Thomas Bayes
Saggio sulla soluzione di un problema nella dottrina del caso.

Philosophical Transactions of the Royal Society of London,
53, 370–418, 1763

PROBLEMA

Sono dati: il numero di volte in cui un evento ignoto si è verificato e il numero di volte in cui non si è verificato;
È richiesta: la probabilità [*chance*] che la probabilità [*probability*] del suo verificarsi in una singola prova [*trial*] sia compresa tra due gradi di probabilità prestabiliti.[1]

SEZIONE I

DEFINIZIONI
1. Un certo numero di eventi sono *inconsistenti*,[2] se, verificandosi uno di essi, nessuno degli altri si può verificare.
2. Due eventi sono *contrari* quando o l'uno o l'altro devono verificarsi, ma non possono verificarsi insieme.
3. Si dice che un evento *fallisce* (*fails*) quando non si verifica o, ma è lo stesso, quando si verifica il suo contrario.
4. Un evento è detto *determinato* se si è verificato oppure no.
5. La *probabilità di un evento qualunque* è il rapporto tra il valore atteso, calcolato in dipendenza del verificarsi dell'evento, e la probabilità (*chance*) della cosa attesa[3] al verificarsi dell'evento.[4]

---

[1] [Bayes esordisce distinguendo tra due probabilità: l'oggettiva e ontologica, o dell'evento (*probability*), e la soggettiva ed epistemica, o del sapere del soggetto (*chance*). Il raddoppiamento e l'autoriferimento probabilistico sono decisamente moderni. Cartesianamente parlando, la probabilità epistemica precede logicamente la probabilità ontologica, come il *cogito* precede il *sum*. (Cfr. nota 11). Bayes parla di "gradi di probabilità" (*degrees of probability*) per indicare quelli che oggi si chiamano "gradi fiduciali" (*degrees of belief*). Riporto la definizione ISO di *degrees of belief*: *In contrast to this frequency-based point of view of probability, an equally valid viewpoint is that probability is a measure of the degree of belief that an event will occur. For example, suppose one has a chance of winning a small sum of money D and one is a rational bettor. One's degree of belief in event A occurring is p* = 0.5 *if one is indifferent to this two betting choices: (1) receiving D if event A occurs but nothing if it does not occur; (2) receiving D if event A does not occur but nothing if it does occur*. N.d.T.]

[2] [Oggi si dice *incompatibili*. N.d.T.]

[3] [*Expectation* si può leggere come *desire*. La teoria delle probabilità, intesa alla De Finetti come logica dell'incerto, si può interpretare psicanaliticamente come logica del desiderio. Il desiderio si "misura" come attesa in uno stato di incertezza. L'attesa dipende a sua volta dal sapere, che tuttavia è incompleto e indeterministico – da qui l'incertezza. N.d.T.]

[4] [Probabilmente la "cosa attesa al verificarsi dell'evento" è il guadagno. Definizione non chiara di probabilità che mescola valori attesi e probabilità. Bayes definirebbe la probabilità condizionata $p(A|B)$, dove la cosa attesa è A e l'evento da cui dipende è B. La confusione spiega la resistenza dei frequentisti ad accettare il modello bayesiano

6. Con *caso* (*chance*) intendo lo stesso che probabilità.[5]

7. Gli eventi sono indipendenti se il verificarsi di uno non aumenta né diminuisce la probabilità degli altri.

PROPOSIZIONE 1. *Se un certo numero di eventi sono inconsistenti, la probabilità che l'uno o l'altro si verifichi è la somma delle probabilità di ciascuno di essi.*

Supponi che esistano tre eventi siffatti e che io riceva [la somma] N quando uno qualunque dei tre si verifichi e che le probabilità del primo, del secondo e del terzo siano rispettivamente a/N, b/N e c/N. Allora, per la definizione di probabilità, il valore della mia attesa dal primo evento è a, dal secondo è b e dal terzo è c. Quindi il valore delle mie attese da tutti e tre gli eventi è in questo caso [la somma del]le attese da tutti e tre e sarà $a+b+c$. Ma la somma delle mie attese da tutti e tre gli eventi è in questo caso l'attesa di ricevere N, dato il verificarsi di uno di loro. Perciò, per la definizione 5, la probabilità di uno o l'altro di essi è ($a+b+c$)/N o a/N + b/N + c/N, o la somma delle probabilità di ciascuno di essi.

COROLLARIO. *Se è certo che uno o l'altro degli eventi si verifichi, allora* $a+b+c$ = N.

Infatti, poiché in questo caso tutte le attese insieme ammontano all'attesa certa di ricevere N, [la somma de]i loro valori deve essere uguale a N. Da qui risulta evidente che la probabilità di un evento, sommata alla probabilità del suo fallimento (o del suo contrario) è il rapporto di eguaglianza [*sic*]. Infatti, i due sono eventi inconsistenti, dei quali avviene necessariamente o l'uno o l'altro. Perciò, se la probabilità di un evento è P/N, la probabilità del suo fallimento sarà (N–P)/N.

PROPOSIZIONE 2. *Se una persona ha un'attesa in dipendenza dal verificarsi di un evento, la probabilità dell'evento sta alla probabilità del suo fallimento come la perdita, se fallisce, sta al guadagno, se si verifica.*

Supponi che una persona abbia l'attesa di ricevere N, in dipendenza da un evento che ha probabilità P/N. Allora, per la definizione 5, il valore della sua attesa è P. Perciò se l'evento fallisce, la persona perde il corrispondente di P. Se si verifica riceve N e la sua attesa cessa. Perciò il suo guadagno [netto] è N–P. Allo stesso modo, dato che la probabilità di un evento è N/P, la probabilità del suo fallimento è (N–P)/P per il corollario della proposizione 1. Ma (N–P)/P sta a N/P come P sta a N–P, cioè la

---

delle probabilità. L'interpretazione più plausibile, basata sull'uso stesso di Bayes, è che la probabilità sia il rapporto tra due valori attesi: il valore atteso corrispondente all'evento A, dato l'evento B, e il valore atteso corrispondente all'evento B, supposto verificato. La definizione sembra presupporre che $p(A|B) = p(AB)/p(B)$. In effetti, più che di una definizione si tratta di un assioma, che formula *in nuce* il teorema di Bayes. La definizione confusa "dimostra" come il soggetto moderno della scienza sappia lavorare con la propria ignoranza. Bayes ignora l'essenza della probabilità, tanto che non sa darne una buona definizione. Ma dal cappello della propria ignoranza tira fuori il coniglio di una nuova teoria delle probabilità – nuova rispetto alla teoria del passato (Bernoulli) e nuova rispetto alla teoria del futuro (Laplace). N.d.T.]

[5] [Appena distinte, le due probabilità vengono subito unificate. Sapere ed essere, dopo essere stati separati, confluiscono. Tanto va ricordato a coloro i quali accusano il cartesianesimo di dualismo. N.d.T.]

probabilità di un evento sta alla probabilità del suo fallimento come la perdita se fallisce sta al guadagno se si verifica.[6]

PROPOSIZIONE 3. *La probabilità che due eventi successivi si verifichino entrambi è un rapporto composto della probabilità del primo e della probabilità del secondo, supposto che il primo si sia verificato*.[7]

Supponi che, se entrambi gli eventi si verificano, riceverò N e che la probabilità di verificarsi di entrambi sia P/N e che il primo accada con probabilità *a*/N (quindi sia (N–a)/N la probabilità che non accada) e che il secondo accada nell'ipotesi che si verifichi il primo con la probabilità *b*/N. Allora, per la definizione 5, P sarà il valore della mia attesa di ricevere *b*, se si verifica il primo. Di conseguenza, se il primo si verifica, il mio guadagno è *b*–P e, se il primo fallisce, la mia perdita è P. Quindi, componendo all'inverso, *a* sta a N come P sta a *b*.[8] Ma il rapporto di P su N è composto dal rapporto di *a* su N e dal rapporto di *b* su N. Perciò lo stesso rapporto di P su N è composto dal rapporto di *a* su N e dal rapporto di *b* su N. In altri termini, la probabilità che entrambi gli eventi accadano insieme è composta dalla probabilità del primo e dalla probabilità del secondo, supposto che accada il primo.[9]

COROLLARIO. *Se di due eventi successivi la probabilità del primo è a/*N *e la probabilità di entrambi è* P/N. *allora la probabilità del secondo, supposto che si verifichi il primo, è* P/*a*.

PROPOSIZIONE 4. *Dati due eventi successivi determinati ogni giorno, se ogni giorno la probabilità del secondo è b/*N *e la probabilità di entrambi* P/N, *e se ricevo* N *se entrambi gli eventi si verificano nel primo giorno in cui il secondo si verifica, allora, date queste condizioni, dico che la probabilità di ottenere* N *è* P/*b*.

Infatti, in caso contrario sia *x*/N la probabilità di ottenere N e *y* stia a *x* come N–*b* sta a N. Allora, siccome *x*/N è, per la definizione 1, la mia probabilità di ottenere N, *x* è il valore della mia attesa. Poiché, in base alle condizioni date il primo giorno ho un'attesa di ottenere N che dipende dal verificarsi di entrambi gli eventi insieme, la cui probabilità è P/N, il valore di questa attesa P. Allo stesso modo, se questa coincidenza

[6] [Bayes non pubblicò il suo saggio. Resisteva alla propria stessa innovazione scientifica? Una misura dell'eventuale resistenza è il ricorso al linguaggio euclideo delle proporzioni. Versare vino nuovo nelle botti vecchie doveva sembrare rassicurante al buon presbitero. Del resto anche Newton nei suoi *Principi di filosofia naturale* diede una macchinosa dimostrazione euclidea dell'ellitticità delle orbite planetarie, nell'ipotesi che su di essi agisse una forza centrale. Non adottò il calcolo delle flussioni che era più rapido e adatto al problema. N.d.T.]
[7] [Il teorema porge la probabilità della concatenazione di eventi. Considerando gli eventi come significanti, il teorema tratta della probabilità della metonimia. N.d.T.]
[8] [Come già Newton, dicevo, Bayes non abbandona l'assetto euclideo della proporzionalità, forse perché connesso alla nozione di misura, che ai fisici del tempo appariva nozione irrinunciabile. Anche i moderni fisici quantistici hanno problemi ad abbandonare la nozione di misura, che pure crea loro tanti grattacapi (teoria del collasso). Facendo intervenire un infinito superiore al numerabile, una teoria dotata di misura sembra più potente di una teoria più qualitativa. Pregiudizi prescientifici.]
[9] [Il trattato di Hume *Ricerca sull'intelletto umano* è di 15 anni prima. Bayes dà una versione quantitativa dell'indebolimento del principio di ragion sufficiente. La causa è ridotta a evento ipotetico, dotato di una certa probabilità di verificarsi. N.d.T.]

non si verifica io ho un'attesa di essere reintegrato nelle condizioni precedenti, cioè di ricevere ciò che vale x, in dipendenza del non verificarsi del secondo evento, la cui probabilità, per il corollario della proposizione 1., è (N–b)/N o y/x, poiché y sta a x come N–b sta a N. Quindi, dato che x è la cosa attesa [il guadagno atteso] e y/x la probabilità di ottenerlo, il valore di questa attesa è lo stesso di prima, cioè x. Di conseguenza la probabilità di ottenere N è, per la definizione 5, ancora x/N o P/b.[10] Ma dopo questa scoperta la probabilità che io ottenga N è la probabilità che il primo evento si sia verificato, supponendo che si sia verificato il secondo, le cui probabilità sono state specificate in precedenza.

Ma la probabilità che un evento si sia verificato è pari alla probabilità che io abbia indovinato, avendo scommesso che sarebbe si sarebbe verificato.[11] Perciò è evidente la seguente proposizione.

PROPOSIZIONE 5. *Dati due eventi successivi, essendo la probabilità che si verifichi il secondo* b/N e la *probabilità che si verifichino insieme entrambi* P/N, *se si scopre che il secondo evento si è verificato, la probabilità che io abbia indovinato è ancora* P/b.[12]

PROPOSIZIONE 6. *La probabilità che si verifichino diversi eventi indipendenti è il rapporto composto*[13] *delle rispettive probabilità.*

Data la natura degli eventi indipendenti, la probabilità che uno qualsiasi di loro non varia se uno degli altri si verifica oppure no. Di conseguenza la probabilità che il secondo evento si verifichi, avendo supposto che si sia verificato il primo, coincide con la probabilità originaria. Ma la probabilità che due eventi qualsiasi si verifichino è, per la proposizione 3., il rapporto composto della probabilità del primo evento e della probabilità del secondo evento, supposto che si sia verificato il primo. Quindi la probabilità che due eventi indipendenti si verifichino entrambi è il rapporto composto della probabilità del primo e della probabilità del secondo. Allo stesso modo,

---

[10] Quanto qui si afferma può forse essere chiarito considerando che tutto quanto può essere perso al verificarsi del secondo evento è la probabilità di essere reintegrato nelle condizioni precedenti, se l'evento da cui la mia attesa dipendeva è stato determinato come detto nella proposizione. Ma questa probabilità è sempre tanto a mio favore quanto contro di me. Se si verifica il primo evento è contro di me e pari alla probabilità che il secondo evento non si verifichi. Se il primo evento non si verifica è a mio favore e pari alla probabilità che il secondo evento non si verifichi. La perdita, perciò, non può essere a mio svantaggio.

[11] [Questo è l'assunto di base della concezione soggettivistica o epistemica della probabilità. Questo assunto è qui messo alla prova nel caso della probabilità condizionata. N.d.T.]

[12] Quel che Bayes dimostra in questo e nel precedente teorema equivale alla risposta alla seguente questione. Qual è la probabilità che un certo evento, quando si verifica, sia accompagnato da un altro, da determinare nello stesso momento? In questo caso, essendo dato uno degli eventi, nulla può essere dovuto per la sua attesa. Di conseguenza il valore di un'attesa al verificarsi di entrambi deve essere pari all'attesa che dipende dal verificarsi di uno di essi. In altri termini, la probabilità che, quando uno dei due eventi si verifica, l'altro si verificherà è pari alla probabilità di quest'altro. Detta x la probabilità di quest'altro, se b/N è la probabilità dell'evento dato, e p/N è la probabilità di entrambi, poiché $p/N = b/N \times x$, si ottiene $x = p/b$, la probabilità dei teoremi 4. e 5. [Il teorema 5. È quello che correntemente si chiama teorema di Bayes o, con vezzo eziologico, teorema delle probabilità delle cause. N.d.T.]

[13] [Prodotto. N.d.T.]

considerando il primo e il secondo evento come un evento, la probabilità che tre eventi indipendenti si verifichino tutti insieme è il rapporto composto delle probabilità che si verifichino i primi due e della probabilità del terzo. Così possiamo procedere per un numero qualsiasi di eventi e il teorema è evidente.

Corollario 1. *Dati diversi eventi indipendenti, la probabilità che il primo si verifichi, il secondo non si verifichi, il terzo non si verifichi, il quarto si verifichi ecc. è il rapporto composto delle probabilità che si verifichi il primo, non si verifichi il secondo, non si verifichi il terzo, si verifichi il quarto ecc. Per non verificarsi di un evento si intende sempre il verificarsi del suo contrario*.

Corollario 2. *Dati diversi eventi tali che la probabilità per ciascuno di essi sia a e la probabilità di non verificarsi sia b, la probabilità che il primo si verifichi, il secondo non si verifichi, il terzo non si verifichi, il quarto si verifichi ecc. è il rapporto composto abba ecc.*

Infatti, secondo la notazione algebrica, se *a* indica un rapporto e *b* un altro, *abba* indica il rapporto composto di *a*, *b*, *b*, *a*. Questo corollario è solo un caso particolare del precedente.

DEFINIZIONE. Se, in conseguenza di certi dati,[14] si configura la probabilità che un certo evento possa accadere, il verificarsi o il non verificarsi di quell'evento come conseguenza di questi dati, lo chiamo verificarsi o non verificarsi alla *prima prova* (*trial*).[15] Se gli stessi dati si ripetono, il verificarsi o il non verificarsi di quell'evento come loro conseguenza, lo chiamo verificarsi o non verificarsi dell'evento alla *seconda prova*. E cosi si procede a ogni prova, finché gli stessi dati si ripetono.

È da qui evidente che il verificarsi o il non verificarsi dello stesso evento in così tante e differenti prove è in realtà il verificarsi o il non verificarsi di altrettanti eventi indipendenti esattamente simili l'uno all'altro.

---

[14] [Cos'è un "dato"? A differenza di quello antico, in cui i "dati" erano fissi e avevano autonomia ontologica – in pratica anticamente ciò che era dato erano le cause o i principi primi – nel discorso scientifico moderno i "dati" non sono dati in senso stretto una volta per tutte, ma sono presupposti epistemici variabili da problema a problema. Formano un sapere congetturale, che viene presupposto e da cui si deducono conseguenze vere (teoremi) e che si confutano con controesempi che li falsificano. In matematica i dati si chiamano *assiomi* e le conseguenze *teoremi*, in fisica *condizioni al contorno*, in biologia *risultati sperimentali* o *osservazioni*, in informatica formano l'*archivio* o *database*. Nel calcolo delle probabilità i dati operano da fattori condizionanti di probabilità condizionate. Non avendo valore ontologico, i dati scientifici sono semplici valori di variabili. Va notato che le nozioni di variabile e di funzione si impongono in matematica solo nel XVIII secolo a opera di Eulero, sostituendo e generalizzando la nozione classica di quantità. N.d.T.]

[15] [L'avvento della nozione di *prova* (*trial*) rappresenta il definitivo e pratico indebolimento del principio di ragion sufficiente, che all'epoca era già stato decostruito teoricamente da David Hume. Nella prova l'effetto può conseguire o non conseguire alla causa. La nozione di causa perde così la propria categoricità o il proprio potere determinante, che gli era stato conferito da Aristotele duemila anni prima. Inoltre, la nozione di prova, come evento indipendente da altre prove – vedi proposizione 7. – semplifica i calcoli. In senso astratto, lo spazio delle prove è uno spazio prodotto. N.d.T.]

PROPOSIZIONE 7. *Sia a la probabilità che un evento si verifichi in ogni singola prova e b la probabilità che non si verifichi. La probabilità che l'evento si verifichi p volte e non si verifichi q volte in* $(p+q)$ *prove è* Ea^pb^q, *dove E è il coefficiente del termine* a^pb^q *nell'espansione della potenza del binomio* $(a+b)^{p+q}$.[16]

Per il verificarsi o il non verificarsi di un evento le differenti prove sono altrettanti eventi indipendenti. Per i corollario 2. della proposizione 6. la probabilità che l'evento si verifichi nella prima prova, non si verifichi nella seconda e nella terza, si verifichi nella quarta, non si verifichi nella quinta ecc. (verificandosi e non verificandosi finché il numero dei successi sia p e degli insuccessi sia q) è *abbab* ecc. dove il numero di a è p e il numero di b è q, cioè a^pb^q. Analogamente, considerando l'evento di p successi e q insuccessi in qualsiasi ordine, la sua probabilità è ancora a^pb^q. Ma il numero di differenti ordinamenti in cui un evento può verificarsi p volte e non verificarsi q volte è pari al numero di permutazioni [distinte] che *aaaabbb* ammette, quando p sia il numero delle a e q il numero delle b. Tale numero è pari a E, il coefficiente dell'espansione binomiale di $(a+b)^{p+q}$ in cui compare a^pb^q. L'evento può quindi verificarsi p volte e non verificarsi q volte in E modi e non di più in $p+q$ prove. Il suo verificarsi e non verificarsi sono eventi inconsistenti, la probabilità di ciascuno dei quali è a^pb^q. Perciò, per la proposizione 1., la probabilità che uno o l'altro modo di verificarsi p volte e di non verificarsi q volte in $p+q$ prove è Ea^pb^q.[17]

SEZIONE II.

POSTULATO 1. Supponiamo che la tavola o piano rettangolare ABCD siano fatti e livellati in modo tale che, se una delle biglie O o W è gettata su ABCD, è uguale la probabilità che si fermi in una qualsiasi delle parti uguali di ABCD e che si fermi necessariamente in una di esse.[18]

POSTULATO 2. Suppongo che la biglia W sia gettata per prima. Nel punto in cui si ferma, traccio la parallela *os* al lato AD, che incontra CD e AB in *o* e *s*. Dopo di che si

[16] [Questo è il teorema da cui Jakob Bernoulli (1713) deduce la legge dei grandi numeri. (Cfr. *Ars conjectandi*, Cap. IV). Conserviamo il simbolo E, che oggi si scriverebbe $\binom{p+q}{p}$, perché indica il valore atteso (*expectation*) dell'evento: p successi e q insuccessi. Essendo soggettivistica la forma bayesiana del calcolo delle probabilità, la probabilità è il guadagno che il soggetto si aspetta in condizioni aleatorie. N.d.T.]

[17] [È questo il momento giusto per rilevare la differenza e la simmetria tra l'approccio di Bernoulli e quello di Bayes. Entrambi trattano il problema binomiale ma, mentre il primo tiene fissa la probabilità e fa variare il numero di prove, il secondo tiene fisso il numero di prove e fa variare la probabilità. In termini moderni, il primo è frequentista e dimostra che la frequenza si avvicina alla probabilità al crescere del numero delle prove, mentre il secondo è soggettivista e tratta le probabilità come variabili indipendenti. Tuttavia, la diatriba tra frequentisti e soggettivisti non ha senso. Frequentismo e soggettivismo sono due modi diversi ma complementari di guardare la stessa cosa: la formula newtoniana di espansione del binomio. N.d.T.]

[18] [Questo postulato *non* è intuizionista. L'intuizionismo prevede esplicitamente il caso in cui la biglia non si fermi mai. Tale evento si realizzerebbe in un biliardo perfettamente elastico e senza attrito. È comunque singolare che Bayes scelga come modello per presentare una teoria indeterminista proprio il biliardo, che è il luogo degli eventi deterministi per eccellenza, regolati dalle equazioni di Lagrange, che vedevano la luce proprio a quei tempi. N.d.T.]

getti la biglia O $p+q = n$ volte. Se O si ferma tra *os* e AD in un singolo lancio, si dice che si è verificato l'evento M in una singola prova.

Tutto ciò premesso, valgono i seguenti lemmi.

LEMMA 1. *La probabilità che il punto o cada tra due punti di AB è pari al rapporto tra la distanza di questi punti sulla lunghezza di AB.*

Siano due punti qualsiasi sulla linea AB, che chiamiamo *f* e *b*. Attraverso di essi tracciamo le F*f* e *b*L parallele a AD che incontrano CD rispettivamente in F e L. Si danno due casi.

a) I rettangoli C*f*, F*b* e LA sono reciprocamente commensurabili.

Allora possono essere divisi in parti uguali. Gettata la biglia W, la probabilità che si fermi su un numero qualunque di queste parti uguali di queste parti uguali è la somma delle probabilità di fermarsi su una qualunque di esse, in quanto la fermata su una qualunque delle parti di AC è un evento che fa parte di un insieme di molti eventi inconsistenti e quindi per la proposizione 1. le probabilità si sommano. (La probabilità che la biglia si fermi su una qualsiasi delle parti uguali è la stessa). È la probabilità che si fermi su una qualunque di queste parti uguali, moltiplicata per il loro numero.

Di conseguenza, la probabilità che la biglia W si fermi da qualche parte in F*b* è la probabilità che si fermi su una delle parti uguali in cui F*b* è suddivisa, moltiplica per il loro numero. Inoltre la probabilità che W si fermi da qualche parte in C*f* o LA, cioè la probabilità che non si fermi in f*b* (dato che deve fermarsi da qualche parte in AC per il primo postulato) è la probabilità che si fermi sopra una delle parti uguali, moltiplicata per il numero delle parti uguali che compongono C*f* e LA.

Pertanto la probabilità che la biglia W si fermi in F*b* sta alla probabilità che non si fermi in F*b* come il numero di parti uguali di F*b* sta al numero di parti uguali di C*f* e LA, presi insieme o, ma è lo stesso, come F*b* sta a C*f* e LA, presi insieme o, ancora, come *fb* sta a B*f e* A*b*, presi insieme. Componendo all'inverso, la probabilità che la biglia W si fermi in F*b* sommata alla probabilità che non si fermi in F*b* sta come *fb* a AB o come il rapporto di F*b* su AB sta al rapporto di AB su AB. Ma la probabilità di qualunque evento, sommata alla probabilità di non verificarsi è il rapporto sull'uguaglianza. Perciò la probabilità che la biglia W si fermi su F*b* sta al rapporto sull'eguaglianza come il rapporto di *fb* su AB sta al rapporto di AB su AB, ossia al rapporto sull'eguaglianza. Perciò la probabilità che la biglia si fermi su F*b* è il rapporto di *fb* su AB. Ma, per ipotesi, a seconda che la biglia W si fermi su F*b* oppure no, il punto *o* giacerà tra *f* o *b*, e pertanto la probabilità che il punto *o* giaccia tra *f* e *b* è il rapporto di *fb* su AB.

b) I rettangoli C*f*, F*b* e LA *non* sono reciprocamente commensurabili. Anche in questo caso si dimostra che la probabilità non può essere né maggiore né minore del rapporto di *fb* su AB. Ammettiamo che sia minore e pari al rapporto di *fc* su AB. Sulla linea *fb* prendiamo i punti *p* e *t* tali che *pt* sia maggiore della metà di *cb* e *p* e *t* siano i punti di divisione più vicini rispettivamente a *f* e *c* su *fb*.[19] Allora, essendo B*p*, *pt* e *t*A commensurabili, lo sono anche i rettangoli C*p*, D*t* e il rettangolo su *pt*, che completa AB. Per quanto detto, la probabilità che il punto *o* giaccia tra *p* e *t* è il rapporto tra *pt* e AB. Ma se *o* giace tra *p* e *t* deve giacere tra *f* e *b*. Pertanto la probabilità che giaccia tra *f* e *b* non può essere minore del rapporto di *fc* su AB, perché *pt* è maggiore di *fc.*

Allo stesso modo si dimostra che la citata probabilità non può essere superiore al rapporto di *fb* su AB.

---

[19] [Ciò è sempre possibile. Basta dividere AB in parti uguali minori della metà di *cb* e prendere come punti *p* e *t* i punti sul segmento *fb* più vicini agli estremi *f* e *b*. N.d.T.]

Deve quindi essere uguale.

LEMMA 2. *Gettata la biglia W e tracciata la linea os, la probabilità dell'evento M in una singola prova è il rapporto di Ao su AB.*

Infatti, come nel lemma precedente, la probabilità che la biglia O si fermi da qualche parte in D*o* o tra AD e *so* è il rapporto di A*o* su AB. Ma il fermarsi della biglia O tra AD e *so* dopo un singolo lancio è il verificarsi dell'evento M in una singola prova. E il lemma è evidente.

COSTRUZIONE. Su BA si disegna la figura B*ghikm*A con la proprietà $y = Ex^p r^q$,[20] con il seguente significato dei simboli. Con il punto *b* si divide arbitrariamente BA in due parti A*b* e B*b*. A partire da *b* si traccia la perpendicolare limitata dalla figura in *m*. *y, x, r* rappresentano rispettivamente i rapporti a AB di *bm*, A*b* e B*b*. *E* ha il solito significato del coefficiente del termine $x^p r^q$ nell'espansione del binomio $(x+r)^{p+q}$.

Tutto ciò premesso, affermo la seguente

PROPOSIZIONE 8. Prima di lanciare la biglia W, la probabilità che il punto *o* cada tra due punti qualunque *f* e *b* di AB e che su *p*+*q* prove l'evento M si verifichi *p* volte e non si verifichi *q* volte, è il rapporto rispetto a CA, il quadrato costruito su AB, di *fghikmb* la parte della figura B*ghikm*A compresa tra le perpendicolari *fg* e *bm*, innalzate sopra AB.

DIMOSTRAZIONE. Ammettiamo che non sia così e sia dapprima il rapporto di D maggiore di quello di *fghimb* su CA. Attraverso i punti *e*, *d*, *c*, tracciamo le perpendicolari a *fb* che incontrano la curva (*sic*) A*mig*B in *h*, *i*, *k*, essendo il punto *d* localizzato in modo che *di* sia la maggiore delle perpendicolari comprese tra la linea *fb* e la curva (*risic*) A*mig*B, ed essendo i punti *e*, *d*, *c*, così tanti e così localizzati che i rettangoli *bk*, *ci*, *ei* e *fb*, presi insieme differiscono da *fghikmb* meno di quanto differisca D. Tutto ciò si può agevolmente fare con l'ausilio dell'equazione della curva (basta *sic*) e la differenza tra D e la figura *fghikmb* data. Allora, essendo *di* l'ordinata

[20] [La traduzione tedesca porge "curva", termine che Bayes (o il suo *editor*) usa più avanti. È qui evidente come a Bayes manchi la nozione di funzione, che all'epoca Eulero stava ancora elaborando. Bayes, come già Newton, è costretto a dimostrare i suoi teoremi *more geometrico*, che non è il massimo della vita, nonostante l'opinione di Spinoza. In un certo senso, all'alba dell'era scientifica la geometria euclidea diventa la matematica ufficiale, quella garantita da professori e accademici, molto prudenti nell'accettare la nuova matematica: o quella infinitesimale di Leibniz o quella infinita di Maurolico. C'è una ragione strutturale per ciò, che va al di là delle preferenze personali. La geometria euclidea è completa. La completezza è l'eredità del modo di pensare classico, che nel Rinascimento gli umanisti strappano al pensiero religioso e vorrebbero trasferire al pensiero laico (che per altro non ne avvertiva la mancanza). Nella fattispecie non esistono verità euclidee indimostrabili a partire dagli assiomi di Euclide (o meglio da quelli di Hilbert). I ricorrenti tentativi fallimentari di dimostrare il V postulato di Euclide erano animati dal terrore dell'incompletezza. Era una paura vana. Nel XIX secolo si dimostrerà che esistono verità non euclidee sì, ma in geometrie non euclidee. Ma la nuova scienza che nel XVII secolo, che esordisce con il dubbio cartesiano, quella sì minaccia di essere veramente incompleta. Da qui la levata di scudi contro Cartesio e l'asserragliarsi dei benpensanti (aristotelici) dietro i bastioni euclidei. N.d.T.]

perpendicolare maggiore che insiste su *fb*, le altre saranno progressivamente minori man mano che se ne allontanano ai due lati di *fb*, come si vede dalla costruzione della figura. Di conseguenza *eb* sarà maggiore di *gf* e di ogni altra ordinata che insiste su *ef*.

Ora se A*o* fosse uguale a A*e*, allora per il lemma 2. La probabilità dell'evento M in una singola prova sarebbe il rapporto di A*e* su AB. Di conseguenza, per il corollario della proposizione 1., la probabilità che non si verifichi sarebbe il rapporto di B*e* su AB. Pertanto, se x e r sono rispettivamente i citati rapporti, per la proposizione 7., la probabilità di verificarsi p volte e di non verificarsi q volte in $p+q$ prove sarà $Ex^p r^q$. Ma x e r sono i rapporti a AB rispettivamente di A*e* e B*e*. Se y è il rapporto di *eb* su AB, allora per costruzione della figura A*i*B, $y = Ex^p r^q$. Quindi, se A*o* fosse uguale ad A*e*, allora la probabilità che l'evento M si verifichi p volte e non si verifichi q volte in $p+q$ prove sarebbe y, cioè il rapporto di *eb* su AB. E se A*o* fosse uguale a A*F* o fosse qualunque altro valore intermedio tra A*e* e A*f*, l'ultima probabilità menzionata sarebbe per la stessa ragione sarebbe il rapporto su AB di *fg* o di qualunque altra ordinata che insista su *ef*. Ma *eh* è la maggiore di tutte le ordinate che insistono su *ef*. Perciò, supponendo che il punto giaccia in un posto qualsiasi tra *f* e *e*, la probabilità che l'evento M si verifichi p volte e non si verifichi q volte in $p+q$ prove, non può superare il rapporto di *eh* su AB. Qui, allora, si verificano conseguentemente due eventi. Il primo è che il punto *o* sarà fra *e* e *f*; il secondo che l'evento M si verifica p volte e non si verifica q volte in $p+q$ prove. La probabilità del primo, per il primo lemma, è il rapporto di *ef* su AB e, supposto che il primo si verifichi, per quanto appena dimostrato, segue che la probabilità del secondo non può superare il rapporto di *eh* su AB, per la proposizione 3. Ne consegue evidentemente che a probabilità che si verifichino entrambi non può superare il rapporto composto di *ef* su AB e di *eh* su AB, cioè il rapporto di *fh* su CA. Perciò la probabilità che il punto *o* giaccia tra *f* e *e* e che l'evento M si verifichi p volte e non si verifichi q volte non supera il rapporto di *fh* si CA.

Analogamente la probabilità che il punto *o* sia compreso tra *e* e *d* e che l'evento M si verifichi e non si verifichi come prima non può superare il rapporto di *bk* su CA. Sommiamo tutte queste diverse probabilità e, per la proposizione 1., la loro somma darà la probabilità che il punto *o* giaccia tra *f* e *b* e che l'evento M si verifichi p volte e non si verifichi q volte in $p+q$ prove. Allo stesso modo sommiamo i corrispondenti rapporti e otterremo il rapporto della somma degli antecedenti alla somma dei conseguenti, cioè il rapporto di $fb+ei+ci+bk$ su CA. Tale rapporto è minore di D su CA, poiché D è maggiore di $fb+ei+ci+bk$. Pertanto, la probabilità che il punto *o* stia tra *f* e *b* e che l'evento M si verifichi p volte e non si verifichi q volte in $p+q$ prove è minore del rapporto di D su CA. Ma si era supposto uguale, il che è assurdo.

Allo stesso modo, iscrivendo nella figura i rettangoli *eg*, *dh*, *dk*, *cm*, si dimostra che l'ultima probabilità menzionata supera qualunque rapporto inferiore al rapporto di *fghikmb* su CA.

Si conclude che la probabilità è il rapporto di *fghikmb* su CA.

COROLLARIO. Prima di gettare la biglia W la probabilità che il punto *o* sia da qualche parte tra A e B o da qualche parte sulla linea AB e l'evento M si verifichi p e non si verifichi q volte in $p+q$ prove è il rapporto dell'intera figura A*i*B su ZCA.

Ma è certo che il punto *o* stia da qualche parte sulla linea AB. Perciò prima di lanciare la biglia W, la probabilità che l'evento M si verifichi p e non si verifichi q volte in $p+q$ prove è il rapportodi A*i*B su AC.

PROPOSIZIONE 9. *Se, prima di conoscere il posto del punto o, si ammette che l'evento M si è verificato p volte e non si è verificato q volte in p+q prove, allora posso congetturare che il punto o giace tra due punti qualsiasi f e b della linea AB e* *conseguentemente che la probabilità dell'evento M in una singola prova è compresa tra il rapporto di Ab su AB e di Af su AB. La probabilità di essere nel giusto è il rapporto di quella parte della figura AiB, prima descritta, che è compresa tra le perpendicolari condotte sopra AB in corrispondenza dei punti f e b, rispetto all'intera figura AiB.*

Infatti, si danno successivamente i seguenti due eventi. Il primo è che il punto *o* sia compreso tra *f* e *b*; il secondo è che l'evento M si verifichi *p* volte e non si verifichi *q* volte in $p+q$ prove. Per la proposizione 8. la probabilità originaria del secondo evento è il rapporto di A*i*B su CA. Perciò, per la proposizione 5., se in primo luogo si scopre che si è verificato il secondo evento, e da qui congetturo che si è verificato anche il primo. La probabilità che io sia nel giusto è il rapporto di *fghimb* su A*i*B, come volevasi dimostrare.

COROLLARIO. Date le stesse ipotesi, congetturo che la probabilità che l'evento M è compresa tra *o* e il rapporto di A*b* su AB, la probabilità [*chance*] che io sia nel giusto è il rapporto di A*bm* su A*i*B.

SCOLIO. Dalle proposizioni precedenti risulta chiaro che nel caso di un evento come quello che qui chiamo M, dal numero delle prove in cui si verifica e dal numero di prove in cui non si verifica, dato un certo numero di prove totali, senza conoscere molto di più concernente tale evento, si può congetturare la sua probabilità.[21] Con il metodo usuale di valutare l'ampiezza delle aree si può vedere la probabilità [*chance*] che la congettura sia giusta. E che la stessa regola sia quella giusta da usare nel caso di un evento della cui probabilità non sappiamo assolutamente nulla prima di qualsiasi prova che lo riguardi, sembra evidente dalla seguente considerazione. (Intendo un evento per il quale non ho ragione di pensare che in un certo numero di prove si verifichi un certo numero di prove piuttosto che un altro). Infatti, su questa base, posso giustamente ragionare su di esso come se la probabilità non fosse stata fissata prima e poi sia determinata in modo da non darmi motivo di pensare che in un certo numero di prove dovrebbe verificarsi un certo numero di volte piuttosto che qualsiasi altro. Ma questo è proprio il caso dell'evento M. Infatti, prima del lancio della biglia W, che per il corollario della proposizione 8. determina la probabilità in una singola prova, la probabilità che l'evento M ha di verificarsi *p* volte e di non verificarsi *q* volte in $p+q$ prove è il rapporto di A*i*B su CA. Come si può vedere con il metodo delle flussioni, questo rapporto è lo stesso, una volta dati $p+q$ o *n*, qualunque sia *p*.[22] Di conseguenza, prima di scoprire il posto del punto *o* o il numero di volte che l'evento M si è verificato non ho motivo di pensare che debba verificarsi un certo numero di volte piuttosto che un altro.

In quanto segue, pertanto, darò per acquisito che la regola data nella proposizione 9. per l'evento M sia la regola da usare in relazione a qualunque evento della cui

[21] [In Bayes più che in Bernoulli è evidente la maturazione epistemologica del calcolo delle probabilità, che diventa autoreferenziale. Con Bayes ha senso parlare di probabilità delle probabilità. Il calcolo delle probabilità diventa il calcolo delle probabilità delle congetture, che Bernoulli aveva in animo di fondare (*Ars conjectandi*), ma si fermò al calcolo delle frequenze più probabili. Laplace riprenderà Bayes più che Bernoulli, per esempio calcolando in $(p+1)/(p+q+2)$ la probabilità del $(p+1)$-esimo successo dopo *p* successi e *q* insuccessi nelle $p+q$ prove precedenti. N.d.T.]

[22] Si prova con il metodo qui menzionato che A*i*B contratto al rapporto di E a 1 sta a CA come 1 a $(n+1)\times$E. Da qui segue che prima della contrazione A*i*B deve stare in rapporto a CA come 1 a $(n+1)$, che è costante, dato *n*, e non dipende da *p*.

probabilità non si sappia assolutamente nulla prima di qualsiasi prova o osservazione. Lo chiamerò *evento ignoto*.

Supponendo di contrarre le ordinate nella figura A*i*B fino al rapporto di E su uno, senza alterare la proporzione delle parti di figura comprese tra loro, e applicando quanto detto per l'evento M a un evento ignoto, abbiamo la seguente proposizione che dà le regole per trovare la probabilità di un evento a partire dal numero di volte che si verifica e non si verifica.

Sia una figura, iscritta sulla base AH, con equazione $y = x^p r^q$, dove y, x, r sono rispettivamente i rapporti delle ordinate di un punto della figura, del segmento della base compreso tra il piede dell'ordinata e A, inizio della base, dell'altro segmento compreso tra il piede dell'ordinata e H, termine della base.

PROPOSIZIONE 10. *Se un evento ignoto si è verificato p volte e non si è verificato q volte in p+q prove, e da due punti f e t della base si tracciano le perpendicolari al curva fc e tF, la probabilità* [chance] *che la probabilità di un evento sia compresa tra il rapporto di Af su AH e il rapporto di At su AH è il rapporto di tFCf, la parte di figura compresa tra le ordinate, e ACFH, l'intera figura che insiste sulla base AH.*

Ciò è evidente dalla proposizione 9. e dall'osservazione nello scolio.

Per rendere tale regola applicabile in pratica dobbiamo trovare il valore dell'area della figura descritta e delle diverse parti ritagliate dalle diverse ordinate, perpendicolari alla base.

1. A tale scopo poniamo AH = 1 e HO = 1, HO il quadrato costruito su AH. Sia Cf = y e Af = x e Hf = r, denotando y, x, r rispettivamente i rapporti di Cf, Af e Hf su AH. Dall'equazione della curva $y = x^p r^q$ e da $r+x = 1$ (poiché Af+fH = AH), si ottiene $y = x^p(1–x)^q = x^p – qx^{p+1} + q.(q–1)/2.x^{p+2} – q.(q–1)/2.(q–2)/3.x^{p+3}$ + ecc. Ora, se l'ascissa è x e l'ordinata x^p, l'area corrispondente è $x^{p+1}/(p+1)$ (per la proposizione 10, cas. 1. Quadrat. Newt.) e se l'ordinata è qx^{p+1} l'area è $qx^{p+2}/(p+2)$ e analogamente per gli altri termini.[23] Perciò se l'ascissa è x e l'ordinata $y = x^p – qx^{p+1} + q.(q–1)/2.x^{p+2} – q.(q–1)/2.(q–2)/3.x^{p+3}$ + ecc., l'area corrispondente è la somma $x^{p+1}/(p+1) – qx^{p+2}/(p+2) + q.(q–1)/2.x^{p+3}/(p+3)$ – ecc. Quindi, se x = Af = Af/AH e y = Cf/AH, allora ACf =ACf/HO = $x^{p+1}/(p+1) – qx^{p+2}/(p+2) + q.(q–1)/2.x^{p+3}/(p+3)$ – ecc.

Da questa equazione, se q è piccolo, è facile calcolare il valore del rapporto di ACf su HO. Analogamente si calcola il valore del rapporto di HCf su HO che è $r^{q+1}/(q+1) – pr^{q+2}/(q+2) + p.(p–1)/2.r^{q+3}/(q+3)$ – ecc. Anche questa serie consiste di pochi termini se p è piccolo.

2. Alle stesse condizioni si calcola il rapporto di ACf su HO.

3. Analogamente per il rapporto di HCf su HO.

4. Se E è il coefficiente del termine $a^p b^q$ nell'espansione binomiale di $(a+b)^{p+q}$, il rapporto dell'intera figura ACFH su HO è $1/(n+1) \times 1/$E, con $n = p+q$, poiché, quando Af =AH, x = 1, r = 0. Pertanto tutti i termini della serie in 2. che si esprimono come rapporti di ACf su HO si annullano tranne l'ultimo, che diventa $1/(n+1)\times q/(p+1)\times(q–1)/(p+2)\times$ ecc. $\times 1/n$. Ma, essendo E il coefficiente di $a^p b^q$ pari a $(p+1)/q\times(p+2)/(q–1)\times$ ecc. $\times\, n/1$, siccome si suppone che AF diventi AH, ACf =ACH. Che dimostra il punto.

[23] [Bayes esegue un'integrazione per serie, che è possibile essendo finito il numero dei termini. Storicamente, Newton inventò il suo metodo delle flussioni proprio a partire dalla formula del binomio. Tale formula si situa all'intersezione dei due calcoli: infinitesimale e probabilistico. Che sia in gioco lo stesso sapere in entrambi i calcoli? Un sapere sull'infinito, nelle due varianti numerabile e continuo? N.d.T.]

5. Per i punti 1. e 4. il rapporto di ACf dell'intera figura ACFH è $(n+1)\times E\times(x^{p+1}/(p+1) - q\times x^{p+2}/(p+2) + q\times(q-1)/2\times x^{p+3}/(p+3) -$ ecc.) e se, come x esprime il rapporto di Af su AH, così X esprime il rapporto di At su AH. IL rapporto di Aft su ACFH sarà $(n+1)\times E\times X^{p+1}/(p+1) - q\times X^{p+2}/(p+2) + q\times(q-1)/2\times X^{p+3}/(p+3) -$ ecc. Di conseguenza il rapporto di tFCf su ACFH è $(n+1)\times EX^d$ nella differenza delle due serie. Per confronto con la proposizione 10. otteniamo la seguente regola pratica.

REGOLA 1. Se nulla è noto[24] riguardo a un evento ecco che si è verificato p volte e non si è verificato q volte su $p+q$ prove, e da qui congetturo che la probabilità che si verifichi in una singola prova sia compresa tra due gradi di probabilità come X e x, la mia probabilità di indovinare è $(n+1)\times EX^d$ nella differenza tra le due serie $X^{(p+1)}/(p+1) - q\,X^{(p+2)}/(p+2) + q\times(q-1)/2\times X^{(p+3)}/(p+3) -$ ecc. e $x^{(p+1)}/(p+1) - qx^{(p+2)}/(p+2) + q\times(q-1)/2\times x^{(p+3)}/(p+3) -$ ecc., dove E è il coefficiente di $a^p bq$ nell'espansione di $(a+b)^{p+q}$.

Quesrta è la regola appropriata da usare quando q è piccolo. Se q è grande e p è piccolo basta scambiare p con q e q con p e x con r o $1-x$ e X con R o $1-X$, che non modifica le differenze delle due serie.

(*Qui termina il contributo di Bayes*).

[24] [Il moderno soggetto della scienza sa operare con la propria ignoranza. In una regola pratica si parte dal presupposto di ignorare, che è la situazione pratica normale o di base. Il primo effetto dell'ignoranza è la formulazione di una congettura. Il secondo è l'attribuzione ad essa di una probabilità. N.d.T.]