Prefazione

La prima reazione a questo libro potrebbe essere:
Perché un altro libro di probabilità? Perché questo titolo?

Una risposta, non particolarmente originale, è che la probabilità costituisce un linguaggio ed uno strumento tecnico e concettuale ormai ben consolidato la cui importanza difficilmente può essere sopravvalutata. Tanto per non tirarla troppo per le lunghe si può citare il grande J. Clerk Maxwell:

The true logic for this world is the calculus of Probabilities.

Nonostante il calcolo delle probabilità sia presente in quasi tutti i campi della fisica, per nostra esperienza sappiamo che molto spesso gli studenti di fisica presentano gravi lacune (sia tecniche che concettuali) anche su aspetti di base della probabilità. Paradossalmente anche chi si occupa di meccanica statistica a volte non è immune da questi difetti di preparazione. Queste carenze sono dovute, a nostro avviso, all' organizzazione didattica che tipicamente relega la presentazione, spesso frammentaria e che utilizza solo matematica elementare, dei concetti e tecniche di base del calcolo delle probabilità ai primi anni nei corsi di laboratori e qualche cenno nel corso di meccanica statistica. Mentre una trattazione sistematica e più rigorosa è disponibile solo in corsi avanzati (non obligatori) della laurea specialistica, in genere con un taglio matematico o fisico-matematico.

Anche sui punti fondamentali, come la legge dei grandi numeri ed il teorema del limite centrale, è facile imbattersi con idee vaghe (se non errate) sulla reale validità dei risultati. Tra le tante possiamo citare la ridicola affermazione (attribuita a Poincaré[1]) che circola sulla diffusa presenza della funzione gaussiana in molti fenomeni, ammantata di una non necessaria aura di mistero: *Gli sperimentali pensano sia un teorema matematico, mentre i matematici lo credono un fatto sperimentale.*

[1] Ci rifiutiamo di credere che il grande scienziato possa aver detto, se non con intento scherzoso, una tale stupidagine.

Questo diffuso disinteresse per la probabilità tra i fisici è per certi aspetti inspiegabile e suona quasi paradossale in quanto i moderni sviluppi della probabilità sono stati chiaramente ispirati dalla fisica. Anche senza essere esperti di storia delle scienze si può tranquillamente sostenere che nella seconda metà dell' ottocento la vecchia probabilità classica non aveva possibilità di sviluppo, sia per problemi interni, ma soprattutto per mancanza di applicazioni serie. Sono stati proprio gli stimoli provenienti dalla fisica, a cominciare con lo sviluppo della meccanica statistica da parte di J.C. Maxwell e L. Boltzmann, ed il moto browniano (con A. Einstein, M. Smoluchowski e P. Langevin) che hanno permesso lo sviluppo moderno del calcolo della probabilità e la teoria dei processi stocastici.

In fisica il calcolo delle probabilità ha un ruolo centrale e questo per diversi motivi, oltre a quelli ovvi (analisi dei dati) possiamo elencare:
a) Chiarire alcuni aspetti fondamentali della meccanica statistica, ad esempio: il significato matematico degli insiemi statistici, l' importanza dei tanti gradi di libertà coinvolti negli oggetti macroscopici, il principo di massima entropia (tanto spesso citato a sproposito).
b) Districarsi nell'apparente dicotomia tra la descrizione deterministica (in termini di equazioni differenziali) della fisica classica e l' uso di approcci probabilistici.
c) Orientarsi nei problemi di modellizazione di fenomeni "complessi", ad esempio quelli che coinvolgono gradi di libertà con tempi caratteristici molto diversi.

Lo schema del libro è il seguente:
La **Prima Parte** (Capitoli 1, 2 e 3) è costituita da un' *Introduzione generale alla probabilità*. Particolare enfasi è dedicata alla probabilità condizionata, le densità marginali ed i teoremi limite (legge dei grandi numeri, teorema del limite centrale e teoria delle grandi deviazioni). Alcuni esempi sono stati introdotti con lo scopo esplicito di evindenziare come molti risultati della meccanica statistica non sono altro che applicazioni di aspetti generali del calcolo delle probabilità.
Nella **Seconda Parte** (Capitoli 4, 5 e 6) presentiamo i *Concetti fondamentali dei processi stocastici*. Dopo una discussione del moto Browniano, introduciamo le catene di Markov ed i processi stocastici la cui densità di probabilità è regolata dall' equazioni di Fokker- Planck. Due brevi parentesi, sul metodo Montecarlo e l'uso delle equazioni differenziali stocastiche per i modelli climatici, danno un'idea dell' importanza applicativa dei processi stocastici in fisica.
La **Terza Parte** (Capitoli 7, 8 e 9) è una *Selezione di argomenti avanzati*: analisi dei sistemi deterministici caotici in termini probabilistici; generalizzazione del teorema del limite centrale per variabili con varianza infinita (funzioni stabili di Lévy); rilevanza delle distribuzioni non gaussiane nei processi di diffusione. Discutiamo infine alcuni dei molti aspetti dell' entropia, dalla meccanica statistica, alla teoria dell' informazione, al caos deterministico. Inutile

dire che la scelta degli argomenti di questa terza parte è dettata dagli interessi degli autori.
Per completezza in ogni Capitolo abbiamo incluso alcuni esercizi e proposto semplici esperimenti numerici.

L' **Appendice** è divisa in due parti, La prima metà è un sillabo che contiene definizioni e concetti di base, è stata inserita alla fine del libro per non appesantire il testo con uno stile troppo pedante con materiale che per qualche lettore è sicuramente superfluo. Nella seconda parte discutiamo tre argomenti interessanti, anche se in parte un po' a margine della fisica: un' applicazione (tecnicamente elementate ma con conseguenze non banali) della probabilità alla genetica; la statistica degli eventi estremi ed un' applicazione (al limite del lecito) del calcolo delle probabilità alla distribuzione dei numeri primi.

Infine, sono riportate le **soluzioni** degli esercizi.

I prerequisiti richiesti al lettore sono solo la matematica di base a livello universitario, per intendersi derivate, integrali, serie, calcolo combinatorio elementare e trasformate di Fourier.

Il primo ringraziamento è per Luca Peliti che ci ha incoraggiato in questo progetto con suggerimenti e consigli che hanno migliorato il libro. Stefano Berti, Massimo Cencini, Fabio Cecconi, Filippo De Lillo, Massimo Falcioni, Giacomo Gradenigo, Miguel Onorato, Davide Vergni e Dario Villamaina hanno letto parti del testo suggerendo miglioramenti, a loro tutti il nostro grazie. Un ringraziamento particolare ad Alessandro Sarracino che ha scovato molti punti poco chiari, refusi ed errori.

Torino e Roma,
Autunno 2011

Guido Boffetta
Angelo Vulpiani

Indice

1

Un'introduzione

1.1 Un po' di storia: gli albori

Come è consueto iniziamo con un breve excursus storico, ovviamente senza pretese di completezza e rigore metodologico. Nel nostro caso l' introduzione storica non è solo un omaggio ai padri fondatori ma anche un'opportunità per sottolineare alcuni aspetti concettuali.

La teoria della probabilità è tra le discipline matematiche una delle più recenti, formalizzata solo nel XX secolo, ha visto i suoi albori molto tardi (rispetto a branche come la geometria, l'algebra e l'analisi), ed inizialmente per motivi piuttosto frivoli: i giochi d'azzardo. Come origine del calcolo delle probabilità è spesso citato il problema, sollevato dal cavalier A.G. de Méré (un accanito giocatore) e risolto da B. Pascal. Il problema era il seguente: spiegare perché (come da evidenza empirica) puntando sull'uscita di almeno un 6 in 4 lanci di un dado (non truccato) sia più facile vincere che perdere, mentre puntando sull' uscita di almeno un doppio 6 in 24 lanci di una coppia di dadi è più facile perdere che vincere. De Méré era convinto che le probabilità di vittoria nei due casi dovessero essere uguali[1]. La risposta di Pascal fu di calcolare bene le probabilità ed ottenere così il risultato giusto. La soluzione è piuttosto semplice: poiché il dado ha 6 facce equivalenti la probabilità di ottenere 6 in un lancio è $p = 1/6$, la probabilità di avere un risultato diverso da 6 è $q = 1 - p = 5/6$, quindi la probabilità di non avere neanche un 6 in 4 lanci è $q^4 = (5/6)^4$ e la probabilità di avere almeno un 6 in 4 lanci è $1 - q^4 = 1 - (5/6)^4 = 671/1296 \simeq 0.517$.

[1] Per la curiosità del lettore, l'argomento (errato) di de Méré era il seguente: se in un gioco ripetuto la probabilità di vincere in un singolo tentativo è $1/N$, allora ci sarà un numero n^* (proporzionale ad N) tale che se $n > n^*$ la probabilità di vincita puntando su almeno un evento in n prove è maggiore della probabilità di perdita. Nel primo gioco $N = 6$ mentre nel secondo $N = 36$ quindi prendendo rispettivamente $n = 4$ ed $n = 24$ si ha che il rapporto n/N è $4/6 = 24/36$ quindi non sarebbe possibile che nel primo gioco la probabilità di vincere sia maggiore di quella di perdere, mentre nel secondo sia il contrario.

Per il secondo gioco si procede in modo analogo: la probabilità di un doppio 6 nel lancio di una coppia di dadi è $p = 1/36$, quindi la probabilità di avere un risultato diverso è $q = 1 - 1/36 = 35/36$, e la probabilità di avere almeno una coppia di 6 in 24 lanci è $1 - q^4 = 1 - (35/36)^{24} \simeq 0.491$. È notevole il fatto che de Méré, pur modesto matematico, sia stato in grado di valutare dalle osservazioni una differenza di solo il 3%.

La prima opera "seria" (che anticipa uno dei risultati fondamentali del calcolo delle probabilità, cioè la legge dei grandi numeri) è il trattato *Ars conjectandi* (Arte del congetturare) di Jakob Bernoulli, pubblicato postumo nel 1713.

1.1.1 La probabilità come frequenza

La legge dei grandi numeri ci assicura che, in un senso che discuteremo in dettaglio nel Capitolo 3, se P è la probabilità che un certo evento accada (ad esempio esca testa nel lancio di una moneta), allora se si effettuano N prove (indipendenti), indicando con N^* il numero di volte che l'evento accade, allora a "parte eventi rari" si ha che

$$\lim_{N\to\infty} \frac{N^*(N)}{N} = P \ . \tag{1.1}$$

In termini più precisi, per ogni $\epsilon > 0$ la probabilità che $N^*(N)/N$ si discosti più di ϵ da P diventa arbitrariamente piccola al crescere di N:

$$\lim_{N\to\infty} P\Big(\Big|\frac{N^*(N)}{N} - P\Big| > \epsilon\Big) = 0 \ . \tag{1.2}$$

Una proposta che suona naturale è di definire probabilità dell' evento come la sua frequenza nel limite di tante prove. A rigor di logica la cosa non è senza difetti e si potrebbe obiettare che:
a) si ha una circolarità (autorefenzialità) nella (1.2);
b) c'è il problema che la (1.2) non assicura la convergenza di N^*/N a P per tutte le successioni, ma solo per "quasi tutte le successioni", ad esempio nel lancio di una moneta l'uscita della sequenze con N volte testa può ovviamente accadere;
c) quanto deve essere grande N?
d) come si decide che l'approssimazione è buona?

Non è facile rispondere in termini elementari ai precedenti problemi. Risposte, almeno parziali, saranno discusse quando tratteremo i teoremi limite e la teoria delle grandi deviazioni.

1.1.2 La probabilità classica

Nel 1716 A. de Moivre in *Doctrine de Changes* introduce la cosiddetta definizione classica della probabilità: la probabilità di un evento è il rapporto tra il numero di casi favorevoli e quelli possibili, supposto che tutti gli eventi siano equiprobabili[2]. Inoltre de Moivre mostra, in una situazione specifica, un caso particolare del teorema del limite centrale [3].

Un ruolo fondamentale per la probabilità classica è giocato, all'inizio del XIX secolo da P.S. Laplace con il suo *Théorie analytique des probabilités*. Nell' edizione del 1814 è incluso il saggio *Essai Philosophique des probabilites* che contiene il celebre manifesto sul determinismo:
Dobbiamo dunque considerare lo stato presente dell'universo come effetto del suo stato anteriore e come causa del suo stato futuro. Un'intelligenza che, per un dato istante conoscesse tutte le forze di cui è animata la natura e la situazione rispettiva degli esseri che la compongono, se per di più fosse abbastanza profonda per sottomettere questi dati all'analisi, abbraccerebbe nella stessa formula i movimenti dei più grandi corpi dell'universo e dell'atomo più leggero: nulla sarebbe incerto per essa e l'avvenire, come il passato, sarebbe presente ai suoi occhi.
Può sembrare paradossale trovare quella che è considerata la formulazione per antonomasia del punto di vista deterministico, in un'opera dedicata alla probabilità.

Poche pagine dopo Laplace cerca di dare una spiegazione all' apparente contraddizione tra l'esistenza di fenomeni irregolari ed il determinismo:

La curva descritta da una semplice molecola di aria o di vapore è regolata con la stessa certezza delle orbite planetarie: non v'è tra di esse nessuna differenza, se non quella che vi pone la nostra ignoranza. La probabilità è relativa in parte a questa ignoranza, in parte alle nostre conoscenze.

La definizione classica di probabilità, che è basata su eventi discreti, ha evidenti difficoltà nel caso si considerino variabili continue. Tuttavia l'approccio può essere generalizzato, almeno in certe situazioni, e portare alla probabilità geometrica. Per darne un esempio consideriamo il seguente problema: una stanza è pavimentata con piastrelle quadrate di lato L, si lanci una moneta di diametro $d < L$, ci si chiede la probabilità (che, si badi bene, non è ancora stata definita) che la moneta cada a cavallo di almeno 2 piastrelle. Nella Fig.1.1 è mostrata la zona in cui deve cadere il centro della moneta per avere l'evento voluto. È naturale (o almeno sembra) supporre che la probabilità sia il rapporto tra l'area della parte tratteggiata e l'area della piastrella, quindi $p = 1 - (L - d)^2/L^2$. Quindi, nell'ambito della probabilità geometrica si definisce come probabilità il rapporto tra l'area dell'evento favorevole e quella totale [4]. Ovviamente in una dimensione, invece dell' area, si usa la lunghezza

[2] Un minimo di riflessione porta a sospettare che in questa definizione ci sia un punto debole perchè il concetto di equiprobabile è autoreferenziale.

[3] Come notato da Kac e Ulam, per qualche purista il risultato di de Moivre non sarebbe da considerare molto profondo in quanto solo un'applicazione piuttosto semplice di formule combinatorie elementari e dell'approssimazione di Stirling $n! \simeq \sqrt{2\pi n}\, n^n e^{-n}$.

[4] Notare che, in termini moderni, nel problema del lancio della moneta la definizione equivale ad assumere che la densità di probabilità congiunta delle variabili x ed

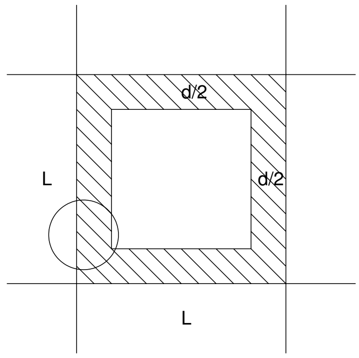

Figura 1.1. Il problema del lancio della moneta. L'area tratteggiata, di larghezza $d/2$, rappresenta la regione in cui deve cadere il centro della moneta affinchè essa sia cavallo di almeno due piastrelle.

ed in tre dimensioni il volume.

A prima vista tutto sembra sensato, purtroppo, come vedremo in seguito, l'idea di fondo della probabilità geometrica nasconde un aspetto sottile che non può essere superato senza un radicale ripensamento del problema su solide basi matematiche.

Come calcolare π lanciando aghi sul pavimento

Un' interessante applicazione della probabilità geometrica è il cosiddetto problema dell'ago di Buffon . Un pavimento è ricoperto di parquet con listelli uguali di larghezza d, un ago di lunghezza $L < d$ è lasciato cadere in modo casuale. Ci si domanda la probabilità che l'ago cada a cavallo tra due listelli. Indichiamo con x la distanza dalla linea di separazione tra due listelli ed il centro dell' ago, e con θ l'angolo che forma l'ago con la retta perpendicolare alla separatrice, vedi Fig.1.2 ed assumiamo che x e θ siano distribuiti in modo uniforme in $(0, d]$ e $(0, \pi/2]$ rispettivamente. Si ha intersezione, vedi Fig.1.2, quando

$$\frac{L}{2} \cos\theta > x \ , \tag{1.3}$$

questo corrisponde all' ago che interseca la linea di separatrice superiore, oppure

$$\frac{L}{2} \cos\theta > d - x \ , \tag{1.4}$$

corrispondente all' intersezione con la linea inferiore. Quindi la probabilità è

y (rispettivamente ascissa ed ordinata del centro della moneta) sia costante, e quindi si ricorre ancora di fatto l'utilizzo del concetto di "equiprobabile" della probabilità classica.

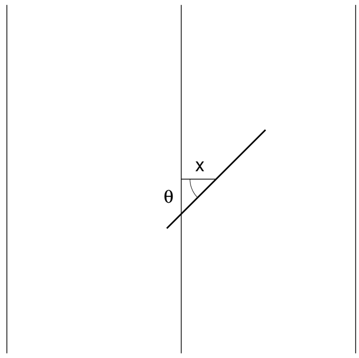

Figura 1.2. Il problema dell'ago di Buffon di lunghezza L lanciato a caso su un pavimento formato da listelli di larghezza d.

data da

$$P = \frac{Area(F)}{Area(E)} ,$$

ove $Area(F)$ e $Area(E)$ sono rispettivamente l'area della regione per la quale vale la (1.3) oppure la (1.4) e l'area totale. Un facile calcolo fornisce

$$P = \frac{4}{\pi d} \int_0^{\frac{\pi}{2}} \frac{L}{2} \cos\theta \, d\theta = \frac{2L}{\pi d} . \tag{1.5}$$

A questo punto, invocando la legge dei grandi numeri si può pensare di calcolare π lanciando un grande numero di volte ($N \gg 1$) un ago e contando il numero di volte (N^*) che l'ago interseca la separatrice si ottiene una buona stima di $P \simeq N^*/N$ e quindi dalla (1.5) una stima di π:

$$\pi \simeq \frac{2LN}{dN^*} .$$

Questo risultato può essere considerato l'antenato del metodo Montecarlo: facendo ricorso alla legge dei grandi numeri ed usando una metodologia stocastica è possibile risolvere un problema (il calcolo di π) che non ha niente di aleatorio.

1.1.3 Il paradosso di Bertrand.

Il linguaggio colloquiale non è sempre in grado di evitare situazioni paradossali, questo è ben evidente nel seguente esempio (dovuto a Bertrand). Si consideri il problema: dato un cerchio di raggio unitario si disegni una corda a caso. Calcolare la probabilità che la lunghezza della corda sia maggiore di

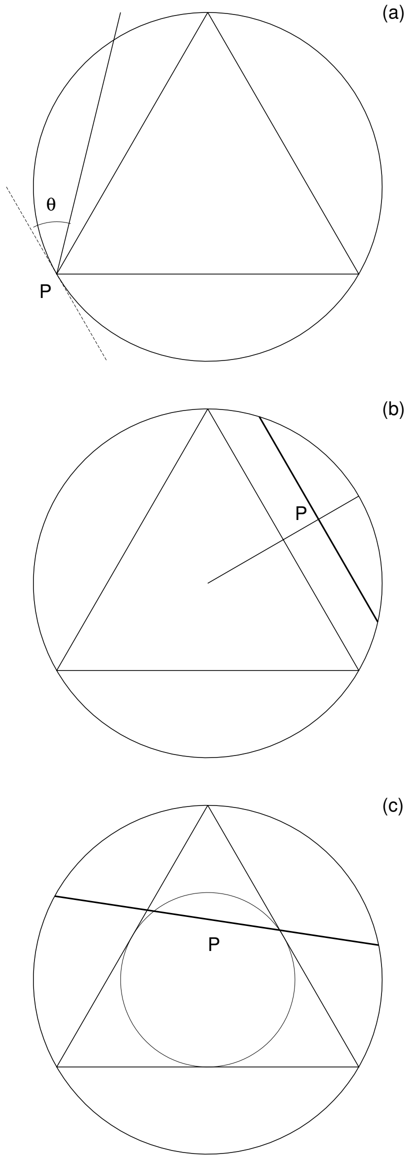

Figura 1.3. Tre possibili soluzioni del problema di Bertrand.

$\sqrt{3}$ (il lato del triangolo equilatero inscritto).

Prima risposta: prendiamo un punto P sul bordo del disco. Tutte le corde che partono da P sono parametrizzate da un angolo θ, vedi Fig.1.3a. Se si vuole che la corda sia più lunga di $\sqrt{3}$ l'angolo θ deve essere compreso in un settore di 60 gradi in un range di 180, quindi la probabilità è $60/180 = 1/3$.

Seconda risposta: consideriamo un punto P su un raggio e la corda passante per P e perpendicolare al raggio, vedi Fig.1.3b. La corda è più lunga di $\sqrt{3}$ se il suo centro P è nella parte interna (di lunghezza $1/2$), quindi poiché il raggio è 1 la probabilità è $1/2$.

Terza risposta: se il centro della corda cade nel disco di raggio $1/2$ allora la corda è più lunga di $\sqrt{3}$, vedi Fig.1.3c, poichè l'area di questo cerchio è $\pi/4$ mentre l' area totale è π, la probabilità è $1/4$.

Qual è la risposta giusta? Semplicemente la domanda è mal posta, perchè "si disegni una corda a caso" è decisamente troppo vago, ed in ognuna delle tre risposte c' è un'assunzione nascosta che sembra naturale, ma è invece arbitraria. Nella prima si è assunto che l'angolo θ sia uniformemente distribuita, nella seconda che il centro della corda sia uniformemente distribuito sul diametro, mentre nella terza che il centro della corda sia uniformemente distribuito all'interno del cerchio.

È chiaro che il paradosso di Bertrand mostra l'ambiguità di alcune idee apparentemente intuitive, che spesso vengono invocate (a sproposito) in ambito fisico. Ad esempio non ha alcun senso, senza qualche specifico argomento dettato dalla fisica od altro, dire *è naturale assumere che una densità di probabilità sia uniforme.*

1.2 La teoria della probabilità diventa una scienza matura

Alla fine del XIX secolo era ormai ben chiaro che il calcolo delle probabilità necessitasse di una profonda e solida sistematizzazione sia tecnica che concettuale. Non è certo un caso che D. Hilbert nel suo celebre discorso al congresso mondiale dei matematici del 1900 a Parigi, nell'elenco dei 23 problemi aperti pone (come sesto) quello di *trattare in modo assiomatico quelle parti delle scienze fisiche in cui la matematica gioca un ruolo importante; in particolare la teoria della probabilità ..... e lo sviluppo rigoroso e soddisfacente del metodo delle medie in fisica matematica, in particolare nella teoria cinetica dei gas.*

L'iniziatore di questo progetto è stato E. Borel che intuì che la teoria della misura di Lebesgue dovesse essere la base matematica della teoria della probabilità. Il programma di formalizzazione può essere considerato concluso nel 1933 con la pubblicazione del libro di A.N. Kolmogorov *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Concetti fondamentali di teoria delle probabilità) [5]. Discutiamo brevemente gli assiomi introdotti da Kolmogorov per formalizzare il calcolo delle probabilità ed il loro significato.

Consideriamo un insieme Ω di eventi elementari ω e sia $\mathcal{F}$ una famiglia di sottoinsiemi di Ω. Chiamiamo Ω spazio degli eventi ed eventi casuali (o semplicemente eventi) gli elementi di $\mathcal{F}$:

I- $\mathcal{F}$ è un'algebra d' insiemi, cioè $\Omega \in \mathcal{F}$, ed $\mathcal{F}$ è chiuso rispetto all'operazione di unione, intersezione e complemento, cioè se $A \in \mathcal{F}$, e $B \in \mathcal{F}$, allora anche $A \cap B$, $A \cup B$ e $\overline{A} = \Omega - A$ sono contenuti in $\mathcal{F}$ [6].

II- Ad ogni elemento A di $\mathcal{F}$ si associa un numero reale non negativo (la probabilità di A) $P(A)$.

III- $P(\Omega) = 1$.

IV- Se due insiemi A e B sono disgiunti (cioè $A \cap B = \emptyset$) allora $P(A \cup B) = P(A) + P(B)$.

La terna $(\Omega, \mathcal{F}, P)$ è detta spazio di probabilità. È un facile esercizio mostrare che

$$P(\overline{A}) = 1 - P(A) \ , \ P(\emptyset) = 0 \ , \ 0 \leq P(A) \leq 1 \ .$$

Discutiamo ora il significato concettuale (ed empirico) dei quattro assiomi di Kolmogorov, cosa importante se si vuole che il calcolo delle probabilità non sia solo una branca della matematica ma sia anche utilizzabile nelle scienze.

Consideriamo un "esperimento" $\mathcal{S}$ che può essere ripetuto un numero di volte praticamente illimitato, e consideriamo un dato gruppo di eventi possibili come risultato del realizzarsi dell' esperimento $\mathcal{S}$.

L'assioma **I** specifica gli "oggetti" per i quali ha senso definire la probabilità. Ad esempio se $\mathcal{S}$ è costituito dal lancio di una coppia di monete distinguibili, allora gli eventi elementari sono le facce visibili delle due monete, quindi $\Omega = \{TT, TC, CT, CC\}$ ove TC indica l'uscita di testa per la prima moneta e croce per la seconda e così via [7].

Le proprietà della probabilità di un evento $P(A)$ devono essere tali che:

a) si è praticamente certi che se $\mathcal{S}$ è ripetuto un numero molto grande di volte ($N \gg 1$) e l'evento A accade M volte allora M/N è molto vicino a $P(A)$;

b) se $P(A)$ è molto piccola allora è praticamente certo che l' evento A non avviene in una singola realizzazione di $\mathcal{S}$.

---

[5] L'opera di Kolmogorov può essere vista come la summa finale che riassume il lungo processo di sistematizzazione e formalizzazione che ha visto impegnati molti matematici, tra i quali (oltre a Borel e Kolmogorov) F.P. Cantelli, M. Fréchet, A. A. Khinchin, P. Levy e M. von Mises.

[6] $B - A$ è l'insieme che contiene gli elementi di B ma non quelli di A, quindi $\overline{A} = \Omega - A$ è costituito dagli elementi non contenuti in A.

[7] "Ovviamente" se le monete non sono truccate si avrà $P(TT) = P(TC) = P(CT) = P(CC) = 1/4$.

Poiché $0 \leq M/N \leq 1$ e per l' evento Ω si ha sempre $M = N$ sono naturali gli assiomi **II** e **III**.

Se A e B sono incompatibili (i.e. A e B sono disgiunti) allora $M = M_1 + M_2$ ove M, M_1 e M_2 sono rispettivamente il numero di volte che accadono gli eventi $A \cup B$, A e B allora $M/N = M_1/N + M_2/N$ che suggerisce l'assioma **IV**.

Nel caso, particolarmente importante, che l'evento elementare ω sia un numero reale allora Ω è la retta numerica reale R, e la scelta naturale per $\mathcal{F}$ sono gli intervalli semiaperti $[a, b)$. È comodo introdurre la funzione di distribuzione:

$$F(x) = P([-\infty, x)) \ ,$$

cioè la probabilità che la variabile aleatoria X sia minore di x, e la densità di probabilità

$$p_x(x) = \frac{dF(x)}{dx} \ ,$$

ovviamente si ha

$$P([a, b)) = \int_a^b p_x(x')dx' \ .$$

La notazione $p_x(x)$ o $p_X(x)$ indica che la densità è relativa alla variabile aleatoria X. A voler essere rigorosi la definizione di densità di probabilità ha senso solo se $F(x)$ è derivabile; tuttavia se accettiamo il fatto che $p_x(x)$ possa essere una funzione generalizzata (ad esempio con delta di Dirac) il problema non si pone[8].

Notiamo inoltre che gli assiomi di Kolmogorov sono perfettamente compatibili con la definizione della probabilità classica e di quella geometrica; inoltre l'insieme degli assiomi non è contraddittorio[9]. Aggiungiamo che Kolmogorov era un convinto frequentista nel senso che pensava che l' interpretazione della probabilità in termini di frequenza fornisse la migliore connessione tra il formalismo matematico e la realtà fisica.

1.2.1 Il concetto di indipendenza

Due eventi A e B sono detti indipendenti se

$$P(A \cap B) = P(A)P(B) \ , \tag{1.6}$$

[8] Se la variabile aleatoria è discreta allora $F(x)$ è costante a tratti. Per il lancio di un dado non truccato abbiamo $F(x) = 0$ per $x < 1$, $F(x) = 1/6$ per $1 \leq x < 2$, $F(x) = 2/6$ per $2 \leq x < 3$, etc, e quindi

$$p_x(x) = \sum_{n=1}^{6} \frac{1}{6}\delta(x - n) \ .$$

[9] Basta considerare il caso in cui l'unico evento possibile è Ω, quindi $\mathcal{F}$ è costituito solo da Ω e $\emptyset$ ed inoltre $P(\Omega) = 1$, $P(\emptyset) = 0$.

più in generale $A_1, A_2, ..., A_N$ sono indipendenti se

$$P(A_1 \cap A_2 \cap ... \cap A_N) = \prod_{k=1}^{N} P(A_k) \ . \tag{1.7}$$

Questa definizione suona piuttosto intuitiva, comunque vista l' importanza del concetto è opportuno rafforzare l'intuizione. La probabilità di $A \cap B$ se A e B sono indipendenti deve essere una funzione solo di $P(A)$ e $P(B)$:

$$P(A \cap B) = F(P(A), P(B)) \ , \tag{1.8}$$

dobbiamo ora determinare la forma di $F(x, y)$. Consideriamo il seguente esperimento: il lancio di una moneta, opportunamente truccata in modo che la probabilità di avere testa sia p, e di un dado con quattro facce numerate da 1 a 4, anche il dado è truccato in modo tale che le facce 1, 2, 3 e 4 appaiono rispettivamente con probabilità p_1, p_2, p_3 e p_4 (ovviamente $p_1+p_2+p_3+p_4 = 1$). Assumiamo che il lancio della moneta e del dado sia due eventi indipendenti e consideriamo l' evento $T \cap (1 \cup 2)$, cioè che venga testa (evento A) e che appaia il lato numerato con 1, oppure quello numerato con 2 (evento B). Dall' assioma **IV** abbiamo $P(B) = p_1 + p_2$ e quindi dalla (1.8) si ha

$$P(T \cap (1 \cup 2)) = F(p, p_1 + p_2) \ , \tag{1.9}$$

ove $p = P(A)$. Poiché $T \cap (1 \cup 2) = (T \cap 1) \cup (T \cap 2)$ ed inoltre gli eventi $T \cap 1$ e $T \cap 2$ sono disgiunti, per l'assioma **IV** e la (1.8) si ha

$$P(T \cap (1 \cup 2)) = F(p, p_1) + F(p, p_2) \, .$$

Quindi $F(x, y)$ deve soddisfare l'equazione

$$F(x, y_1 + y_2) = F(x, y_1) + F(x, y_2) \ . \tag{1.10}$$

A questo punto, notando che $F(1, y) = y$ e $F(x, 1) = x$, assumendo (cosa che sembra naturale) che $F(x, y)$ sia continua in x ed y, dalla (1.10) si ottiene $F(x, y) = xy$.

Un altro argomento per "convincersi" della (1.6): supponiamo che in $N \gg 1$ prove l'evento A avvenga $N(A)$ volte, B avvenga $N(B)$ volte e $A \cap B$ avvenga $N(A \cap B)$ volte. Possiamo scrivere

$$\frac{N(A \cap B)}{N} = \frac{N(A \cap B)}{N(B)} \frac{N(B)}{N} \ ,$$

a questo punto se ha A e B sono indipendenti è sensato assumere che la realizzazione di B non influenzi l'occorrenza di A e quindi per N grandi $N(A \cap B)/N(B)$ non deve essere diverso da $N(A)/N$, ora identificando le frequenze con le probabilità segue la (1.6).

1.2.2 Un altro assioma

Kolmogorov aggiunge un quinto assioma (apparentemente innocente), quello di continuità o **additività numerabile**
V- se $\{A_j\}$, con $j = 1, 2, ...$ è una collezione numerabile di eventi in $\mathcal{F}$ a due a due disgiunti allora

$$P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j) \ .$$

Per la precisione nel libro del 1933 Kolmogorov introdusse un assioma equivalente: se $\{A_j\}$ è una successione decrescente di eventi tali che $A_1 \supseteq A_2 \supseteq ..$ con $\lim_{N\to\infty} \bigcap_{j=1}^{N} A_j = \emptyset$ allora $\lim_{N\to\infty} P(A_N) = 0$.

A questo punto il lettore attento si sarà reso conto della struttura matematica che si nasconde dietro gli assiomi di Kolmogorov: siamo in presenza della teoria della misura con opportuno "travestimento". Ricordiamo infatti, per completezza (e comodità), che una funzione non negativa di A, $\mu(A)$ è chiamata misura se valgono le seguenti proprietà:

Proprietà 1 se $A_1, A_2, ...$ sono insiemi disgiunti e misurabili allora anche la loro unione $A_1 \cup A_2 \cup ...$ è misurabile e

$$\mu(A_1 \cup A_2 \cup ...) = \mu(A_1) + \mu(A_2) + ...$$

Proprietà 2 se A e B sono misurabili e $A \subset B$ allora l'insieme $B - A$ è misurabile e, per la Proprietà 1, si ha $\mu(B - A) = \mu(B) - \mu(A)$

Proprietà 3 un certo insieme E ha misura 1: $\mu(E) = 1$.

Proprietà 4 se due insiemi misurabili sono congruenti (ad esempio sovrapponibili con rotazioni e/o traslazioni) hanno la stessa misura.

M. Kac sintetizzò l'approccio di Kolmogorov con lo slogan *la teoria della probabilità è la teoria della misura più un'anima*. L'anima è la nozione di dipendenza statistica e lo strumento matematico che quantifica questa nozione è la probabilità condizionata.

L'additività numerabile è un'assunzione delicata. Come esplicitamente ammette Kolmogorov è difficilmente possibile spiegare il suo significato empirico in quanto nella descrizione di ogni processo aleatorio sperimentalmente osservabile possiamo ottenere solo degli spazi di probabilità finiti. Con l'assioma **V** (che in teoria della misura corrisponde alla proprietà di σ- additività, o additività numerabile) di fatto decidiamo di limitare (arbitrariamente) la teoria ad una sottoclasse di modelli[10].

[10] L' importanza di assumere, o meno, l'assioma **V** è nota nell' ambito della teoria della misura. Per esempio G. Vitali nel 1905 fornì un esempio di sottoinsieme della retta reale che non è misurabile rispetto a nessuna misura che sia positiva, invariante per traslazioni e σ-additiva (in particolare la misura di Lebesgue). Per la costruzione dell'insieme di Vitali è indispensabile l'assioma della scelta (di Zermelo), un aspetto delicato e controverso della teoria degli insiemi. Ad esempio per Lebesgue l'esistenza dell'insieme non misurabile trovato da Vitali non è accettabile per gli "empiristi" che rigettano l'assioma della scelta.

1.3 Probabilità e mondo reale

Aldilà degli aspetti tecnici è naturale chiedersi perchè un fisico dovrebbe occuparsi di probabilità. In fondo lanciare dadi o monete su pavimenti piastrellati non sono cose propriamente eccitanti anche se magari non banali.

Ci sono almeno tre situazioni interessanti in fisica in cui il ricorso al calcolo delle probabilità sembra inevitabile [11]:

a) Il numero dei gradi di libertà coinvolti nel problema è elevato e non si è interessati ai dettagli del sistema ma solo al comportamento collettivo di poche variabili. Questo è il caso della meccanica statistica (vedi Capitoli 7 e 9).

b) Non si ha un controllo completo delle condizioni iniziali del sistema che, pur deterministico, ha un comportamento "irregolare" (instabile), questo è il caso dei sistemi con caos deterministico che può manifestarsi indipendentemente dal numero dei gradi di libertà (vedi Capitolo 7).

c) In un fenomeno "complesso" [12] non si ha il controllo delle "tante cause in gioco", l'esempio paradigmatico è il moto Browniano in cui una particella colloidale (molto più grande delle molecole del fluido in cui è immersa) si muove in modo irregolare. In presenza di una netta separazione di scala tra il tempo caratteristico delle molecole e quello della particella colloidale, è possibile descrivere il fenomeno con un processo stocastico (equazione di Langevin, vedi i Capitoli 4 e 6).

I casi sopraelencati hanno un elemento comune: per qualche motivo non si ha una conoscenza completa del fenomeno e ci si "accontenta" di una descrizione non dettagliata, cioè non dello stato completo del sistema ma di una sua proiezione o di descrizione a "grana grossa"[13]. Questo modo di concepire la probabilità in fisica è molto antica e la si ritrova anche in Laplace che sosteneva che la probabilità è relativa alla nostra ignoranza.

Usando la terminologia della filosofia della scienza si può dire che si è di fronte ad una concezione epistemica della probabilità, che è vista come qualcosa meramente legato alla "nostra ignoranza" e non alla natura intrinseca dei sistemi classici (che sono deterministici) [14].

Non è questa la sede per una dettagliata analisi epistemologica. Per quei lettori che si preoccupassero del significato apparentemente "negativo" di epistemico (in quanto non legato alla vera natura del sistema, ma solo alla limitate capacità umane) facciamo notare che epistemico non è affatto sinonimo di soggettivo, inoltre spesso il livello ontico e quello epistemico si possono incrociare in modo sottile. Come esempio possiamo citare il caso dei sistemi

[11] In tutto il libro non tratteremo fenomeni quantistici che sono descritti da un formalismo in cui la probabilità gioca un ruolo fondamentale, anche più che in ambito classico.

[12] La terminologia è volutamente vaga.

[13] In inglese si usa il termine *coarse graining*.

[14] Al contrario, almeno nell' interpretazione ortodossa della scuola di Copenhagen, in meccanica quantistica la probabilità ha un carattere ontico, cioè intrinseco al fenomeno e non dipendente dalla mancanza di informazione dell' osservatore.

dinamici deterministici caotici, la cui natura deterministica è sicuramente una proprietà ontica in quanto intrinseca al sistema ed indipendente dalla bravura dello scienziato che studia il sistema. Al contrario la limitata predicibilità, e più in generale l'essere caotico, è da considerarsi una proprietà epistemica, infatti nel calcolo degli esponenti di Lyapunov e dell' entropia di Kolmogorov-Sinai si fà ricorso ad una descrizione non arbitrariamente accurata dello stato iniziale (ad esempio per l'entropia di Kolmogorov-Sinai si introduce una partizione dello spazio delle fasi in celle di grandezza finita). Ma non per questo il caos è un fenomeno non oggettivo[15].

Spesso i termini "probabilità", "risultato statistico" e simili sono intesi come qualcosa di vago ed impreciso, opposti alla certezza. La speranza delle certezze assolute (in scienze diverse dalla matematica), se mai c'è veramente stata (a parte in qualche vulgata semplicistica) è tramontata da tempo. Scriveva J.C. Maxwell nel 1873:

Il fatto che dagli stessi antecedenti seguano le stesse conseguenze è una dottrina metafisica. Nessuno può negarlo. Ma non è molto utile nel mondo in cui viviamo, ove non si verificano mai gli stessi antecedenti e nulla accade identico a se stesso due volte. Infatti, per quanto possiamo saperne, uno degli antecedenti potrebbe essere la data precisa e la località dell'evento, in questo caso la nostra esperienza sarebbe del tutto inutile. [...] L'assioma della fisica che ha, in un certo senso, la stessa natura è "che da antecedenti simili seguono consequenze simili". Ma qui siamo passati da uguaglianza a somiglianza, dall'accuratezza assoluta ad una piú o meno rozza approssimazione.

Il ricorso ad un approccio probabilistico sembra quindi inevitabile: l'austero Maxwell divenne il paladino della teoria delle probabilità una scienza che, nata per motivi *futili ed immorali* (nelle parole dello stesso Maxwell), è diventata la guida necessaria per l'interpretazione del mondo.

Si potrebbe osservare che abbiamo parlato di probabilità ma senza una definizione inattaccabile ed esplicita. La definizione classica e quella geometrica hanno i loro problemi, anche il punto di vista in termini di frequenze di serie molto lunghe non sembra immune da critiche. Nell' approccio assiomatico di Kolmogorov semplicemente la probabilità non è definita, ma vengono solo enunciate le proprietà che le probabilità devono avere.

Per quanto riguarda la possibile definizione, consideriamo le seguenti domande in linguaggio comune che contengono un riferimento al concetto di probabilità:
a) qual è la probabilità che la Roma vinca lo scudetto nel prossimo campionato di calcio?
b) qual è la probabilità che il governo cada la prossima settimana?
c) qual è la probabilità di avere sempre testa in 50 lanci di una moneta?

[15] Indicando con ϵ la taglia delle celle della partizione, l' entropia di Kolmogorov-Sinai, che misura il grado di impredicibilità del sistema, diventa indipendente da ϵ nel limite di piccoli ϵ (vedi Capitolo 9).

d) qual è la probabilità che una molecola di elio in un gas a temperatura 400 Kelvin e pressione di un'atmosfera abbia una velocità superiore a $100\,m/s$?

Non c'è un unanime consenso sul fatto che le quattro domande siano ben poste. In a) e b) si hanno eventi singoli (non ripetibili) e quindi si deve escludere l'interpretazione di probabilità in termini di frequenze e cercare di definire la probabilità come "grado di fiducia" che un individuo nutre. Nell'approccio soggettivista, sviluppato soprattutto da B. de Finetti, è proposta una definizione di probabilità applicabile ad esperimenti casuali i cui eventi elementari non siano ritenuti ugualmente possibili e che non siano necessariamente ripetibili un numero arbitrario di volte sotto le stesse condizioni. La probabilità di un evento viene definita come il prezzo che un individuo razionale ritiene giusto pagare per ricevere 1 se l'evento si verifica, 0 se l'evento non si verifica. Per rendere concretamente applicabile la definizione, si aggiunge un criterio di coerenza: le probabilità devono essere attribuite in modo tale che non sia possibile ottenere una vincita o una perdita certa. Nell' ambito dell'approccio soggettivista tutte le affermazioni a), b), c) e d) hanno senso.

Al contrario Kolmogorov è molto netto nel sostenere che non per tutti gli eventi si può definire una probabilità *L'assunzione che una definita probabilità esiste per un dato evento sotto certe condizioni è un'ipotesi che deve essere verificata e giustificata in ciascun caso individuale.*

Una volta che si è deciso che ha senso parlare di probabilità dell'evento X, rimane il problema di rispondere alla domanda *quanto vale la probabilità che accada l'evento X ?* Caso per caso si deve cercare una risposta. Nell'esempio paradigmatico della moneta non truccata l'ipotesi che la probabiltà che esca testa è $1/2$ segue da ovvie considerazioni di simmetria, che, ovviamente, non sempre sono possibili[16].

Noi adottiamo il punto di vista per cui solo domande del tipo c) e d) siano scientificamente rilevanti, assumendo (almeno come ipotesi di lavoro) la validità dell' interpretazione frequentistica. Pur con tutti i suoi limiti questo modo di intendere la probabilità ha un grande vantaggio fisico: le frequenze contengono un' informazione obiettiva, e sperimentalmente verificabile, che non ha niente a che fare con le nostre credenze o conoscenze riguardo la natura.

Riportiamo due citazioni, che condividiamo in pieno:
Uno dei compiti più importanti della teoria delle probabilità è di identificare quegli eventi la cui probabilità è vicino a zero od a uno (A.A. Markov);
Tutto il valore epistemologico della teoria delle probabilità è basato su questo: i fenomeni aleatori, considerati nella loro azione collettiva a grande scala, generano una regolarità non aleatoria (B.V. Gnedenko e A.N. Kolmogorov).

Un aspetto molto importante per la fisica è la stretta relazione tra il punto di vista frequentistico ed i fondamenti della meccanica statistica (in particolare

[16] Come determinare la probabilità che esca testa se la moneta è truccata? È possibile farlo evitando il ricorso al metodo "empirico" in cui si ricorre a tanti lanci?

il problema ergodico). Per L. Boltzmann (ed anche A. Einstein) la probabilità di un evento non è altro che la percentuale di tempo nella quale si ha l'evento, idea questa che permette di definire in modo non ambiguo la probabilità anche in un singolo sistema, senza invocare gli insiemi statistici.

Concludiamo il capitolo con una brevissima discussione sul fatto che le asserzioni probabilistiche non sarebbero falsificabili. Su questo Popper è molto netto:

Le stime probabilistiche **non** *sono falsificabili. E, naturalmente non sono neppure verificabili ... i risultati sperimentali, per quanto numerosi e favorevoli, non potranno mai stabilire in modo definitivo che la frequenza relativa "testa" è* 1/2 *e sarà sempre* 1/2.

Alla lettera queste affermazioni sono ovviamente vere anche se viene spontaneo domandarsi:

a) veramente si fà ricerca per falsificare le teorie?

b) qualcuno può sostenere, in buona fede, che la distribuzione di Maxwell-Boltzmann per le velocità delle molecole di un gas (un' asserzione probabilistica), non sia stata "verificata" (direttamente od indirettamente) oltre ogni ragionevole dubbio?

c) è vero che non sapremo mai il valore "esatto" della probabilità di un certo evento, ma questo è un fatto che non sconvolge più di tanto: come si può essere sicuri che la lunghezza di un oggetto sia proprio 22.33 centimetri?

d) i problemi sollevati da Popper sulla probabilità non sono comuni a tutte le scienze? non siamo di fronte al vecchio problema (irrisolubile) della giustificazione dell'induzione?

Stiamo inesorabilmente scivolando verso la questione troppo generale della connessione tra le teorie da noi costruite, l'interpretazione delle osservazioni empiriche e la realtà fisica. In accordo con Reichenbach a noi sembra accettabile che, se crediamo che le asserzioni sul mondo fisico abbiano significato, in modo altrettanto sicuro possiamo credere nel significato del concetto di probabilità. Per ora non troviamo scappatoia migliore che una citazione di A. Einstein:

Quando le leggi della Natura si riferiscono alla realtà non sono certe. Quando sono certe non si riferiscono alla realtà.

Esercizi

1.1. Dimostrare che:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{E.1}$$

$$P(A \cup B \cup C) = \tag{E.2}$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C) \ .$$

1.2. Dati due eventi A e B tali che $P(A) = 3/4$ e $P(B) = 1/3$ mostrare che $P(A \cap B) > 1/12$.

1.3. Mostrare che non possono esistere due insiemi A e B tali che $P(A) = 4/10$, $P(B) = 3/10$ e $P((\Omega - A) \cap (\Omega - B)) = 2/10$.

1.4. Trovare le probabilità di vincita al lotto per il singolo estratto, l'ambo, il terno etc. Si ricordi che nel gioco del lotto si estraggono 5 numeri da un contenitore che ne contiene 90.

Letture consigliate

Per la storia del moderno calcolo delle probabilità:

J. von Plato *Creating Modern Probability* (Cambridge University Press, 1994).

Per una breve discussione su probabilità e teoria della misura:

M. Kac and S. Ulam *Mathematics and Logic* (Dover Publications, 1992).

Lettura obbligata per approfondire gli aspetti concettuali:

A.N. Kolmogorov *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933); traduzione inglese *Foundations of the Theory of Probability* (Chelsea Publ. Comp. 1956), consultabile su: `http://www.kolmogorov.com/Foundations.html`

Per un' introduzione alle varie interpretazioni della probabilità:

A. Hájek " Interpretations of Probability", in *Stanford Encyclopedia of Phylosophy*, consultabile su:
`http://www.seop.leeds.ac.uk/entries/probability-interpret`