27), scritta nella forma

$$F_1(u, z) = F_2(x, v) \frac{\frac{\partial v}{\partial y}}{\frac{\partial u}{\partial y}},$$

permette di eliminare sia  $F_1(u, z)$  che  $F_2(x, v)$  e ricavare

$$\frac{\frac{\partial^2 u}{\partial x \partial y}}{\frac{\partial u}{\partial x}} - \frac{\frac{\partial^2 u}{\partial y^2}}{\frac{\partial u}{\partial y}} = \frac{\frac{\partial^2 v}{\partial y \partial z}}{\frac{\partial v}{\partial z}} - \frac{\frac{\partial^2 v}{\partial y^2}}{\frac{\partial v}{\partial y}}. \quad (10.36)$$

Ora, una semplice verifica permette di rendersi conto che

$$\frac{\partial}{\partial y} \ln \left( \frac{\frac{\partial u}{\partial x}}{\frac{\partial u}{\partial y}} \right) = \frac{\partial}{\partial y} \ln \frac{\partial u}{\partial x} - \frac{\partial}{\partial y} \ln \frac{\partial u}{\partial y} = \frac{\frac{\partial^2 u}{\partial x \partial y}}{\frac{\partial u}{\partial x}} - \frac{\frac{\partial^2 u}{\partial y^2}}{\frac{\partial u}{\partial y}}$$

e che

$$-\frac{\partial}{\partial y} \ln \left( \frac{\frac{\partial v}{\partial y}}{\frac{\partial v}{\partial z}} \right) = \frac{\frac{\partial^2 v}{\partial y \partial z}}{\frac{\partial v}{\partial z}} - \frac{\frac{\partial^2 v}{\partial y^2}}{\frac{\partial v}{\partial y}}$$

per cui (10.36) si può riscrivere nella forma

$$\frac{\partial}{\partial y} \ln \left( \frac{\frac{\partial u}{\partial x}}{\frac{\partial u}{\partial y}} \right) = -\frac{\partial}{\partial y} \ln \left( \frac{\frac{\partial v}{\partial y}}{\frac{\partial v}{\partial z}} \right). \quad (10.37)$$

Se ricordiamo il significato di  $u$  e  $v$  dato dalle definizioni (10.24), possiamo scrivere

$$\frac{\frac{\partial u}{\partial x}}{\frac{\partial u}{\partial y}} = \frac{F_1(x, y)}{F_2(x, y)} \quad \frac{\frac{\partial v}{\partial y}}{\frac{\partial v}{\partial z}} = \frac{F_1(y, z)}{F_2(y, z)}$$

cosicché la (10.37) diventa

$$\frac{\partial}{\partial y} \ln \left[ \frac{F_1(x, y)}{F_2(x, y)} \right] = -\frac{\partial}{\partial y} \ln \left[ \frac{F_1(y, z)}{F_2(y, z)} \right] :$$

poiché solo il membro di sinistra contiene  $x$  e solo quello di destra contiene  $z$ , ambo i membri dovranno essere uguali ad una stessa funzione  $\Phi(y)$ , per cui

$$\frac{\partial}{\partial y} \ln \left[ \frac{F_1(x, y)}{F_2(x, y)} \right] = \frac{d}{dy} \ln \Phi(y) \quad \frac{\partial}{\partial y} \ln \left[ \frac{F_1(y, z)}{F_2(y, z)} \right] = -\frac{d}{dy} \ln \Phi(y). \quad (10.38)$$

Permutiamo gli argomenti  $y \rightarrow x$  e  $z \rightarrow y$  nell'ultima equazione, ottenendo

$$\frac{\partial}{\partial x} \ln \left[ \frac{F_1(x, y)}{F_2(x, y)} \right] = -\frac{d}{dx} \ln \Phi(x)$$

che, combinata con la prima delle equazioni (10.38), mostra che

$$\frac{\partial}{\partial x} \ln \left[ \frac{F_1(x, y)}{F_2(x, y)} \right] dx + \frac{\partial}{\partial y} \ln \left[ \frac{F_1(x, y)}{F_2(x, y)} \right] dy = d \ln \frac{\Phi(y)}{\Phi(x)}$$

e quindi

$$\left[ \frac{F_1(x, y)}{F_2(x, y)} \right] = h \frac{\Phi(y)}{\Phi(x)}, \quad (10.39)$$

dove  $h$  è una costante di integrazione. Ora, dividendo (10.28) per (10.27) si ricava

$$\frac{F_1(u, z)}{F_2(u, z)} \frac{\partial u}{\partial y} = \frac{\frac{\partial v}{\partial y}}{\frac{\partial v}{\partial z}} = \frac{F_1(y, z)}{F_2(y, z)}$$

che, grazie a (10.39), si può porre nella forma

$$\frac{\Phi(z)}{\Phi(u)} \frac{\partial u}{\partial y} = \frac{\Phi(z)}{\Phi(y)}$$

ovvero, ricordando il significato di  $u$  dato dalla (10.24),

$$\frac{\partial F(x, y)}{\partial y} = \frac{\Phi[F(x, y)]}{\Phi(y)}.$$

Operando allo stesso modo sulle equazioni (10.26) e (10.27), ricaviamo anche

$$\frac{\partial F(y, z)}{\partial y} = \frac{\Phi[F(y, z)]}{\Phi(y)}$$

ovvero, permutando gli argomenti  $y \rightarrow x$  e  $z \rightarrow y$ , diventa

$$\frac{\partial F(x, y)}{\partial x} = \frac{\Phi[F(x, y)]}{\Phi(x)}.$$

Dunque

$$dF(x, y) = \frac{\partial F(x, y)}{\partial x} dx + \frac{\partial F(x, y)}{\partial y} dy = \Phi[F(x, y)] \left( \frac{dx}{\Phi(x)} + \frac{dy}{\Phi(y)} \right)$$

e, se si pone

$$\ln g(p) := \int \frac{dp}{\Phi(p)},$$

si ottiene, introducendo una costante di integrazione  $C$

$$Cg(F(x, y)) = g(x)g(y).$$

# Bibliografia

- [1] S. Banach: Sur le problème de la mesure. *Fundamenta Mathematicae*, **4**, 7-33, (1923).
- [2] J. Barone, A. Novikoff: A History of the Axiomatic Formulation of Probability from Borel to Kolmogorov: Part I. *Archive for History of Exact Sciences*, **18**, 123-190, (1978).
- [3] J. Bertrand: *Calcul des Probabilités*. Gauthier-Villars, Paris, (1889).
- [4] H. Blume: Mathematische Begründung und Entwicklung einer Wahrscheinlichkeitsrechnung. *Zeitschrift für Physik*, **92**, 232-256, (1934).
- [5] H. Blume: Zur Anwendung der Wahrscheinlichkeitsrechnung finiter Kollektive. *Zeitschrift für Physik*, **94**, 192-203, (1935).
- [6] G. Bohlmann: *Lebensversicherungs-Mathematik*. In *Encyklopädie der mathematischen Wissenschaften*, I Band, II Theil. Teubner, Leipzig, 852-917, (1901).
- [7] G. Bohlmann: Die Grundbegriffe der Wahrscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung. In *Atti del IV Congresso Internazionale dei Matematici*, vol. III, 244-278, (1908).
- [8] É. Borel: Remarques sur certaines questions de probabilité. *Bulletin de la Société Mathématique de France*. **3**, 123-128, (1905).
- [9] É. Borel: Sur les principes de la théorie cinétique des gaz. *Annales Scientifiques de l'École Normale Supérieure*, **23**, 9-32, (1906).
- [10] É. Borel: Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, **27**, (1909), 247-271.
- [11] É. Borel: *Le Hazard*. Librairie Felix Alcan, Paris, (1920).
- [12] U. Broggi: *Matematica attuariale. Teoria statistica della mortalità; matematica delle assicurazioni sulla vita*. Hoepli, Milano, (1906).

- [13] U. Broggi: *Die Axiome der Wahrscheinlichkeitsrechnung*. Dieterich, Göttingen, (1907).
- [14] M.-C. Bustamante, M. Cléry, L. Mazliak: *Le Traité du calcul des probabilités et de ses applications*. Étendue et limites d'un projet borélien de grande envergure (1921-1938). *North-Western European Journal of Mathematics*, **1**, 111-167, (2015).
- [15] F.P. Cantelli: Sulla probabilità come limite della frequenza. *Rendiconti della Reale Accademia dei Lincei*, **26** (S. V), 39-45, (1917).
- [16] G. Castelnuovo: *Calcolo delle probabilità*. Società Editrice Dante Alighieri, Milano-Roma-Napoli, (1919).
- [17] D. Costantini: *Fondamenti del calcolo delle probabilità*. Feltrinelli, Milano, (1970).
- [18] R.T. Cox: Probability, Frequency, and Reasonable Expectation. *American Journal of Physics*, **14**, 1-13, (1946).
- [19] R.T. Cox: *The Algebra of Probable Inference*. The Johns Hopkins Press, Baltimore, (U.S.A.), (1961).
- [20] A.I. Dale: A Study on Some Early Investigations into Exchangeability. *Historia Mathematica*, **12**, 323-336, (1985).
- [21] B. De Finetti: Fondamenti logici del ragionamento probabilistico. *Bollettino dell'Unione matematica italiana*, **9**, 1-3, (1930).
- [22] B. De Finetti: Problemi determinati e indeterminati nel calcolo delle probabilità. *Rendiconti della Reale Accademia Nazionale dei Lincei*, **12**, (1930), 367-373.
- [23] B. De Finetti: Funzione caratteristica di un fenomeno aleatorio. *Memorie della Reale Accademia Nazionale dei Lincei*, **4** (S. VI), (1930), 86-133.
- [24] B. De Finetti: Probabilismo. Saggio critico sulla teoria delle probabilità e sul valore della scienza. *Logos. Biblioteca di Filosofia*, **14**, 163-219, (1931).
- [25] B. De Finetti: Sul significato soggettivo della probabilità. *Fundamenta Mathematicae*, **17**, 298-329, (1931).
- [26] B. De Finetti: Statistica e probabilità nella concezione di R. von Mises. *Supplemento statistico i nuovi problemai di Politica, Storia ed Economia*, **2**, 9-19, (1936).
- [27] B. De Finetti: La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, **7**, 1-68, (1937).
- [28] B. De Finetti: Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Triestini, sez. II: scienza ed ingegneria*, **19** 29-81, (1949).

- [29] D. Gillies: *Philosophical Theories of Probability*. Routledge, London and New York, (2000).
- [30] K. Grelling, A. Herzberg: Diskussion über Wahrscheinlichkeit. *Erkenntnis*, **1**, 260-285, (1930).
- [31] F. Hausdorff: *Grundzüge der mengenlehre*. Veit & Comp., Leipzig, (1914).
- [32] D. Hilbert: Mathematische Probleme. *Archiv für Mathematik und Physik*, **1** (S. 3), 213-237, (1901). In: *Gesammelte Abhandlungen*, III Band. Springer, Berlin, (1935), 290-329.
- [33] W.E. Johnson: *Logic*. Part III. *The Logical Foundations of Science*. Cambridge University Press, Cambridge (U. K.), (1924).
- [34] W.E. Johnson: The Deductive and Inductive Problems. *Mind*, **41**, 409-423, (1932).
- [35] F. Klein: Ueber den allgemeinen Functionsbegriff und dessen Darstellung durch eine willkürliche Curve. *Mathematische Annalen*, **22**, (1883), 249-259.
- [36] A. N. Kolmogorov: *Grundbegriffe der Wahrscheinlichkeitsrechnung*. In: *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol **2**. Springer, Berlin, (1933).
- [37] U. Krengel: On the contributions of Georg Bohlmann to probability theory. *Journ@l Electronique d'Histoire des Probabilités et de la Statistique*, **7**, 1-13, (2011).
- [38] H. Lebesgue: *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthier-Villars, Paris, (1904).
- [39] F. P. Ramsey: *Truth and Probability*. In *The Foundations of Mathematics and other Logical Essays*. R.B. Braithwaite, Editor. Kegan Paul, Trench, Trubner & Co., London, (1931), pp. 156-198.
- [40] I. Schneider: *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Einführungen und Texte. Wissenschaftliche Buchgesellschaft Darmstadt, (1988).
- [41] G. Shafer, V. Vovk: The origins and legacy of Kolmogorov's *Grundbegriffe*. pp. 1-104, (2003).
- [42] H. Steinhaus: Les probabilités dénombrables et leur rapport à la théorie de la mesure. *Fundamenta Mathematicae*, **4**, 286-310, (1923).
- [43] R. von Mises: Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **4**, 52-99, (1919).

- [44] R. von Mises: On the Foundations of Probability and Statistics. *Annals of Mathematical Statistics*, **12**, 191-205, (1941).
- [45] R. von Mises: *Wahrscheinlichkeit Statistik und Wahrheit*. Springer, Berlin, (1928).
- [46] J. von Plato: *Creating modern probability*, Cambridge University Press, Cambridge (U.K.), 1994.