

Note introduttive sulla teoria delle probabilità.

M. Falcioni, A. Vulpiani

Un indispensabile strumento matematico per affrontare lo studio della Meccanica Statistica è il calcolo delle probabilità. Lo scopo di queste note è di aiutare lo studente ad apprendere (o ripassare) le nozioni di base.

1 Introduzione

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 (un minimo di riflessione mostra che in questa definizione c'è un punto debole perchè il concetto di equiprobabilità è autoreferenziale).

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 esempio si consideri il seguente problema: una stanza è pavimentata con piastrelle quadrate di lato L , si lancia 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. La Fig. 1 mostra la zona, di profondità $d/2$ sotto ai lati della piastrella, in cui deve cadere il centro della moneta perché si realizzi l'evento voluto. È naturale (o almeno sembra) supporre che la probabilità sia il rapporto tra l'area di questa zona e l'area della piastrella, cioè $p = 1 - (L - d)^2/L^2$. Quindi, nell'ambito della probabilità geometrica si definisce come probabilità il rapporto tra l'area relativa all'evento favorevole e quella totale. Ovviamente in una dimensione, invece dell'area, si usa la lunghezza ed in tre dimensioni il volume.

A prima vista tutto sembra sensato, purtroppo l'idea di fondo della probabilità geometrica nasconde degli aspetti sottili che non possono essere superati senza un ripensamento del problema su solide basi matematiche. Un esempio dei problemi che possono nascere da una definizione "ingenua" di probabilità, quando si hanno eventi continui è il seguente, dovuto a Bertrand.

Il paradosso di Bertrand – Si consideri il problema: dato un cerchio

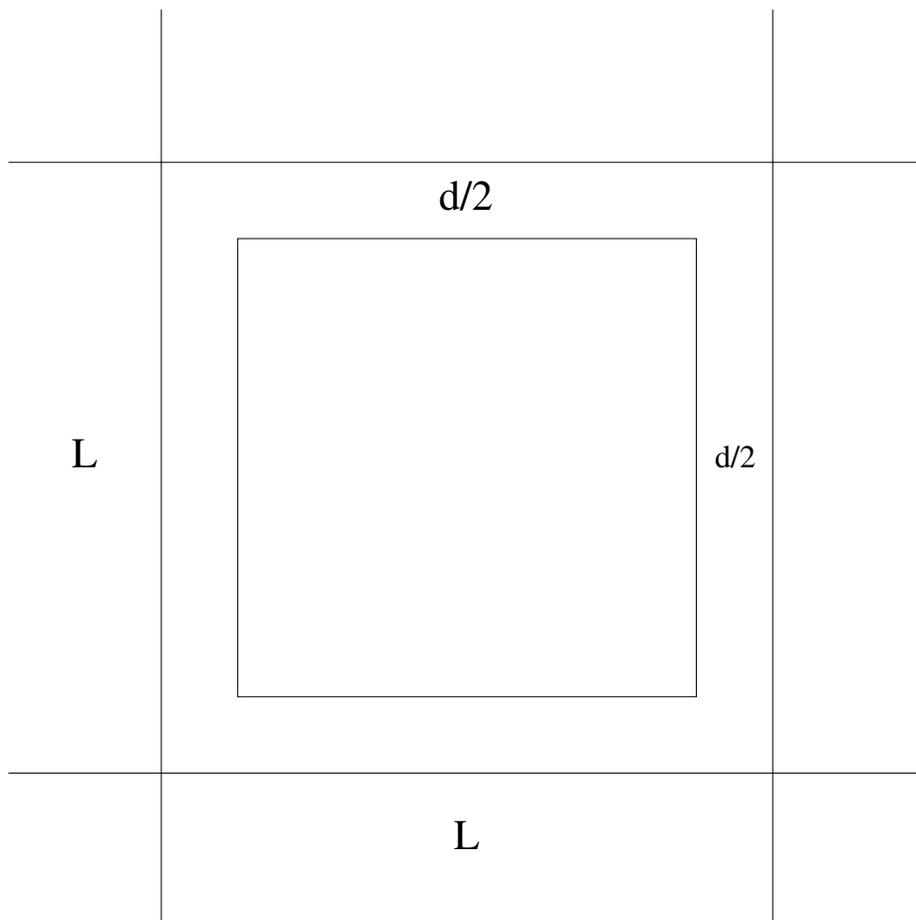


Figure 1: lancio di una moneta di raggio d su una mattonella quadrata di lato L : la regione degli eventi favorevoli.

di raggio unitario si disegni una corda a caso. Calcolare la probabilità che la lunghezza della corda sia maggiore di $\sqrt{3}$ (il lato del triangolo equilatero iscritto).

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. 2. 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 intervallo di 180, quindi la probabilità

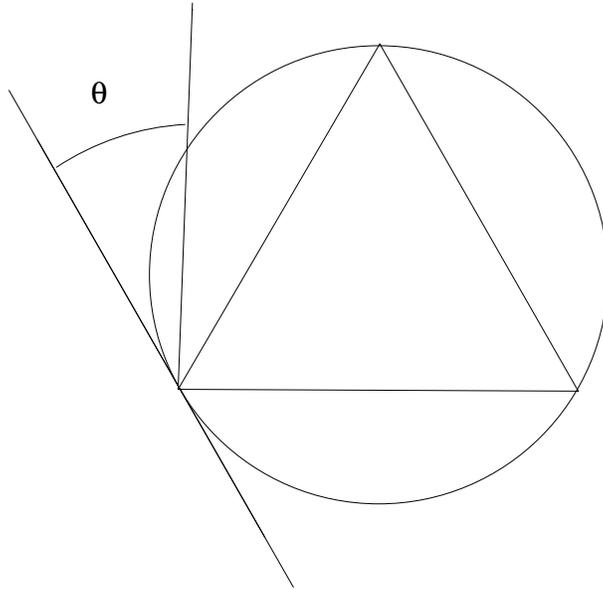


Figure 2: Lunghezza di una corda tracciata a caso in un cerchio: prima risposta.

è $60/180 = 1/3$.

Seconda risposta: consideriamo un punto P su un diametro, e la corda passante per P e perpendicolare al diametro, vedi Fig. 3. La corda è più lunga di $\sqrt{3}$ se il suo centro P è nella parte centrale (di lunghezza 1), quindi poiché il diametro è 2 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. 4, poiché l'area di questo cerchio è $\pi/4$ mentre l'area totale è π la probabilità è $1/4$.

Qual è la risposta giusta? In realtà la domanda è mal posta: “si disegni una corda a caso” è decisamente troppo vago cosicché si possono fare assunzioni diverse, che possono sembrare “naturali”, ma sono arbitrarie e conducono a risposte diverse. Nella prima risposta si è assunto che θ sia uniformemente distribuita, nella seconda che il centro della corda sia uniformemente distribuito sul diametro, mentre nella terza che il centro della corda

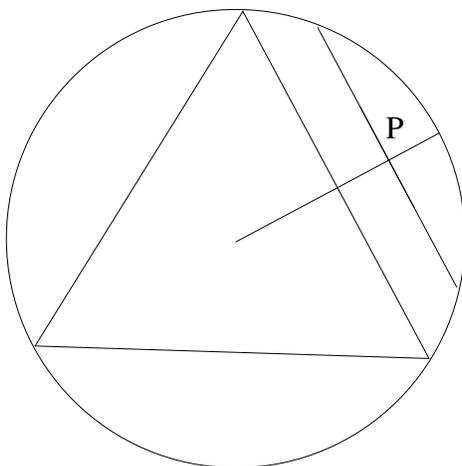


Figure 3: Lunghezza di una corda tracciata a caso in un cerchio: seconda risposta.

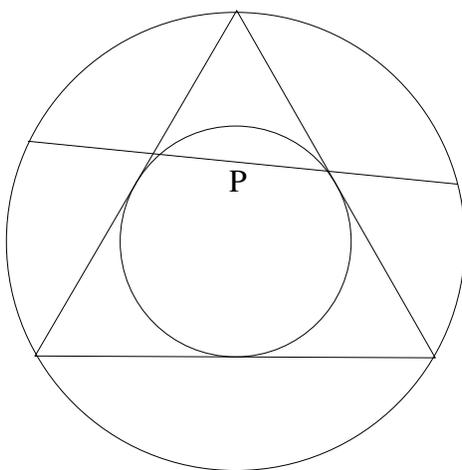


Figure 4: Lunghezza di una corda tracciata a caso in un cerchio: terza risposta.

sia uniformemente distribuito all'interno del cerchio.

2 Un approccio sistematico

È chiaro che il paradosso precedente getta una luce sinistra su alcune idee apparentemente intuitive, che spesso vengono invocate (a sproposito) in ambito fisico. Ad esempio è molto discutibile dire *è naturale assumere che una densità di probabilità (di una qualche variabile) sia uniforme*, senza qualche specifico argomento dettato dalla fisica o altro,

L' iniziatore della sistematizzazione sia tecnica che concettuale della teoria delle probabilità 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) ¹.

2.1 Gli assiomi

Discutiamo brevemente gli assiomi introdotti da Kolmogorov e 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 $\bar{A} = \Omega - A$ sono contenuti in \mathcal{F} ².

II- Ad ogni elemento A di \mathcal{F} si associa un numero reale non negativo (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 (Ω, \mathcal{F}, P) è detta spazio di probabilità. È un facile esercizio

¹L' opera di Kolmogorov può essere vista come la summa finale di un lungo processo 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.

² $B - A$ è l'insieme che contiene gli elementi di B ma non quelli di A , quindi $\bar{A} = \Omega - A$ è costituito dagli elementi non contenuti in A .

mostrare che

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

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

Le considerazioni che seguono tengono conto dell'idea, che si era sviluppata a partire dalla legge dei grandi numeri, di definire la probabilità di un evento come la sua frequenza nel limite di tante prove ripetute (v. Sez. **8.1**).

Assumiamo dato un certo complesso \mathcal{S} di condizioni che ammettono un numero praticamente illimitato di ripetizioni, e si studi un dato gruppo di eventi possibili come risultato del realizzarsi delle condizioni \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 ³.

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} .

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:

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

³"Ovviamente" se le monete non sono truccate si avrà $P(TT) = P(TC) = P(CT) = P(CC) = 1/4$.

cioè la probabilità che l'evento che si realizza abbia valore minore di x , e la densità di probabilità

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

ovviamente si ha

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

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 ⁴.

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 ⁵. 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.

2.2 Il concetto di indipendenza

Due eventi A e B sono detti indipendenti se

$$P(A \cap B) = P(A)P(B) , \tag{1}$$

più in generale A_1, A_2, \dots, A_N sono indipendenti se

$$P(A_1 \cap A_2 \cap \dots \cap A_N) = \prod_{k=1}^N P(A_k) . \tag{2}$$

⁴Se gli eventi elementari costituiscono un sottoinsieme discreto dei reali 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

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

⁵Basta considerare il caso in cui l'unico evento possibile è Ω , quindi \mathcal{F} è costituito solo da Ω e \emptyset ed inoltre $P(\Omega) = 1$, $P(\emptyset) = 0$.

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 $P(B)$:

$$P(A \cap B) = F(P(A), P(B)) , \quad (3)$$

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 dia luogo a due eventi indipendenti e consideriamo l'evento $T \cap (1 \cup 2)$, cioè che venga testa e che appaia il lato numerato con 1, oppure quello numerato con 2. Dall'assioma **IV**, applicato all'evento $(1 \cup 2)$ e dalla formula (3) si ha

$$P(T \cap (1 \cup 2)) = F(p, p_1 + p_2) . \quad (4)$$

D'altra parte, poiché $T \cap (1 \cup 2) = T \cap 1 + T \cap 2$ e gli eventi $T \cap 1$ e $T \cap 2$ sono disgiunti ancora per l'assioma **IV** e la (3) si ha

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

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

$$F(x, y_1 + y_2) = F(x, y_1) + F(x, y_2) . \quad (5)$$

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 (5) si ottiene $F(x, y) = xy$.

Un altro argomento per "convincersi" della (1): 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 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).

2.3 Un altro assioma

Kolmogorov aggiunge un quinto assioma (apparentemente innocente), quello di continuità o additività numerabile

V- se $\{A_j\}$, con $j = 1, 2, \dots$ è una collezione numerabile di eventi in \mathcal{F} a due a due disgiunti allora

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

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

A questo punto possiamo osservare che la struttura matematica dietro agli assiomi di Kolmogorov è quella della teoria della misura, con opportuno "travestimento". M. Kac sintetizzò l'approccio di Kolmogorov con lo slogan *la teoria della probabilità è 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.

Per completezza (e comodità) ricordiamo che una funzione non negativa di A , $\mu(A)$ è chiamata misura se valgono le seguenti proprietà:

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

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

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 hanno la stessa misura.

3 Probabilità condizionata

Uno dei concetti più importanti di tutto il calcolo delle probabilità è sicuramente quello di probabilità condizionata. Saperla utilizzare permette di non cadere in insidiosi tranelli (spesso presentati come paradossi): possiamo dire che il primo livello di comprensione del calcolo delle probabilità può dirsi raggiunto se si è in grado di utilizzare correttamente la probabilità condizionata.

Se $P(B) > 0$ allora la probabilità di A condizionata a B è:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (6)$$

La motivazione della formula precedente si può capire facendo ricorso all'interpretazione classica della probabilità: sia M il numero dei possibili risultati ed indichiamo con M_A , M_B e M_{AB} il numero di quelli in cui si realizzano gli eventi A , B e $A \cap B$ rispettivamente, allora

$$P(A|B) = \frac{M_{AB}}{M_B} = \frac{M_{AB}}{M} \frac{M}{M_B}$$

poiché $P(B) = N_B/M$ e $P(A \cap B) = M_{AB}/M$ si ha la (6).

Come esempio consideriamo il lancio di un dado non truccato: sia B l'insieme dei numeri dispari $B = \{1, 3, 5\}$ ed A il numero 1, la probabilità di avere 1 sapendo che il risultato è dispari è

$$P(A|B) = P(1|B) = \frac{1}{3},$$

se invece A è il numero 2 si ha

$$P(A|B) = P(2|B) = 0,$$

in accordo con l'intuizione.

Notare che se A e B sono indipendenti allora $P(A|B) = P(A)$ (vale anche il viceversa), in questo caso sapere che è avvenuto l'evento B non cambia le informazioni su A .

Scrivendo nella (6) $P(A \cap B) = P(B|A)P(A)$ è immediato ottenere:

$$P(B|A) = P(A|B) \frac{P(B)}{P(A)}, \quad (7)$$

la relazione precedente è detta formula di Bayes.

Ricaviamo un paio di utili risultati in cui la probabilità condizionata gioca un ruolo rilevante

Teorema della probabilità completa Se $B_i \cap B_j = \emptyset$ per $i \neq j$ e $\cup_{i=1}^N B_i = \Omega$ allora

$$P(A) = \sum_{i=1}^N P(A|B_i)P(B_i) . \quad (8)$$

Teorema di moltiplicazione

$$P(A_1 \cap A_2 \cap \dots \cap A_N) = \quad (9)$$

$$P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \dots P(A_N|A_{N-1} \cap A_{N-2} \cap \dots \cap A_2 \cap A_1)$$

Per dimostrare la (8) basta osservare che gli eventi $C_i = A \cap B_i$ sono indipendenti quindi $P(\cup_i C_i) = \sum_i P(C_i)$, inoltre, poiché $\cup_i B_i = \Omega$ si ha $\cup_i C_i = A$, quindi $P(A) = \sum_i P(C_i)$ a questo punto dall'identità $P(C_i) = P(A \cap B_i) = P(B_i)P(A|B_i)$ segue (8).

La (9) si ottiene utilizzando ripetutamente la definizione di probabilità condizionata:

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_N) &= P(A_N|A_{N-1} \cap A_{N-2} \cap \dots \cap A_2 \cap A_1)P(A_{N-1} \cap A_{N-2} \cap \dots \cap A_2 \cap A_1) \\ &= P(A_N|A_{N-1} \cap A_{N-2} \cap \dots \cap A_2 \cap A_1)P(A_{N-1}|A_{N-2} \cap \dots \cap A_2 \cap A_1)P(A_{N-2} \cap \dots \cap A_2 \cap A_1) \end{aligned}$$

e così via.

Come esercizio, discutiamo *Il problema dei compleanni* che si risolve facilmente con l' aiuto della (9): date N persone ($N < 365$) determinare la probabilità che almeno 2 siano nate lo stesso giorno (ignoriamo gli anni bisestili ed assumiamo che le nascite siano indipendenti ed uniformemente distribuite nell'arco dell'anno). Indichiamo con A l'evento *almeno 2 persone sono nate lo stesso giorno*, e con \bar{A} l'evento complementare *non esistono coppie di persone nate lo stesso giorno*, allora

$$P(A) = 1 - P(\bar{A}) ,$$

indichiamo con A_1 l'evento *la seconda persona non è nata nello stesso giorno della prima*, con A_2 l'evento *la terza persona non è nata nello stesso giorno della prima e della seconda* e così via, dal teorema di moltiplicazione abbiamo

$$P(\bar{A}) = P(A_1)P(A_2|A_1)\dots P(A_{N-1}|A_1 \cap \dots \cap A_{N-2})$$

un momento di riflessione convince che

$$P(A_1) = \left(1 - \frac{1}{365}\right), \quad P(A_2|A_1) = \left(1 - \frac{2}{365}\right),$$

$$P(A_3|A_1 \cap A_2) = \left(1 - \frac{3}{365}\right) \dots$$

$$P(A_{N-1}|A_1 \cap \dots \cap A_{N-2}) = \left(1 - \frac{N-1}{365}\right).$$

Quindi

$$P(A) = 1 - \prod_{j=1}^{N-1} \left(1 - \frac{j}{365}\right). \quad (10)$$

Una formula approssimata si può ottenere notando che

$$\begin{aligned} \prod_{j=1}^{N-1} \left(1 - \frac{j}{365}\right) &= \exp \sum_{j=1}^{N-1} \ln \left(1 - \frac{j}{365}\right) \\ &\simeq \exp - \sum_{j=1}^{N-1} \frac{j}{365} = \exp - \frac{N(N-1)}{730}. \end{aligned} \quad (11)$$

Dalla (2.5) per $N = 5$ si ha $P = 0.027$, per $N = 10$, $P = 0.117$; $N = 20$, $P = 0.411$; $N = 22$, $P = 0.476$; $N = 23$, $P = 0.507$; $N = 60$, $P = 0.994$; $N = 64$, $P = 0.997$. Quindi già con 23 persone si ha una probabilità maggiore di 1/2 che almeno due persone abbiano lo stesso compleanno, con 60 si ha la “quasi certezza”. Un risultato non così intuitivo. L' approssimazione (11) è già buona per N dell'ordine di 20, ad esempio per $N = 23$ usando (11) si ha 0.500 invece di 0.502.

3.1 Usare la probabilità condizionata

Molti di quelli che vengono spesso presentati come paradossi del calcolo delle probabilità nascono dalla mancata comprensione del concetto di probabilità

condizionata.

A) L'esempio più banale è quello dei numeri ritardatari al gioco del lotto. Su giornali e reti televisive è spesso data grande rilevanza al fatto che un certo numero (diciamo il 21) non esce su una data ruota (ad esempio Roma) da un grande numero di estrazioni (diciamo 150). La conclusione (errata) è che alla prossima estrazione l'uscita del 21 dovrebbe essere "quasi sicura" in quanto "è difficile che un numero non esca per 151 volte di seguito". L'ultima affermazione è sicuramente vera, ma non è interessante per il problema che ci si pone, in quanto non bisogna confondere:

- P , probabilità dell'evento E : "il 21 non esce 151 volte di seguito"

con

- \tilde{P} , probabilità dell'evento \tilde{E} : "il 21 non esce dopo che non è uscito 150 volte".

La probabilità che interessa è $1 - \tilde{P}$ (dell'evento complementare di \tilde{E}) non $1 - P$ (dell'evento complementare di E).

Poiché la probabilità del singolo estratto è $1/18$, e le estrazioni sono indipendenti, nel primo caso si ha

$$P = \left(1 - \frac{1}{18}\right)^{151} = \left(\frac{17}{18}\right)^{151} \simeq 0.000178 ,$$

da cui si ricava $1 - P \simeq 0,999822$, numero che rappresenta la probabilità che in 151 estrazioni il 21 esca *una o più* volte, evento che *contiene anche* quello che interessa, oltre a tanti altri. Infatti per il calcolo di \tilde{P} sapere che il 21 non è uscito 150 volte è irrilevante (in quanto le estrazioni sono indipendenti), quindi la probabilità che il 21 esca nella 151-ma estrazione resta $1/18 \simeq 0,055555 \ll 1 - P$, la stessa che ha in ogni estrazione.

B) Negli anni 90 J. Tooby e L.Cosmides (due influenti ricercatori di psicologia) in un interessante esperimento posero ad un gruppo di medici e studenti di medicina dell'università di Harvard la seguente domanda

Una malattia ha un tasso di incidenza di 1/1000. Esiste un test che permette di individuarne la presenza. Questo test ha un tasso di falsi positivi

del 5%. Un individuo si sottopone al test. L'esito è positivo. Qual è la probabilità che l'individuo sia effettivamente malato?

La risposta esatta, che si ottiene facilmente dalla formula di Bayes, è circa 2%. Solo il 18% dei partecipanti al test diede la risposta esatta e ben il 58% rispose che la probabilità era del 95%. Il fatto che la maggioranza abbia dato come risposta (errata) $P(M|p) = 95\%$ è “comprensibile”: l'argomento (sbagliato) seguito sarebbe il seguente. *In una popolazione di 100000 individui ci sono circa 100 positivi, ma c'è un errore del 5% quindi il numero dei malati veri e quindi la probabilità cercata è circa 95%.*

Ecco la soluzione: indichiamo con $P(M) = 0.001$ la probabilità di essere malato, con $P(S) = 1 - P(M) = 0.999$ la probabilità di essere sano, $P(p|S) = P_e = 0.05$ è la probabilità un falso positivo cioè di risultare positivo essendo sano, e $P(n|M)$ la di probabilità di risultare negativo essendo malato, per semplicità assumiamo $P(n|M) = P(p|S) = P_e$. La probabilità cercata è $P(M|p)$, usando la formula di Bayes:

$$P(M|p) = P(p|M) \frac{P(M)}{P(p)},$$

poiché $P(p|M) = 1 - P(n|M) = 1 - P_e$ e, per il teorema della probabilità completa, $P(p) = P(p|S)P(S) + P(p|M)P(M) = P_e(1 - P(M)) + (1 - P_e)P(M)$ otteniamo

$$P(M|p) = \frac{(1 - P_e)P(M)}{P_e(1 - P(M)) + (1 - P_e)P(M)},$$

la formula può essere semplificata nel caso che (come accade) sia $P(M)$ che P_e siano piccole rispetto ad 1:

$$P(M|p) \simeq \frac{1}{1 + [P_e/P(M)]}. \quad (12)$$

Con i valori numeri del problema si ha che la probabilità cercata è circa il 2%.

Dalla (12) risulta chiaro che per un test di laboratorio la cosa importante non è tanto la probabilità di errore del test P_e quanto il rapporto $P_e/P(M)$: tanto più una malattia è rara tanto più il test deve essere accurato, altrimenti il risultato non è significativo⁶.

⁶Il lettore può verificare che anche se $P(n|M) \neq P_e$ ma $P(n|M) \ll 1$ la (12) non cambia.

Non è difficile arrivare alla risposta giusta anche senza scomodare il formalismo. Su 100000 soggetti circa 100 sono malati e 99900 sani. Poiché il test sbaglia nel 5% dei casi si avranno circa 4995 soggetti sani che risultano positivi e circa 95 malati che risultano positivi. Quindi la probabilità di essere malati risultando positivi è circa $95/(95 + 4995) \simeq 2\%$.

C) Il problema di Monty Hall nasce dal gioco a premi della TV americana *Let's Make a Deal*. Al giocatore vengono mostrate tre porte chiuse; al di là di una c'è un'automobile e dietro le altre due una capra.

Dopo che il giocatore ha scelto una porta, ma non l'ha aperta, il conduttore dello show (che conosce cosa c'è dietro ogni porta) apre un'altra porta, rivelando una delle due capre, e offre al giocatore la possibilità di cambiare la propria scelta iniziale, passando all'unica porta restante. Passare all'altra porta migliora le chance del giocatore di vincere l'automobile? La risposta è sì: le probabilità di vittoria passano da $1/3$ a $2/3$.

Un'analisi del problema attraverso il teorema di Bayes mostra che è conveniente cambiare porta. Si consideri, senza perdita di generalità, il caso in cui la porta scelta è la 3, e non è stata ancora aperta alcuna porta.

La probabilità che l'automobile si trovi dietro la porta 2, che indichiamo con $P(A_2)$, è ovviamente $1/3$, in quanto l'auto ha la stessa probabilità di trovarsi dietro ciascuna porta. La probabilità che il conduttore dello show apra la porta 1, $P(C_1)$, è $1/2$, infatti l'auto ha la stessa probabilità di trovarsi dietro la porta 1 (il che costringerebbe il conduttore ad aprire la porta 2) come dietro la porta 2 (il che costringerebbe il conduttore ad aprire la porta 1); se poi l'auto non si trova dietro nessuna delle due porte (1 oppure 2), si può ipotizzare che il conduttore ne apra una a caso, con uguale probabilità. Notare che se l'auto si trova dietro la porta 2, in base a queste ipotesi il conduttore aprirà sicuramente la porta 1 cioè $P(C_1|A_2) = 1$. Utilizzando la formula di Bayes si ha:

$$P(A_2|C_1) = \frac{P(C_1|A_2)P(A_2)}{P(C_1)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

quindi conviene cambiare porta.

D) *Problema dei tre prigionieri*. Tre uomini A, B e C sono in prigione. Il prigioniero A ha saputo che due saranno giustiziati ed uno liberato, ma non sa chi. Il giudice ha deciso "a caso" il graziato quindi la probabilità

che A sia liberato è $1/3$. Al secondino, che conosce il nome del graziato, A dice *poiché due di noi saranno giustiziati, certamente almeno uno sarà B o C . Tu non mi dirai nulla sulla mia sorte, ma mi dirai chi tra B e C sarà giustiziato*. Il secondino accetta e dice che B sarà giustiziato. Il prigioniero A si sente un po' più sollevato pensando che sarà giustiziato C oppure lui e conclude che la sua probabilità di essere liberato è salita da $1/3$ ad $1/2$. Ha ragione ad essere ottimista?

Indichiamo con $P(A)$ la probabilità che A sarà liberato e $P(b)$ la probabilità che il secondino dica che B sarà giustiziato. Dalla formula di Bayes la probabilità $P(A|b)$ che A sarà liberato, sapendo che B sarà giustiziato è data da

$$P(A|b) = \frac{P(A \cap b)}{P(b)} = P(b|A) \frac{P(A)}{P(b)},$$

ove $P(b|A)$ è la probabilità che B sarà giustiziato sapendo che A sarà liberato e, per il teorema della probabilità completa $P(b) = P(b|A)P(A) + P(b|B)P(B) + P(b|C)P(C)$. Ovviamente $P(A) = P(B) = P(C) = 1/3$ mentre $P(b|A) = 1/2$ (infatti se A sarà liberato il secondino dirà B o C con uguale probabilità), ed inoltre (non hanno bisogno di commento) $P(b|B) = 0$ e $P(b|C) = 1$, si ottiene

$$P(b) = \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{1}{2}$$

e quindi

$$P(A|b) = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

4 Variabili casuali

4.1 Una variabile casuale

Alcuni esempi di eventi elementari, $\omega \in \Omega$, su cui si definiscono le probabilità sono: le due facce di una moneta, le sei facce di un dado (con sopra sei simboli diversi arbitrariamente scelti), il tempo domani, il colore di una palla estratta da un'urna, ecc.; cioè avvenimenti non necessariamente riconoscibili mediante un valore numerico. Supponiamo che Ω sia un insieme discreto: se si assegna un numero reale a ogni evento elementare, $X(\omega) \in \mathfrak{R}$, si definisce una variabile aleatoria. I valori della variabile (che indicheremo con il carattere

minuscolo, x) hanno probabilità che si desumono dalle probabilità degli eventi ω su cui X è definita: se x è un particolare valore di X , esso eredita la probabilità complessiva degli eventi ω_k ai quali il valore è associato. Cioè: $P_X(x) = \sum_k P(\omega_k)$, essendo $X(\omega_k) = x$. Se la $X(\omega)$ è biunivoca: $P_X(x) = P(\omega)$ e la variabile semplicemente rinomina gli eventi elementari (per es., questo si fa quando si distinguono le facce di un dado con i numeri da 1 a 6).

Quanto esposto sopra vale sia per eventi elementari discreti che continui, nel qual caso la $X(\omega)$ deve rispettare opportune condizioni, abbastanza generali, e la $P_X(x)$ va pensata come densità di probabilità.

È utile sottolineare quanto segue.

La conoscenza di $P_X(x)$ è una descrizione probabilistica dettagliata della variabile aleatoria X . Ma una volta che si sia introdotta una variabile numerica è possibile definire una vasta serie di indicatori o funzioni numeriche (valori medi di vario tipo) che forniscono una descrizione parziale o alternativa della distribuzione completa $P_X(x)$.

Se quel che interessa non sono tanto gli eventi elementari quanto i valori della variabile X a essi associata, è chiaro che, una volta ricavata la distribuzione $P_X(x)$ ci si può dimenticare degli eventi ω di partenza. Si tenga conto però del fatto che, se la funzione $X(\omega)$ non è biunivoca (come succede in genere), così facendo, si perde una parte dell'informazione iniziale: attraverso la $X(\omega)$ si realizza una partizione dello spazio iniziale Ω mettendo insieme (non distinguendo più) gli eventi associati allo stesso valore di X : $P(\omega)$ è più dettagliata di $P_X(x)$.

Per esempio, in meccanica statistica la descrizione di un sistema in equilibrio è data da una densità di probabilità definita nello spazio delle fasi del sistema: gli eventi elementari sono quindi gli stati dinamici del sistema, individuati da opportune variabili canoniche, indicate collettivamente con \mathbf{Q} , che danno nome agli eventi (se si hanno N particelle in uno spazio tridimensionale, \mathbf{Q} è un vettore costituito da posizioni e impulsi di tutte le particelle, quindi $\mathbf{Q} \in \mathfrak{R}^{6N}$).

Sugli stati del sistema è anche definita una funzione energia $H(\mathbf{Q})$ che, in questo contesto, è una variabile casuale. All'equilibrio la densità di probabilità $p_{\mathbf{Q}}(\mathbf{q})$ dipende dallo stato \mathbf{Q} solo attraverso la sua energia, $p_{\mathbf{Q}}(\mathbf{q}) = p_{\mathbf{Q}}(H(\mathbf{q}))$ e, in generale, ciò che interessa (ciò che si può studiare) è l'energia del sistema. È quindi utile ricavare la distribuzione di probabilità dell'energia del sistema. Indicando con E i possibili valori di $H(\mathbf{Q})$, in base alla

regola data sopra, si avrà

$$p_H(E) = G(E) \cdot p_{\mathbf{Q}}(H(\mathbf{q}))|_{H(\mathbf{q})=E} ,$$

dove la densità degli stati $G(E)$ tiene conto di tutti gli stati del sistema che hanno energia E (che, in base all'ipotesi fatta, hanno tutti la stessa probabilità). Se solo la variabile E interessa ci si può dimenticare della $p_{\mathbf{Q}}(H(\mathbf{q}))$ iniziale e ragionare in termini della $p_H(E)$.

4.2 Più variabili casuali

Consideriamo il caso in cui su un dato insieme di eventi elementari si possono definire più variabili casuali; per es., oltre a $X(\omega)$ anche $Y(\omega)$. In questo caso oltre a $P_X(x)$ e $P_Y(y)$ si potrà definire anche la probabilità congiunta $P_{X,Y}(x, y)$, in modo evidente: $P(x, y) = \sum_k P(\omega_k)$, essendo contemporaneamente $X(\omega_k) = x$ e $Y(\omega_k) = y$ (nel caso continuo la somma eventualmente sostituita da un integrale). In generale l'informazione contenuta in $P_{X,Y}(x, y)$ sarà minore di quella iniziale ma più dettagliata di quella contenuta in $P_X(x)$ o $P_Y(y)$; cioè la partizione di Ω realizzata mettendo insieme gli eventi elementari con ugual valore della coppia (x, y) è più raffinata di quella realizzata usando solamente x o y . A questo punto è anche chiaro che si può ottenere $P_X(x)$ (o $P_Y(y)$) sommando la probabilità congiunta su tutti i valori di y (o di x) mentre si tiene fissato x (o y):

$$P_X(x) = \sum_y P_{X,Y}(x, y) . \quad (13)$$

A ciò è legato l'appellativo di distribuzioni marginali per $P_X(x)$ e $P_Y(y)$.

Una volta definita la distribuzione congiunta di due variabili si può introdurre il concetto di variabili indipendenti, in modo del tutto analogo al concetto di eventi indipendenti: X e Y sono (mutuamente) indipendenti se $P_{X,Y}(x, y) = P_X(x) \cdot P_Y(y)$. E il concetto di probabilità condizionata: la probabilità di X condizionata a un certo valore di Y è

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x, y)}{P_Y(y)} , \quad (14)$$

e questa definizione vale anche nel caso in cui le variabili siano continue, e quindi le P siano densità di probabilità.

4.3 Valori medi

Considerando direttamente il caso di variabili continue, supponiamo che sia data una variabile aleatoria X con densità di probabilità $p_X(x)$. Il valor medio di una funzione $f(X)$ è definito come

$$\langle f(X) \rangle = E(f(X)) = \int f(x)p_X(x)dx .$$

Nel caso di più variabili:

$$\begin{aligned} \langle f(X_1, \dots, X_N) \rangle &= E(f(X_1, \dots, X_N)) \\ &= \int f(x_1, \dots, x_N)p_{X_1, \dots, X_N}(x_1, \dots, x_N)dx_1 \dots dx_N . \end{aligned} \quad (15)$$

Particolarmente importanti, da un punto di vista pratico, sono i casi $f(X) = X$, $f(X) = X^2$ e la quantità associata σ_X^2 , la varianza, definita come

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2 = \langle (X - \langle X \rangle)^2 \rangle .$$

La media $E(X)$ dà un' indicazione sulla localizzazione della densità e σ_X^2 , dà un' idea della dispersione dei valori della variabile intorno alla media. Da un punto di vista teorico è importante la media di $f(X) = \exp(itX)$:

$$\phi_X(t) = \int \exp(itx)p_X(x)dx \quad (16)$$

è la funzione caratteristica della distribuzione $p_X(x)$ e ne fornisce, sotto ipotesi abbastanza generali, una descrizione equivalente (essendo la sua trasformata di Fourier).

Date N variabili aleatorie X_1, \dots, X_N e N costanti reali a_1, \dots, a_N , indicando con Y la somma $a_1X_1 + \dots + a_NX_N$, le seguenti proprietà sono di facile dimostrazione (lasciata al lettore),

$$E(Y) = E(a_1X_1 + \dots + a_NX_N) = a_1E(X_1) + \dots + a_NE(X_N) , \quad (17)$$

$$\sigma_Y^2 = \sum_{j=1}^N a_j^2 \sigma_{X_j}^2 + 2 \sum_{j < n} a_j a_n E\left((X_j - E(X_j))(X_n - E(X_n))\right) . \quad (18)$$

Nel caso particolare che X_1, \dots, X_N siano variabili mutuamente indipendenti si ha anche:

$$E\left(\prod_{j=1}^N f_j(X_j)\right) = \prod_{j=1}^N E(f_j(X_j)) , \quad \sigma_Y^2 = \sum_{j=1}^N a_j^2 \sigma_{X_j}^2 . \quad (19)$$

4.4 Valori medi condizionati

Considerando, per semplicità, il caso di due variabili X ed Y con densità di probabilità congiunta $p_{X,Y}(x,y)$, definiamo il valore di aspettazione di una funzione $f(X)$ condizionato a un dato valore di Y come

$$E(f(X)|y) = \int f(x)p_{X|Y}(x|y)dx \quad ,$$

dove $p_{X|Y}(x|y)$ indica la densità di X condizionata a un dato valore di Y . Dall' eq. (14) è chiaro che il valor medio non condizionato $E(f(X))$ si scrive come segue

$$E(f(X)) = \int f(x)p_{X,Y}(x,y)dxdy = \int E(f(X)|y)p_Y(y)dy \quad .$$

4.5 Distribuzione di una somma di variabili

Date le variabili X e Y e la variabile somma $Z = X + Y$, abbiamo visto che per i valor medi si ha $E(Z) = E(X) + E(Y)$. Un' informazione più accurata su Z è data dalla sua densità che, partendo dalla densità congiunta di X e Y : $p_{X,Y}(x,y)$, si scrive

$$p_Z(z) = \int p_{X,Y}(x,y) \delta(z - x - y) dxdy$$

ovvero (per es., integrando su y)

$$p_Z(z) = \int p_{X,Y}(x, z - x) dx \quad . \quad (20)$$

Può essere istruttivo ricavare la formula precedente partendo dal caso di variabili discrete. Se si indica con $p_{i,j}^{(x,y)}$ la probabilità dell' evento $(x = i\Delta x, y = j\Delta y)$, dove i e j sono interi, allora la probabilità di avere $z = k\Delta z$ (con $\Delta x = \Delta y = \Delta z$) è:

$$p_k^{(z)} = \sum_{i,j} p_{i,j}^{(x,y)} \delta_{k,i+j} = \sum_i p_{i,k-i}^{(x,y)} \quad . \quad (21)$$

Nel limite $\Delta x \rightarrow 0$ si ha

$$p_k^{(z)} = p_Z(k\Delta z)\Delta z \quad , \quad p_{i,j}^{(x,y)} = p_{X,Y}(i\Delta x, j\Delta y)\Delta x\Delta y \quad ,$$

e quindi la (20). Nel caso particolare e interessante che X e Y siano indipendenti la formula si riduce a

$$p_Z(z) = \int p_X(x)p_Y(z-x) dx, \quad (22)$$

e quindi la densità della somma è la convoluzione delle densità di partenza.

5 Alcune Distribuzioni Notevoli

5.1 Distribuzione Binomiale

Questa distribuzione, pur elementare, ha un ruolo importante nel calcolo della probabilità. Consideriamo la variabile $Y_N = X_1 + X_2 + \dots + X_N$, ove le $\{X_i\}$ sono variabili indipendenti e identicamente distribuite (*i.i.d.*) che valgono 1 oppure 0, con probabilità p e $1-p$ rispettivamente. La variabile Y_N può quindi assumere i valori interi in $[0, N]$. La probabilità che $y_N = k$ è:

$$P_N(k) = C_{N,k} p^k (1-p)^{N-k}, \quad (23)$$

detta *distribuzione binomiale*, dove $C_{N,k}$ è il numero di modi (combinazioni) in cui si possono disporre k oggetti in una sequenza lunga N . L'espressione, ben nota, per $C_{N,k}$ è

$$C_{N,k} = \frac{N!}{k!(N-k)!},$$

e quindi

$$P_N(k) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}. \quad (24)$$

Il calcolo della media e della varianza di y_N sono elementari: poichè le $\{X_i\}$ sono i.i.d. si ha (vedi eq.i (17) e (19)):

$$E(Y_N) = NE(X) = Np, \quad \sigma_{Y_N}^2 = N\sigma_X^2 = Np(1-p).$$

Dalla distribuzione binomiale nel limite $N \gg 1$ e p finita si ottiene la distribuzione Gaussiana (v. Sez. 8.2 e 8.4).

Un diverso caso limite interessante è quello seguente.

5.2 Distribuzione di Poisson

Consideriamo l'andamento limite della distribuzione Binomiale nelle condizioni $N \gg 1$, $p = \lambda/N$ con $\lambda = O(1)$ e $k \ll N$. Sotto queste condizioni si ha:

$$\frac{N!}{k!(N-k)!} = \frac{N(N-1)\dots(N-k+1)}{k!} \simeq \frac{N^k}{k!}$$

e

$$(1-p)^{N-k} = \left(1 - \frac{\lambda}{N}\right)^{N-k} \simeq \left(1 - \frac{\lambda}{N}\right)^N \simeq e^{-\lambda},$$

dalla (24) si ha

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$

cioè un'espressione indipendente da N . Notare che la probabilità è normalizzata correttamente: $P(0) + P(1) + \dots = 1$. Il significato di λ è chiaro dal calcolo di $E(k) = \langle k \rangle$:

$$E(k) = \sum_{k=0}^{\infty} k P(k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \lambda \frac{\partial}{\partial \lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda,$$

con un calcolo analogo si ha

$$\sigma^2 = E(k^2) - E(k)^2 = \lambda.$$

Un esempio elementare (ma interessante per la meccanica statistica) della distribuzione di Poisson è la probabilità di trovare k particelle in una piccola regione di volume ΔV di un recipiente di volume $V \gg \Delta V$, contenente un numero N molto grande di particelle. Assumendo che le particelle siano distribuite uniformemente allora la probabilità che una data particella sia contenuta in regione di volume ΔV è $p = \Delta V/V$. Trascurando l'interazione tra le particelle e assumendo quindi che esse possano trovarsi in ΔV indipendentemente una dall'altra, la probabilità che k particelle siano in ΔV è data dalla binomiale (24). Introducendo la densità $\rho = N/V$ si può scrivere $Np = \rho\Delta V = \langle k \rangle$; pensando fissate le quantità ρ e ΔV , e quindi $\langle k \rangle$, si potrà considerare la (24) nel limite $N \gg 1$, $p = \langle k \rangle/N \ll 1$, ottenendo una distribuzione di Poisson.

5.3 Distribuzione χ^2 di Pearson

Siano X_1, \dots, X_N variabili i.i.d. con densità di probabilità gaussiana a media nulla e varianza unitaria; abbiamo quindi

$$p_{X_1, \dots, X_N} = \sqrt{\frac{1}{(2\pi)^N}} \exp -\frac{1}{2} \sum_{j=1}^N x_j^2 .$$

Consideriamo la variabile

$$\chi_N^2 = \sum_{j=1}^N X_j^2$$

utilizzando una formula che sarà introdotta in seguito, eq. (35), per la variabile $Y = \chi_N^2$, e ricordando la definizione della funzione gamma di Eulero, si ha

$$p_Y(y) = \frac{y^{(N/2-1)}}{2^{N/2}\Gamma(N/2)} e^{-\frac{y}{2}} . \quad (25)$$

La precedente distribuzione di probabilità è chiamata distribuzione di χ^2 di Pearson per N gradi di libertà; in meccanica statistica descrive la distribuzione di probabilità di energia di un gas di particelle classiche non interagenti.

Analogamente per la variabile $Z = \chi_N$ si ha

$$p_Z(z) = 2 \frac{z^{N-1}}{2^{N/2}\Gamma(N/2)} e^{-\frac{z^2}{2}} .$$

La distribuzione di probabilità di χ^2 (o equivalentemente quella per χ) ha un ruolo importante nel trattamento dei dati sperimentale. È naturale infatti aspettarsi che la differenza tra un'osservazione sperimentale ed il valore "vero" sia una variabile gaussiana.

5.4 Distribuzione multidimensionale di variabili gaussiane

La distribuzione di probabilità di Gauss con media m e varianza σ^2 :

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x - m)^2 ,$$

può essere generalizzata al caso di N variabili indipendenti X_1, \dots, X_N , ciascuna con media m_j e varianza σ_j^2 :

$$p_{X_1, \dots, X_N}(x_1, \dots, x_N) = \prod_n \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp - \sum_n \frac{1}{2\sigma_n^2} (x_n - m_n)^2 .$$

Considerando delle nuove variabili Y_1, \dots, Y_N esprimibili come combinazioni lineari delle X_1, \dots, X_N :

$$\mathbf{Y} = \mathcal{A}\mathbf{X} + \mathcal{B} ,$$

con \mathcal{A} a determinante non nullo, si può vedere che se $A_{ij} = \mathcal{A}_{ij}$ è una matrice simmetrica definita positiva (cioè con autovalori positivi) la distribuzione di probabilità delle \mathbf{Y} è della forma:

$$p_{\mathbf{Y}}(y_1, \dots, y_N) = \sqrt{\frac{|\det \mathcal{A}|}{(2\pi)^N}} \exp - \frac{1}{2} \sum_{i,j} (y_i - b_i)(y_j - b_j) A_{ij} \quad (26)$$

dove $\{b_j\}$ sono i valori medi di $\{Y_j\}$. È facile mostrare⁷ che:

$$\langle (y_i - b_i)(y_j - b_j) \rangle = [\mathcal{A}^{-1}]_{ij} .$$

La (26) è detta gaussiana multivariata.

Nel caso $N = 2$ con $\langle x_1 \rangle = \langle x_2 \rangle = 0$ e $\sigma_1 = \sigma_2 = 1$ la forma più generale di gaussiana bivariata è:

$$p_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp - \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)} ,$$

dove ρ è il coefficiente di correlazione tra X_1 e X_2 : $\langle x_1 x_2 \rangle = \rho$ e $|\rho| \leq 1$. Nel caso generale di due variabili Y_1 e Y_2 , con valori medi m_1 e m_2 , e varianze σ_1^2 e σ_2^2 , la distribuzione gaussiana bivariata si ottiene facilmente dalla precedente con il semplice cambio di variabili:

$$Y_1 = m_1 + \sigma_1 X_1 \quad Y_2 = m_2 + \sigma_2 X_2$$

⁷Basta cambiare variabile:

$$\mathbf{y} \rightarrow \mathbf{z} = \mathcal{C}(\mathbf{y} - \mathbf{b})$$

in modo tale che z_1, \dots, z_N siano indipendenti, calcolare $\langle z_j^2 \rangle$ e poi tornare a $\langle (y_i - b_i)(y_j - b_j) \rangle$.

ottenendo così:

$$p_{Y_1, Y_2}(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp - \frac{1}{2(1-\rho^2)} \left[\left(\frac{y_1 - m_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{y_1 - m_1}{\sigma_1} \right) \left(\frac{y_2 - m_2}{\sigma_2} \right) + \left(\frac{y_2 - m_2}{\sigma_2} \right)^2 \right], \quad (27)$$

e

$$\rho = \frac{1}{\sigma_1\sigma_2} \langle (y_1 - m_1)(y_2 - m_2) \rangle .$$

6 Funzioni generatrici: come contare senza sbagliare

Molti problemi di probabilità con variabili intere sono riconducibili a calcoli combinatori. Consideriamo ad esempio il lancio di 3 dadi truccati in modo tale che per il primo dado il risultato 1, 2, ..., 6 avvenga con probabilità p_1, p_2, \dots, p_6 , per il secondo dado con probabilità q_1, q_2, \dots, q_6 , per il terzo dado con probabilità t_1, t_2, \dots, t_6 (ovviamente con i vincoli $\sum_i p_i = \sum_i q_i = \sum_i t_i = 1$) e ci si chiede la probabilità che la somma faccia 12 oppure 8. Un calcolo esplicito basato solo sulle definizioni elementari è chiaramente possibile, le cose si complicano se invece di 3 dati se ne considerano 5 o 6.

Per fortuna esiste una tecnica semplice e potente, basata sulla funzione generatrice, che permette, diciamo così, di *contare senza l'esplicita enumerazione di tutti i casi possibili*.

Data una sequenza di numeri reali $P_0, P_1, \dots, P_k, \dots$, la funzione generatrice $G(s)$ è definita come segue:

$$G(s) = \sum_{k=0}^{\infty} s^k P_k = P_0 + sP_1 + s^2P_2 + \dots . \quad (28)$$

Se la serie è convergente, $G(s)$ permette di archiviare in modo compatto la sequenza, a causa delle seguenti (evidenti) proprietà:

$$G(0) = P_0, \quad G'(0) = P_1, \quad \dots, \quad \left. \frac{1}{n!} \frac{d^n G(s)}{ds^n} \right|_{s=0} = P_n . \quad (29)$$

Se la sequenza dà le probabilità di una variabile aleatoria X a valori interi, cioè $P_k = P_X(x = k)$, alle precedenti proprietà si aggiunge $G(1) = 1$ e, usando per la funzione generatrice la notazione $G_X(s)$, si ha anche $G_X(s) = E(s^X)$.

Se X_1, \dots, X_N sono variabili indipendenti con funzioni generatrici $G_{X_i}(s)$, allora per la funzione generatrice $G_Z(s)$ della variabile somma $Z = X_1 + X_2 + \dots + X_N$ si ha:

$$G_Z(s) = E(s^Z) = E\left(\prod_{i=1}^N s^{X_i}\right) = \prod_{i=1}^N E(s^{X_i}) = \prod_{i=1}^N G_{X_i}(s); \quad (30)$$

notare che questo risultato vale solo per variabili indipendenti, perché solo in tal caso è vera la terza uguaglianza, v. eq. (19). La formula (30) permette di risolvere senza intralci il problema dei 3 dadi truccati:

$$G_Z(s) = (sp_1 + s^2p_2 + \dots + s^6p_6)(sq_1 + s^2q_2 + \dots + s^6q_6)(st_1 + s^2t_2 + \dots + s^6t_6)$$

la probabilità che il risultato sia k (con $k = 3, 4, \dots, 18$) è semplicemente il coefficiente davanti a s^k nella $G_Z(s)$, calcolo che non presenta nessuna difficoltà.

Se le funzioni generatrici fossero utili solo per problemi di dadi truccati, o cose simili, la cosa non sarebbe poi tanto interessante. L'idea di fondo della funzione generatrice, comune ad altre situazioni della matematica, è una sorta di "cambiamento di base" (molto simile all'uso delle trasformate di Fourier). La conoscenza della $G(s)$ è del tutto equivalente alla conoscenza delle $\{P_k\}$, e a volte è più facile lavorare con le funzioni generatrici e poi tornare alle $\{P_k\}$.

Un risultato facile da ottenersi, ma interessante, è il seguente: se X_1, \dots, X_N sono variabili Poissoniane indipendenti con parametri $\lambda_1, \dots, \lambda_N$, cioè

$$P_{X_i}(x_i = k) = \frac{\lambda_i^k}{k!} e^{-\lambda_i} \quad k = 0, 1, \dots$$

allora la variabile $Z = X_1 + X_2 + \dots + X_N$ è Poissoniana con parametro $\Lambda = \sum_{i=1}^N \lambda_i$. Basta calcolare la $G_{X_i}(s)$:

$$G_{X_i}(s) = \sum_k s^k P_k = \sum_k s^k \frac{\lambda_i^k}{k!} e^{-\lambda_i} = e^{-\lambda_i(1-s)}$$

e usare la (30), ottenendo

$$G_Z(s) = e^{-(\sum_i \lambda_i)(1-s)},$$

cioè la funzione generatrice della Poissoniana con parametro $\Lambda = \sum_{i=1}^N \lambda_i$.

7 Qualche risultato utile

Discutiamo alcuni risultati semplici e utili, particolarmente in meccanica statica.

7.1 Come cambiare variabile

Consideriamo il caso in cui conosciamo la densità di probabilità $p_X(x)$ della variabile X , ci domandiamo la densità di probabilità, $p_Y(y)$ della variabile $Y = f(X)$. Per semplicità consideriamo il caso in cui $f(x)$ sia invertibile, cioè $f' \neq 0$. Ricordando che, dalla definizione di variabile casuale, i valori di Y ereditano le probabilità dei valori di X da cui provengono (come è anche evidente dalla Fig. 5), se $y_1 = f(x_1)$ e $y_2 = f(x_2)$ si ha:

$$P(y \in [y_1, y_2]) = P(x \in [x_1, x_2]) .$$

Se $x_2 = x_1 + \Delta x$ con Δx piccolo, allora ponendo $y_2 = y_1 + \Delta y$ con $\Delta y = f'(x_1)\Delta x$, poichè

$$p_X(x)\Delta x = p_Y(y)|f'(x)|\Delta x ,$$

(il modulo è stato introdotto per tener conto dei casi con $f' < 0$) si ottiene

$$p_Y(y) = \frac{p_X(x^*)}{|f'(x^*)|} , \quad x^* = f^{-1}(y) . \quad (31)$$

Nel caso di f non monotona (lasciato come facile esercizio) si ha:

$$p_Y(y) = \sum_{x^{(k)}:f(x^{(k)})=y} \frac{p_X(x^{(k)})}{|f'(x^{(k)})|} . \quad (32)$$

Tenendo presenti le proprietà delle delta di Dirac, la (31) e la (32) possono essere scritte in forma compatta (e facile da ricordare):

$$p_Y(y) = \int p_X(x)\delta(y - f(x))dx .$$

Nel caso di più variabili, cioè $Y_j = f_j(X_1, \dots, X_N)$, con $j = 1, \dots, N$ si procede in modo analogo

$$p_{Y_1, \dots, Y_N}(\mathbf{y}) = \sum_{\mathbf{x}^{(k)}:\mathbf{f}(\mathbf{x}^{(k)})=\mathbf{y}} \frac{p_{X_1, \dots, X_N}(\mathbf{x}^{(k)})}{|\det \mathcal{J}(\mathbf{x}^{(k)})|}$$

ove \mathcal{J} è la matrice con elementi $\partial f_i / \partial x_j$

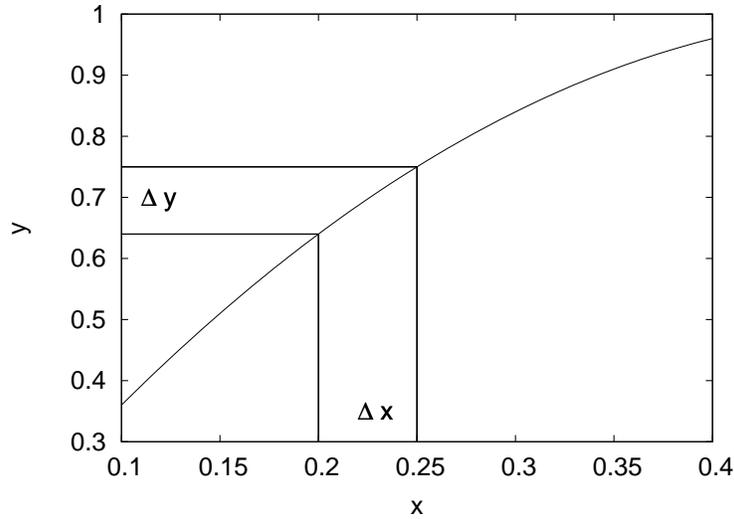


Figure 5: Relazione tra densità di x e di $y(x)$.

7.2 Come eliminare variabili che non interessano

Estendiamo qui un' idea già esposta nella Sez. **3.2**, introducendo le distribuzioni marginali. Supponiamo di conoscere la densità di probabilità congiunta $p_{X_1, \dots, X_N}(x_1, \dots, x_N)$ e di non essere interessati a tutte le variabili X_1, \dots, X_N ma solo ad alcune di esse, oppure ad una funzione $Y = f(X_1, \dots, X_N)$. Vedremo che queste sono situazioni piuttosto comuni in meccanica statistica. Come procedere?

Cominciamo, per semplicità di notazione, con il caso di 2 variabili: data $p_{X_1, X_2}(x_1, x_2)$ come determinare $p_{X_1}(x_1)$? La risposta è evidente:

$$p_{X_1}(x_1) = \int p_{X_1, X_2}(x_1, x_2) dx_2 . \quad (33)$$

Nel caso di 3 variabili si può essere interessati ad una sola (ad esempio X_1) oppure due (ad esempio (X_1, X_2)) quindi

$$\begin{aligned} p_{X_1}(x_1) &= \int p_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_2 dx_3 , \\ p_{X_1, X_2}(x_1, x_2) &= \int p_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_3 . \end{aligned} \quad (34)$$

La generalizzazione è ovvia.

Discutiamo ora la densità di probabilità di una funzione delle variabili aleatorie: $Y = f(X_1, \dots, X_n)$. Anche in questo caso la risposta è evidente:

$$P(y \in [y_1, y_2]) = \int_{y_1 < f(x_1, \dots, x_n) < y_2} p_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_N$$

nel limite di y_1 molto vicino a y_2 si ha

$$p_Y(y)dy = \int_{y < f(x_1, \dots, x_n) < y+dy} p_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_N .$$

La formula precedente può essere scritta nella forma facile da ricordare

$$p_Y(y) = \int p_{X_1, \dots, X_n}(x_1, \dots, x_n) \delta[y - f(x_1, \dots, x_n)] dx_1 \dots dx_N . \quad (35)$$

8 Conseguenze in Meccanica Statistica

In meccanica statistica in diversi casi interessanti ci si disinteressa di una classe di variabili e si utilizza un procedimento di proiezione:

- a) nel passaggio dall' ensemble microcanonico a quello canonico;
- b) nel calcolo della distribuzione di probabilità dell'energia, o di altre quantità macroscopiche;
- c) in teoria cinetica quando si introducono le distribuzioni a una particella, a due, etc.; e in teoria dei liquidi in cui la distribuzione a due particelle gioca un ruolo fondamentale.

8.1 Dall' insieme microcanonico a quello canonico

Indichiamo con $(\mathbf{Q}_1, \mathbf{Q}_2)$ le variabili che descrivono lo stato microscopico di un sistema costituito da N particelle in un volume V , con energia compresa nell' intervallo $[E, E + \Delta]$, per il quale vale da distribuzione microcanonica $\rho_M(\mathbf{Q}_1, \mathbf{Q}_2)$. Le variabili \mathbf{Q}_1 determinano lo stato di N_1 particelle nel volume V_1 (sistema S_1), analogamente le \mathbf{Q}_2 determinano lo stato delle rimanenti $N_2 = N - N_1$ particelle in un volume $V_2 = V - V_1$ (sistema S_2).

La densità di probabilità per il solo sistema S_1 segue dalla (33):

$$\rho_{S_1}(\mathbf{Q}_1) = \int \rho_M(\mathbf{Q}_1, \mathbf{Q}_2) d\mathbf{Q}_2 .$$

Ponendo

$$\omega(E, N, V)\Delta = \int_{E < H < E + \Delta} d\mathbf{Q}_1 d\mathbf{Q}_2 ,$$

per definizione di insieme microcanonico si ha

$$\begin{aligned} \rho_M(\mathbf{Q}_1, \mathbf{Q}_2) &= \frac{1}{\omega(E, N, V)\Delta} \quad \text{se } H((\mathbf{Q}_1, \mathbf{Q}_2)) \in [E, E + \Delta] \\ \rho_M(\mathbf{Q}_1, \mathbf{Q}_2) &= 0 \quad \text{se } H((\mathbf{Q}_1, \mathbf{Q}_2)) \notin [E, E + \Delta] , \end{aligned} \quad (36)$$

e se nell' Hamiltoniana totale $H((\mathbf{Q}_1, \mathbf{Q}_2)) = H_1(\mathbf{Q}_1) + H_2(\mathbf{Q}_2) + H_{12}(\mathbf{Q}_1, \mathbf{Q}_2)$ si trascura il contributo di interazione⁸ H_{12} si ottiene

$$\rho_{S_1}(\mathbf{Q}_1) = \frac{\omega(E - H_1(\mathbf{Q}_1), N - N_1, V - V_1)}{\omega(E, N, V)} . \quad (37)$$

Scrivendo $\omega(E, N, V) = e^{S(E, N, V)/k_B}$, dove $S(E, N, V)$ è interpretabile come l'entropia del sistema e k_B è la costante di Boltzmann, ricordando che $\partial S/\partial E = T^{-1}$ è l'inverso della temperatura, nel limite $H_1 \ll E$, $N_1 \ll N$ e $V_1 \ll V$, con uno sviluppo di Taylor si ottiene la distribuzione canonica per il sottosistema S_1 :

$$\rho_C(\mathbf{Q}_1) = \frac{e^{-\beta H_1(\mathbf{Q}_1)}}{Z(\beta, V_1, N_1)} , \quad (38)$$

dove $\beta = 1/k_B T$ e $Z(\beta, V_1, N_1) = \int e^{-\beta H_1(\mathbf{Q}_1)} d\mathbf{Q}_1$ è la funzione di partizione.

8.2 Densità di probabilità marginali meccanica statistica

Un esempio di applicazione della formula (35) molto importante in fisica è il calcolo della densità di probabilità del modulo della velocità in meccanica

⁸Questo è fisicamente sensato se $N_1 \gg 1$ e inoltre il raggio di interazione tra le coppie di particelle è piccolo rispetto alla grandezza lineare del sistema descritto dalle variabili \mathbf{Q}_1 .

statistica classica. La distribuzione di Maxwell- Boltzmann fornisce la densità di probabilità delle componenti cartesiane della velocità:

$$p_{V_x, V_y, V_z}(v_x, v_y, v_z) = P_{MB}(\mathbf{v}) = B e^{-A(v_x^2 + v_y^2 + v_z^2)} , \quad (39)$$

ove $B = [m/(2\pi k_B T)]^{3/2}$ e $A = m/(2k_B T)$. Per la variabile modulo $V = \sqrt{V_x^2 + V_y^2 + V_z^2}$ dalla (35) si ha

$$p_V(v) = 4\pi B v^2 e^{-Av^2} .$$

Riprendiamo l' esempio introdotto alla fine della Sez. **3.1** usando come densità di equilibrio quella canonica (38). Essendo interessati alla sola variabile energia E si ha:

$$p_H(E)dE = \frac{1}{Z} \int_{E < H < E+dE} e^{-\beta H(\mathbf{Q})} d\mathbf{Q} ,$$

dove \mathbf{Q} indica le variabili che descrivono il sistema con Hamiltoniana $H(\mathbf{Q})$. Possiamo riscrivere l' equazione precedente nella forma

$$p_H(E)dE = \frac{1}{Z} e^{-\beta E} \int_{E < H < E+dE} d\mathbf{Q} = \frac{1}{Z} e^{-\beta E} \omega(E)dE ,$$

da cui, ricordando che $\omega(E) = e^{S(E)/k_B}$ si ottiene infine

$$p_H(E) = \frac{1}{Z} \exp - \left\{ \left(\frac{1}{k_B T} \right) [E - TS(E)] \right\} .$$

È interessante considerare il risultato precedente nel limite $N \gg 1$, in cui è naturale assumere (sotto opportune condizione si può dimostrare esplicitamente) che si abbia:

$$E = eN + o(N) , \quad S(E) = s(e)N + o(N)$$

dove $o(N)$ indica una quantità che cresce meno rapidamente di N , e ed $s(e)$ sono l'energia per particella e l'entropia per particella rispettivamente. Per la densità di probabilità della variabile $h = H/N$, abbiamo quindi

$$p_h(e) = N p_H(E(e)) = \frac{N}{Z} \exp - \left\{ N \left(\frac{1}{k_B T} \right) [e - Ts(e)] \right\} .$$

Questo tipo di densità di probabilità si può considerare una generalizzazione del teorema del limite centrale (v. Sez. **8.3, 8.4, 8.5**).

8.2.1 Dal microcanonico alla distribuzione di Maxwell-Boltzmann

È istruttivo ricavare la distribuzione di Maxwell-Boltzmann utilizzando la (33). Consideriamo un sistema costituito N particelle non interagenti con Hamiltoniana

$$H = \sum_{n=1}^N \frac{\mathbf{p}_n^2}{2m} ,$$

che supponiamo sia descritto dalla densità microcanonica. In modo analogo alla procedura usata per determinare la (38), per la densità di probabilità dell' impulso P di una particella si ha:

$$p_{\mathbf{P}}(\mathbf{p}) = \frac{\omega\left(E - \frac{\mathbf{p}^2}{2m}, N - 1\right)}{\omega(E, N)} .$$

Poichè

$$\omega(E, N) = 3NmC_N(2mE)^{(3N/2)-1} ,$$

dove $C_N = \pi^{3N/2}/\Gamma((3N/2) + 1)$, nel limite $N \gg 1$ si ha

$$p_{\mathbf{P}}(\mathbf{p}) \simeq \frac{C_{N-1}}{C_N} \frac{1}{(2mE)^{3/2}} \left(1 - \frac{\mathbf{p}^2}{2mE}\right)^{3N/2} ,$$

che è valida per $|\mathbf{p}| \leq \sqrt{2mE}$. Utilizzando l' approssimazione di Stirling $\Gamma(n + 1) \simeq n^n e^{-n} \sqrt{2\pi n}$ si ha $C_{N-1}/C_N \simeq (3N/2\pi)^{3/2}$, e ricordando che $E = \frac{3}{2}Nk_B T$ otteniamo

$$p_{\mathbf{P}}(\mathbf{p}) \simeq \frac{1}{\sqrt{(2\pi mk_B T)^3}} \exp - \frac{\mathbf{p}^2}{2mk_B T} , \quad (40)$$

che è la (39) scritta per l'impulso invece che per la velocità.

8.2.2 Densità di probabilità ridotte in teoria cinetica

Consideriamo un sistema di N particelle di massa m ed indichiamo con \mathbf{q}_j il vettore (a 6 dimensioni) che determina posizione e velocità della j -ma particella, cioè $\mathbf{q}_j = (\mathbf{x}_j, \mathbf{v}_j)$. L'informazione "completa" delle proprietà statistiche al tempo t è data dalla densità di probabilità $\rho_N(\mathbf{q}_1, \dots, \mathbf{q}_n, t)$ la cui evoluzione è determinata dall'equazione di Liouville

$$\frac{\partial \rho_N}{\partial t} + \sum_{j=1}^N \frac{\partial \rho_N}{\partial \mathbf{x}_j} \frac{\partial H}{\partial \mathbf{p}_j} - \sum_{j=1}^N \frac{\partial \rho_N}{\partial \mathbf{p}_j} \frac{\partial H}{\partial \mathbf{x}_j} = 0 ,$$

dove $\mathbf{p}_j = m\mathbf{v}_j$ e H è l'Hamiltoniana del sistema. In molte circostanze è sufficiente la conoscenza delle densità di probabilità ridotte:

$$\rho_1(\mathbf{q}_1, t) = \int \rho_N(\mathbf{q}_1, \dots, \mathbf{q}_N, t) d\mathbf{q}_2 d\mathbf{q}_3 \cdots d\mathbf{q}_N ,$$

$$\rho_2(\mathbf{q}_1, \mathbf{q}_2, t) = \int \rho_N(\mathbf{q}_1, \dots, \mathbf{q}_N, t) d\mathbf{q}_3 d\mathbf{q}_4 \cdots d\mathbf{q}_N ;$$

ad esempio nell'equazione di Boltzmann, che sotto opportune ipotesi descrive in modo accurato le proprietà statistiche di gas diluiti, compare solo la ρ_1 .

8.2.3 Densità di probabilità ridotte in fisica dei liquidi

Se le particelle interagiscono con un potenziale centrale a due corpi (cioè dipendente solo dalla distanza) allora la densità ridotta a due particelle è sufficiente per determinare le proprietà termodinamiche del sistema. In presenza di equilibrio termodinamico la ρ_2 non dipende dal tempo ed ha la forma

$$\rho_2(\mathbf{q}_1, \mathbf{q}_2) = P_{MB}(\mathbf{v}_1)P_{MB}(\mathbf{v}_2)F_2(\mathbf{x}_1, \mathbf{x}_2)$$

dove P_{MB} indica la densità di probabilità di Maxwell-Boltzmann per la velocità ed $F_2(\mathbf{x}_1, \mathbf{x}_2)$ è la parte spaziale. Per la simmetria radiale del problema, quest'ultima funzione dipende solo da $r = |\mathbf{x}_1 - \mathbf{x}_2|$ e sarà della forma $4\pi r^2 g_2(r)/V$ dove $g_2(r)$ è la funzione di distribuzione radiale definita come segue: $4\pi r^2 \rho_0 g_2(r) dr$ è la probabilità di trovare una particella a una distanza tra r ed $r + dr$ da una particella data, se $\rho_0 = N/V$ è la densità di particelle del sistema. La conoscenza di $g_2(r)$ è sufficiente a determinare l'energia media del sistema:

$$\frac{U}{N} = \frac{3}{2}k_B T + \frac{\rho_0}{2} \int_0^\infty 4\pi r^2 V_I(r) g_2(r) dr , \quad (41)$$

dove V_I è il potenziale di interazione.

In modo analogo, dall'equazione del viriale per la pressione P si può scrivere l'equazione di stato:

$$P = \rho_0 k_B T - \frac{2\pi\rho_0^2}{3} \int_0^\infty r^3 V_I'(r) g_2(r) dr . \quad (42)$$

Le equazioni (41) e (42) sono formalmente esatte, anche se non è semplice calcolare la $g_2(r)$. Tuttavia è interessante il fatto che $g_2(r)$ è misurabile con

esperimenti di scattering di neutroni (o di luce), inoltre nel limite di gas diluiti si possono ottenere approssimazioni analitiche per la $g_2(r)$.

9 Sistemi con tante variabili: Teoremi Limite

In questa sezione discuteremo gli andamenti limite della somma di un numero molto elevato di variabili indipendenti. I risultati che si ottengono sono di grande importanza da un punto di vista sia pratico (per le applicazioni in fisica, biologia e finanza) che concettuale; infatti mostrano in modo chiaro come il calcolo delle probabilità non è soltanto la scienza dell'incerto, ma è in grado di affermare che alcuni eventi sono praticamente certi, o praticamente impossibili.

9.1 La legge dei grandi numeri

Storicamente il primo esempio di teorema limite è stata la legge dei grandi numeri che, derivata per la prima volta da J. Bernoulli nel 1713, è alla base dell'interpretazione frequentistica della probabilità. Una semplice dimostrazione può essere ottenuta dalla disuguaglianza di Chebyshev ⁹:

$$P(|X - \langle X \rangle| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2} . \quad (43)$$

La dimostrazione della formula precedente è facile:

$$P(|X - \langle X \rangle| > \epsilon) = \int_{-\infty}^{\langle X \rangle - \epsilon} p_X(x) dx + \int_{\langle X \rangle + \epsilon}^{\infty} p_X(x) dx ,$$

notiamo che negli intervalli su cui si effettua l'integrale $|x - \langle X \rangle| > \epsilon$, quindi

$$P(|X - \langle X \rangle| > \epsilon) \leq \int_{-\infty}^{\langle X \rangle - \epsilon} \frac{(x - \langle X \rangle)^2}{\epsilon^2} p_X(x) dx + \int_{\langle X \rangle + \epsilon}^{\infty} \frac{(x - \langle X \rangle)^2}{\epsilon^2} p_X(x) dx$$

⁹Nel calcolo delle probabilità si incontrano frequentemente nomi russi, poiché non esiste una regola universalmente accettata per la trascrizione dall'alfabeto cirillico a quello latino è facile trovare lo stesso nome scritto in modi diversi, ad esempio Chebyshev a volte è scritto pure Tchebichev, analogamente Markov e Markoff sono la stessa persona, stessa cosa per Kolmogorov e Kolmogoroff, così come per Lyapunov, Ljapunov oppure Liapounoff. In queste note abbiamo seguito la trascrizione usata nella letteratura anglosassone, che comunque non è senza eccezioni.

$$\leq \int_{-\infty}^{\infty} \frac{(x - \langle X \rangle)^2}{\epsilon^2} p_X(x) dx = \frac{\sigma^2}{\epsilon^2} .$$

In modo analogo si ottiene la disuguaglianza di Markov: per ogni $k > 0$ si ha

$$P(|X - \langle X \rangle| > \epsilon) \leq \frac{E(|X - \langle X \rangle|^k)}{\epsilon^k} .$$

Consideriamo ora N variabili X_1, \dots, X_N indipendenti identicamente distribuite (i.i.d.), con valor medio $\langle X \rangle$ e varianza $\sigma_X^2 < \infty$. La variabile $Y_N = (X_1 + X_2 + \dots + X_N)/N$ ha valore medio $\langle X \rangle$ e varianza $\sigma_{Y_N}^2 = \sigma_X^2/N$. Usiamo ora la disuguaglianza (43) per la Y_N :

$$P\left(\left|\frac{1}{N} \sum_{n=1}^N X_n - \langle X \rangle\right| > \epsilon\right) \leq \frac{\sigma_{Y_N}^2}{\epsilon^2} = \frac{\sigma_X^2}{N\epsilon^2} . \quad (44)$$

dalla quale si ottiene che per ogni $\epsilon > 0$

$$\lim_{N \rightarrow \infty} P\left(\left|\frac{1}{N} \sum_{n=1}^N X_n - \langle X \rangle\right| > \epsilon\right) = 0 . \quad (45)$$

Il risultato precedente vale anche per variabili non identicamente distribuite, purché siano indipendenti e con varianza limitata: $\sigma_j^2 < B < \infty$. Indicando con m_j il valor medio della variabile X_j si ottiene facilmente:

$$P\left(\left|\frac{1}{N} \sum_{n=1}^N X_n - m_n\right| > \epsilon\right) \leq \frac{1}{N^2\epsilon^2} \sum_{n=1}^N \sigma_n^2 \leq \frac{B}{N\epsilon^2}$$

9.2 Teorema del limite centrale

Variabili indipendenti e identicamente distribuite.– Abbiamo visto che nel limite $N \rightarrow \infty$ la densità di probabilità di $Y_n = (X_1 + X_2 + \dots + X_N)/N$ diventa una delta di Dirac centrata intorno a $\langle X \rangle$. La domanda che segue in modo naturale è chiedersi la forma della densità di probabilità della variabile $X_1 + X_2 + \dots + X_N$ nel limite $N \gg 1$ intorno a $N\langle X \rangle$, vedremo che si ha una forma universale (cioè indipendente da $p(x)$).

Come abbiamo già visto nella Sez. **3.5**, la densità di probabilità della somma $Z = X + Y$, date $p_X(x)$ e $p_Y(y)$, si scrive

$$p_Z(z) = \int p_X(x)p_Y(y)\delta(z-(x+y))dxdy = \int p_X(x)p_Y(z-x)dx = (p_X * p_Y)(z) ,$$

dove \star indica la convoluzione. In generale date N variabili indipendenti X_1, \dots, X_N con densità di probabilità $p_1(x_1), \dots, p_N(x_N)$ per la variabile somma $Z = X_1 + \dots + X_N$ si ha

$$p_Z(z) = (p_1 \star p_2 \star \dots \star p_N)(z) . \quad (46)$$

A parte qualche eccezione¹⁰ la formula precedente non è di uso facile.

Abbiamo quindi bisogno di un metodo per affrontare il problema delle somme di variabili aleatorie indipendenti consentendoci il controllo del comportamento della (46) nel limite $N \gg 1$. Lo strumento tecnico chiave è la funzione caratteristica:

$$\phi_X(t) = \int e^{itx} p_X(x) dx = E(e^{itX}) . \quad (47)$$

Notare che, a parte una costante moltiplicativa, la $\phi_X(t)$ è la trasformata di Fourier della $p_X(x)$; possiamo dire, sotto opportune ipotesi, che $\phi_X(t)$ e $p_X(x)$ sono equivalenti e si ha

$$p_X(x) = \frac{1}{2\pi} \int \phi_X(t) e^{-itx} dt .$$

Mostriamo un'importante proprietà della funzione caratteristica: se X_1, X_2, \dots, X_N sono variabili aleatorie indipendenti con funzioni caratteristiche

¹⁰Citiamo il caso di N variabili gaussiane con media m_1, m_2, \dots, m_N e varianza $\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2$; utilizzando la formula

$$\int_{-\infty}^{\infty} e^{-ax^2+bx} dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}} ,$$

è facile mostrare che z è una variabile gaussiana con media $m_1 + m_2 + \dots + m_N$ e varianza $\sigma_1^2 + \sigma_2^2 + \dots + \sigma_N^2$.

Analogamente se n_1, n_2, \dots, n_N sono variabili Poissoniane con parametri $\lambda_1, \lambda_2, \dots, \lambda_N$:

$$P(n_j = k) = \frac{\lambda_j^k}{k!} e^{-\lambda_j} ,$$

allora z è ancora una variabile Poissoniana:

$$P(z = k) = \frac{\Lambda^k}{k!} e^{-\Lambda}$$

con $\Lambda = \lambda_1 + \lambda_2 + \dots + \lambda_N$; questo risultato era stato già ottenuto nella Sez. 5 usando le funzioni generatrici.

$\phi_{X_1}(t), \phi_{X_2}(t), \dots, \phi_{X_N}(t)$ allora per la somma $Z = X_1 + X_2 + \dots + X_N$ si ha

$$\phi_Z(t) = \prod_{j=1}^N \phi_{X_j}(t), \quad (48)$$

la formula precedente, analogamente a quanto visto per la funzione generatrice, si ottiene notando che per variabili indipendenti si ha

$$\phi_Z(t) = E(e^{it(X_1+X_2+\dots+X_N)}) = \prod_{j=1}^N E(e^{itX_j}) .$$

Un'altra semplice (ma utile) proprietà della funzione caratteristica è la seguente: se la variabile X ha come funzione caratteristica $\phi_X(t)$ allora la funzione caratteristica della variabile $Y = aX + b$ (ove a e b sono costanti reali) è

$$\phi_Y(t) = e^{itb} \phi_X(at) . \quad (49)$$

Siamo ora pronti per il teorema del limite centrale (TLC)¹¹: asintoticamente la densità di probabilità della somma di tante variabili indipendenti è una gaussiana. In forma un po' più precisa: se X_1, \dots, X_N sono variabili i.i.d. con media $\langle X \rangle$ e varianza σ^2 , per grandi N la densità di probabilità di

$$Z_N = \frac{1}{\sigma\sqrt{N}} \sum_{n=1}^N (X_n - \langle X \rangle)$$

è la gaussiana a media nulla e varianza unitaria:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

Consideriamo la variabile $Y_N = X'_1 + \dots + X'_N$ ove $X'_j = X_j - \langle X \rangle$ ed indichiamo con $\phi_{X'}$ la funzione caratteristica di X' ; dalla (48) si ha

$$\phi_{Y_N}(t) = [\phi_{X'}(t)]^N ,$$

dalla (49) per la variabile $Z_N = Y_N/(\sigma\sqrt{N})$ si ha:

$$\phi_{Z_N}(t) = \left[\phi_{X'}\left(\frac{t}{\sigma\sqrt{N}}\right) \right]^N .$$

¹¹Il termine *centrale* è da intendersi come importante, fondamentale e si riferisce a teorema e non a limite, sarebbe forse meno ambiguo dire teorema centrale del limite o teorema limite centrale.

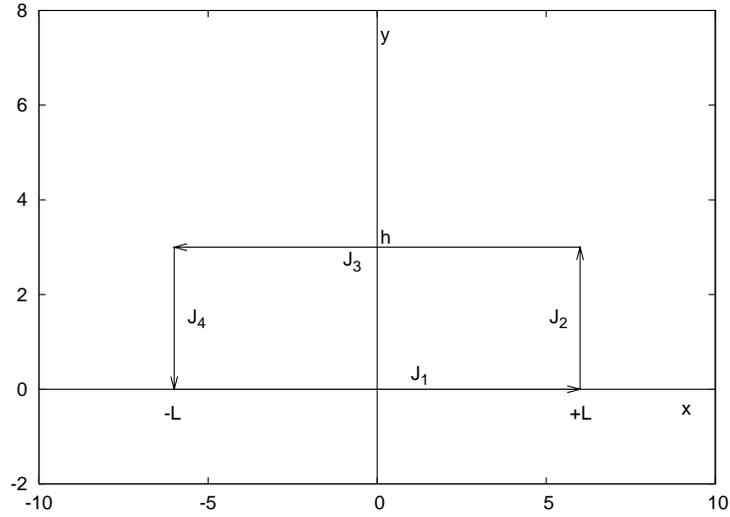


Figure 6: Cammino d'integrazione nel piano complesso per il calcolo della funzione caratteristica della densità gaussiana.

Notiamo che per piccoli valori di t la funzione caratteristica può essere scritta nella forma

$$\phi_X(t) = 1 + it\langle X \rangle - \frac{t^2}{2}\langle X^2 \rangle + O(t^3) ,$$

nel caso della variabile X' che ha media nulla abbiamo

$$\phi_{X'}(t) = 1 - \frac{t^2}{2}\sigma^2 + O(t^3) ,$$

quindi

$$\phi_{Z_N}(t) = \left[1 - \frac{t^2}{2N} + O\left(\frac{t^3}{N^{3/2}}\right) \right]^N$$

che nel limite $N \rightarrow \infty$ diventa¹²

$$\lim_{N \rightarrow \infty} \phi_{z_N} = e^{-\frac{t^2}{2}} .$$

È facile mostrare che $\phi(t) = e^{-\frac{t^2}{2}}$ è la funzione caratteristica della gaussiana

¹²Stiamo assumendo che sia possibile trascurare i termini $O(t^3)$, cioè è corretto se la $p_X(x)$ decade abbastanza rapidamente per grandi $|x|$.

a media nulla e varianza unitaria. Consideriamo la funzione di variabile complessa

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2} + itz}$$

ove $z = x + iy$ e t è una costante reale. Calcoliamo l'integrale di $f(z)$ sulla curva chiusa percorsa in senso antiorario costituito dal rettangolo con lati $J_1 : -L < x < L, y = 0$; $J_2 : 0 < y < h, x = L$; $J_3 : -L < x < L, y = h$ e $J_4 : 0 < y < h, x = -L$, vedi Fig.6. È facile vedere che nel limite $L \rightarrow \infty$ l'integrale su J_2 e J_4 è zero mentre l'integrale su J_1 non è altro che $\phi(t)$ la funzione caratteristica della gaussiana a media nulla e varianza unitaria, in modo analogo un facile calcolo mostra che l'integrale su J_3 è $-\phi(t-h)e^{\frac{h^2}{2}-th}$ poiché la $f(z)$ è analitica l'integrale sul circuito chiuso è zero quindi

$$\phi(t) = \phi(t-h)e^{\frac{h^2}{2}-th} ,$$

ricordando che $\phi(0) = 1$ ponendo $h = t$ si ottiene il risultato

$$\phi(t) = e^{-\frac{t^2}{2}} .$$

Abbiamo quindi che nel limite $N \rightarrow \infty$

$$p_{Z_N}(z) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} .$$

Notare che nel risultato finale i dettagli della $p_X(x)$, a parte $\langle X \rangle$ e σ , sono scomparsi.

In forma matematicamente più precisa abbiamo:

$$\lim_{N \rightarrow \infty} P(a < z_N < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx . \quad (50)$$

Variabili indipendenti.— La condizione che le variabili siano identicamente distribuite non è essenziale, può essere rimossa senza difficoltà, purchè siano indipendenti con varianza limitata: $0 < a < \sigma_j^2 < b < \infty$. Il calcolo si ripete in modo analogo per

$$Z_N = \frac{1}{D_N} \sum_{n=1}^N (X_n - m_n) ,$$

dove m_j è il valor medio della X_j e

$$D_N^2 = \sum_{n=1}^N \sigma_n^2 .$$

Indichiamo con $Y_N = (X_1 - m_1) + (X_2 - m_2) + \dots + (X_N - m_N)$ allora

$$\phi_{Y_N}(t) = \prod_{j=1}^N \phi_{X_j}(t) = \prod_{j=1}^N \left[1 - \sigma_j^2 \frac{t^2}{2} + O(t^3) \right] ,$$

quindi

$$\begin{aligned} \phi_{Z_N}(t) &= \phi_{Y_N} \left(\frac{t}{D_N} \right) = \prod_{j=1}^N \left[1 - \frac{\sigma_j^2}{2} \frac{t^2}{D_N^2} + O(t^3) \right] \\ &= \exp \sum_{n=1}^N \ln \left(1 - \frac{\sigma_n^2}{2} \frac{t^2}{D_N^2} + O(t^3) \right) , \end{aligned}$$

poiché $0 < a < \sigma_j^2 < b < \infty$ si ha che $D_N^2 \sim N$ quindi il coefficiente che moltiplica t^2 è piccolo e si può scrivere

$$\phi_{Z_N}(t) \simeq \exp \left[- \sum_{n=0}^N \frac{\sigma_n^2}{2} \frac{t^2}{D_N^2} \right] = e^{-\frac{t^2}{2}} .$$

Il teorema del limite centrale spiega¹³ il fatto che la distribuzione Gaussiana è presente in situazioni molto diverse ed apparentemente senza alcuna relazione: dalla fisica alla biologia, dall' economia alle scienze sociali. Infatti è piuttosto naturale pensare che in molti casi il valore di una variabile sia il risultato di tante cause indipendenti.

Notiamo che nella dimostrazione del teorema del limite centrale i due ingredienti fondamentali che permettono una dimostrazione semplice sono:

- a) la varianza finita $\sigma_X^2 < \infty$;
- b) l'indipendenza delle variabili $\{X_j\}$.

¹³Diciamo spiega, e non dimostra, in quanto nelle scienze naturali non si ha mai una "vera dimostrazione"; infatti è praticamente impossibile avere il completo controllo delle ipotesi. Ad esempio non è facile avere la certezza empirica dell' indipendenza. Vedremo in seguito che c'è un' altra distribuzione (la lognormale) molto comune nei fenomeni naturali, questa distribuzione ha una stretta connessione con il teorema del limite centrale.

Non è difficile convincersi che se $\sigma_X^2 = \infty$ la somma di tante variabili indipendenti non si avvicina ad una gaussiana. Un esempio facile da trattare analiticamente è il caso di variabili indipendenti la cui distribuzione di probabilità è:

$$p_X(x) = \frac{1}{\pi(1+x^2)} .$$

Questa distribuzione è detta di Cauchy, con un semplice calcolo di analisi complessa si mostra che la sua funzione caratteristica è $\phi_X(t) = e^{-|t|}$. Se si considera la variabile $Y_N = X_1 + \dots + X_N$, dove le $\{X_j\}$ sono indipendenti e distribuite con la funzione di Cauchy, allora $\phi_{Y_N}(t) = e^{-N|t|}$ e quindi la media Y_N/N è distribuita come la singola X indipendentemente da N . Abbiamo quindi che non vale il TCL e neanche la legge dei grandi numeri.

Variabili non indipendenti.– Se le variabili non sono indipendenti intuitivamente ci si aspetta, che se le $\{X_j\}$ sono “debolmente dipendenti” allora il teorema del limite centrale valga ancora e l’unica modifica sia sostituire σ^2 con una varianza efficace σ_{eff}^2 che tenga conto delle correlazioni. Per semplicità di notazione consideriamo il caso con $\langle X_j \rangle = 0$. Assumiamo che la successione X_1, X_2, \dots sia statisticamente stazionaria ¹⁴ e introduciamo la funzione di correlazione $C(k) = E(X_0 X_k) = E(X_n X_{n+k})$. Se le variabili $\{X_j\}$ sono debolmente correlate, cioè se $\sum_{k=1}^{\infty} C(k) < \infty$, allora si può dimostrare che il teorema del limite centrale vale ancora con σ^2 rimpiazzata da

$$\sigma_{eff}^2 = \sigma^2 + 2 \sum_{k=1}^{\infty} C(k) . \quad (51)$$

9.3 Grandi Deviazioni

Nella precedente sezione abbiamo visto che il teorema del limite centrale, sotto opportune ipotesi, dimostra la (50). A volte il TLC viene enunciato dicendo che date N variabili indipendenti per la variabile $Y_N = (X_1 + \dots + X_N)/N$ quando $N \gg 1$ si ha

$$p_{Y_N}(y_N) \simeq \frac{1}{\sqrt{2\pi\sigma^2/N}} e^{-(y_N - m)^2 N / (2\sigma^2)} , \quad (52)$$

questo è sostanzialmente giusto se si aggiunge che $y_N - m$ non deve essere troppo grande diciamo non più di $O(\sigma/\sqrt{N})$. È invece errato assumere la

¹⁴In altre parole il “grado di dipendenza” tra X_k e X_l è funzione solo di $k - l$.

validità della (52) alla lettera, cioè su intervalli arbitrari. Questo può essere chiarito dal seguente esempio. Consideriamo X_1, \dots, X_N variabili i.i.d e limitate $a < X_j < b$. Ci domandiamo la densità di probabilità di

$$Y_N = \prod_{n=1}^N X_n ,$$

in seguito vedremo che questo problema è interessante in diversi contesti. Passando al logaritmo abbiamo

$$Y_N = \exp \left[\sum_{n=1}^N \ln X_n \right] ,$$

essendo le $\{X_n\}$ indipendenti anche le variabili $\{\alpha_j = \ln X_j\}$ lo sono, quindi per la variabile $A_N = \ln Y_N = \alpha_1 + \dots + \alpha_N$, invocando il TLC si conclude che

$$p_{A_N}(a_N) \simeq \frac{1}{\sqrt{2\pi C^2 N}} e^{-(a_N - \langle \alpha \rangle N)^2 / (2C^2 N)} ,$$

dove C^2 è la varianza di α . Utilizzando la regola per il cambio di variabili vista nella Sez. **6.1**, per la Y_N si ha:

$$p_{Y_N}(y_N) \simeq p_{LN}(y_N) = \frac{1}{y_N \sqrt{2\pi C^2 N}} e^{-(\ln y_N - \langle \alpha \rangle N)^2 / (2\sigma^2 N)} , \quad (53)$$

ove $p_{LN}(\)$ è detta distribuzione lognormale. Prendendo alla lettera il risultato precedente, cioè assumendo la validità della lognormale anche fuori dai confini di validità del TLC ($|\delta A_N| < O(C\sqrt{N})$) si ottengono risultati manifestamente inconsistenti. Per esempio, se si usa la (53) per calcolare $E(Y_N^q)$, si ha

$$E_{LN}(Y_N^q) = e^{N(q\langle \alpha \rangle + q^2 C^2 / 2)} , \quad (54)$$

dove LN indica che il valore medio è calcolato con la distribuzione lognormale. Il risultato esatto è invece

$$E(Y_N^q) = [E(X^q)]^N = e^{N \ln E(X^q)}$$

e, poiché $a < X < b$, si ha $E(X^q) < b^q$ e quindi

$$E(Y_N^q) < e^{Nq \ln b} ,$$

in evidente contrasto con la (54) per q sufficientemente grandi. Questo è dovuto al fatto che la vera densità di probabilità di Y_N è esattamente zero per $y_N > b^N$ mentre il contributo dominante per il calcolo di $E_{LN}(y_N^q)$ per grandi q proviene proprio da valori di $y_N > b^N$. In altre parole, indicando con $p_V(y_N)$ la distribuzione di probabilità esatta, è vero che

$$p_{LN}(y_N) \simeq p_V(y_N) ,$$

però non è corretto affermare che

$$p_{LN}(y_N)y_N^q \simeq p_V(y_N)y_N^q ,$$

per valori arbitrari di q . Invece per piccoli valori di q l' approssimazione lognormale (54) è corretta. Basta notare che per q vicino a zero si ha

$$\ln E(X^q) = \ln E(e^{q \ln X}) = \ln E\left(1 + q \ln X + \frac{q^2}{2}(\ln X)^2 + \dots\right) ,$$

ricordando che vale lo sviluppo $\ln(1 + \epsilon) = \epsilon - \epsilon^2/2 + O(\epsilon^3)$ abbiamo

$$\begin{aligned} \ln E(X^q) &= q \langle \ln X \rangle + \frac{q^2}{2} \langle (\ln X - \langle \ln X \rangle)^2 \rangle + O(q^3) \\ &= q \langle \alpha \rangle + C^2 \frac{q^2}{2} + O(q^3) = \ln E_{LN}(X^q) + O(q^3) . \end{aligned}$$

9.4 Oltre il limite centrale: la funzione di Cramer

Dall'esempio precedente è chiara la necessità di andare oltre il TLC, cioè controllare le "grandi deviazioni". Questa teoria è stata introdotta negli anni 30 del XX secolo dal matematico svedese H. Cramer per descrivere la statistica di eventi rari nell' ambito dei rischi assicurativi. L' idea può essere spiegata con un semplice calcolo combinatorio.

Consideriamo una sequenza di lanci di una moneta truccata i cui possibili risultati sono testa (+1), con probabilità p , o croce (-1), con probabilità $1 - p$, ed indichiamo il risultato dell' n -mo lancio con X_n . Se si definisce $Y_N = (X_1 + \dots + X_N)/N$, si ha $\langle Y_N \rangle = 2p - 1$ e $\sigma_{Y_N}^2 = 4p(1 - p)/N$.

Il numero di modi con cui si possono avere k volte testa in N lanci è $N!/ [k!(N - k)!]$ quindi dalla distribuzione binomiale abbiamo

$$P\left(y_N = \frac{2k}{N} - 1\right) = \frac{N!}{k!(N - k)!} p^k (1 - p)^{N - k} . \quad (55)$$

Usando l'approssimazione di Stirling $n! \simeq n^n e^{-n} \sqrt{2\pi n}$ e scrivendo $k = fN$ and $N - k = (1 - f)N$ ove $f = k/N$ è la frequenza dell'evento testa in N lanci si ha

$$P(y_N = 2f - 1) \sim e^{-NI(p,f)}, \quad (56)$$

dove

$$I(p, f) = f \ln \frac{f}{p} + (1 - f) \ln \frac{1 - f}{1 - p}. \quad (57)$$

La quantità $I(p, f)$ è chiamata “entropia relativa” (o divergenza di Kullback-Leibler), e gode delle seguenti proprietà:

- $I(p, f) = 0$ se $f = p$,
- $I(p, f) > 0$ se $f \neq p$.

È facile ripetere l'argomento nel caso multinomiale ove le X_1, \dots, X_N possono prendere m possibili diversi valori a_1, a_2, \dots, a_m con probabilità p_1, p_2, \dots, p_m . Nel limite $N \gg 1$, la probabilità di osservare le frequenze f_1, f_2, \dots, f_m è

$$P_N(\{f_j\}) \sim e^{-NI(\{p\}, \{f\})}$$

dove

$$I(\{p\}, \{f\}) = \sum_{j=1}^m f_j \ln \frac{f_j}{p_j},$$

è l'entropia relativa delle probabilità $\{f\}$, rispetto alle probabilità $\{p\}$. Questa quantità misura la “distanza”¹⁵ tra $\{p\}$ and $\{f\}$ nel senso che $I(\{p\}, \{f\}) = 0$ se e solo se $\{p\} = \{f\}$, e $I(\{p\}, \{f\}) > 0$ se $\{p\} \neq \{f\}$.

Dal calcolo precedente si capisce come sia possibile andare oltre la teoria del limite centrale e controllare le proprietà statistiche degli eventi estremi

¹⁵In effetti non è una vera distanza in senso tecnico. Dati due vettori \mathbf{x} e \mathbf{y} un funzione $d(\mathbf{x}, \mathbf{y})$ è una distanza se

- a) $d(\mathbf{x}, \mathbf{y})$ è positiva a parte il caso $\mathbf{x} = \mathbf{y}$ in cui è zero;
- b) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$;
- c) $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$;

l'ultima disuguaglianza (detta triangolare) non vale per l'entropia relativa. Non vale neanche la b) ma questo non è un problema grave basta simmetrizzare le cose e considerare

$$\frac{1}{2}[I(\{p\}, \{f\}) + I(\{f\}, \{p\})].$$

(code della distribuzione di probabilità) per $N \gg 1$. Scrivendo $I(f, p)$ in termini di $y_N = 2f - 1$, l'eq. (56) diventa

$$p_{Y_N}(y_N) \sim e^{-NC(y_N)} , \quad (58)$$

con

$$C(y) = \frac{1+y}{2} \ln \frac{1+y}{2p} + \frac{1-y}{2} \ln \frac{1-y}{2(1-p)} .$$

La $C(y)$ è detta funzione di Cramer. Per valori di f prossimi a p , e quindi $y_N \simeq \langle Y_N \rangle$, lo sviluppo di Taylor mostra che

$$C(y_N) \simeq \frac{(y_N - \langle Y_N \rangle)^2}{2\sigma^2} ,$$

con $\sigma^2 = 4p(1-p)$, in accordo con quanto ci si aspetta dal teorema del limite centrale, eq. (52).

L'Eq. (58) ha una validità generale (nell'ambito di variabili i.i.d.) e può essere ottenuta con un diverso approccio che permette di esprimere la $C(y)$, che caratterizza la variabile $Y_N = (X_1 + \dots + X_N)/N$, in termini dei momenti della variabile X . In particolare è possibile mostrare che la funzione di Cramer $C(y)$ può essere scritta come una trasformata di Legendre:

$$C(y) = \sup_q \left[qy - L(q) \right] , \quad (59)$$

con $L(q)$ “funzione generatrice dei cumulanti”:

$$L(q) = \ln E(e^{qX}) . \quad (60)$$

Accenniamo all'argomento. Consideriamo i momenti $E(e^{qNY_N})$ che possono essere scritti in due diversi modi:

$$\begin{aligned} E(e^{qNY_N}) &= E(e^{qX})^N = e^{NL(q)} \\ E(e^{qNY_N}) &= \int e^{qNy_N} p_{Y_N}(y_N) dy_N \sim \int e^{[qy - C(y)]N} dy , \end{aligned} \quad (61)$$

nel limite di grandi N , usando il metodo di Laplace (v. Sez. 9), si ottiene

$$L(q) = \sup_y \left[qy - C(y) \right] , \quad (62)$$

che è l'inverso della (59).

Poichè è possibile mostrare che $C(y)$ è una funzione concava ($d^2C/dy^2 \geq 0$) le equazioni (59) e (62) sono equivalenti.

Notiamo che la funzione di Cramer deve obbedire ad alcuni vincoli:

- a) $C(y) > 0$ per $y \neq \langle y \rangle$;
- b) $C(y) = 0$ per $y = \langle y \rangle$;
- c) se y è vicino a $\langle y \rangle$ si ha $C(y) \simeq (y - \langle y \rangle)^2 / (2\sigma^2)$, ove $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$;
- d) $C(y)$ è una funzione concava ($d^2C/dy^2 \geq 0$).

Ovviamente a) e b) sono espressione della legge dei grandi numeri e la c) non è altro che il teorema del limite centrale. La d) è meno intuitiva, in seguito vedremo il suo significato in meccanica statistica.

9.5 Grandi e piccole fluttuazioni in meccanica statistica

Come accennato precedentemente (Sez. 7.2) in meccanica statistica le grandi deviazioni appaiono in modo naturale nel problema delle fluttuazioni di energia per particella di un sistema con N particelle a temperatura T :

$$p(e) \simeq \frac{1}{\mathcal{C}_N} \exp \left\{ -N \left(\frac{1}{k_B T} \right) [e - Ts(e)] \right\}, \quad (63)$$

dove $s(e)$ è la densità di entropia microcanonica. Poiché $\int p(e)de = 1$, la quantità \mathcal{C}_N (funzione di partizione) è un fattore di normalizzazione che si può esprimere come:

$$\mathcal{C}_N \sim \exp \left\{ -N \left(\frac{1}{k_B T} \right) f(T) \right\},$$

dove $f(T)$, energia libera per particella, è definita dalla relazione

$$f(T) = \min_e \{e - Ts(e)\}.$$

Il valore e^* , per il quale la funzione $e - Ts(e)$ è minima, è determinato dall'equazione

$$\frac{1}{T} = \frac{\partial s(e)}{\partial e}, \quad (64)$$

cioè il valore dell'energia tale che il corrispondente insieme microcanonico ha la temperatura T . È quindi chiaro come si scrive la funzione di Cramer per

la distribuzione (63) e qual è il suo significato fisico:

$$C(e) = \left(\frac{1}{k_B T} \right) \left[e - T s(e) - f(T) \right] .$$

Notare che il valore di e per il quale $C(e)$ è minimo (zero) è proprio $e^* = \langle e \rangle$ determinato dalla (64). L'approssimazione gaussiana intorno a e^* è

$$C(e) \simeq \frac{1}{2} C''(e^*) (e - e^*)^2 ,$$

quindi $\langle (e - e^*)^2 \rangle = 1/[N C''(e^*)]$ ricordando che

$$\langle (e - e^*)^2 \rangle = \frac{k_B}{N} T^2 c_V , \quad (65)$$

ove $c_V = \partial \langle e \rangle / \partial T$ è il calore specifico per particella, abbiamo che la concavità della $C(e)$ ha un chiaro corrispondente fisico: $c_V(T)$ essere positivo in modo tale che $\langle (e - e^*)^2 \rangle$ sia positivo.

9.6 Non sempre tante cause indipendenti portano alla gaussiana: la distribuzione lognormale.

Consideriamo il seguente processo moltiplicativo:

$$m_n = X_n m_{n-1} = \left[\prod_{j=1}^n X_j \right] m_0 \quad (66)$$

ove X_j sono variabili aleatorie i.i.d., positive e limitate. Utilizzando il risultato della Sez.8.3, abbiamo che in prima approssimazione la quantità $Y_N = m_N/m_0$ per $N \gg 1$ ha una distribuzione di probabilità lognormale:

$$p_{Y_N}(y_N) \simeq p_{LN}(y_N) = \frac{1}{y_N \sqrt{2\pi C^2 N}} e^{-(\ln y_N - \langle \alpha \rangle N)^2 / (2C^2 N)}$$

ove $\langle \alpha \rangle = \langle \ln X \rangle$ e C^2 è la varianza di $\alpha = \ln X$. Ovviamente valgono le considerazioni precedentemente fatte per le grandi deviazioni, quindi per valori estremi di $\ln Y_N/N - \langle \alpha \rangle$ è necessaria una trattazione più dettagliata in termini della funzione di Cramer che dipende della densità di probabilità della X .

È interessante il fatto che la distribuzione lognormale è presente in molti situazioni: dalla geologia alla biologia e la finanza. Ad esempio descrive in modo ragionevolmente accurato:

- a) il prezzo delle assicurazioni contro incendi ed incidenti industriali;
- b) i giorno di malattia dei lavoratori di un'azienda in un dato periodo;
- c) il numero di batteri sulle foglie;
- d) la grandezza di particelle di terriccio;
- e) la massa dei pezzi di carbone (ed altri minerali) estratti nelle miniere;
- f) la densità di energia dissipata in turbolenza sviluppata.

Non esiste un spiegazione universalmente accettata per questa diffusa presenza della distribuzione lognormale. Comunque è possibile dare un argomento di plausibilità basato sui processi moltiplicativi (66) che sono piuttosto comuni. Come esempio possiamo pensare a m_n come massa di un sasso in montagna. Si può assumere che i sassi presenti in superficie siano il risultato di fragmentazioni che accadono con frequenza annuale: l'acqua si infiltra nelle fenditure e durante l'inverno gelando il masso si può rompere, ad esempio rimane intatto con probabilità p oppure dividersi in due pezzi uguali con probabilità $1 - p$, abbiamo quindi $x = 1$ con probabilità p e $x = 1/2$ con probabilità $1 - p$. Pensando a questo processo ripetuto su tanti anni la (66) è un modello di formazione dei sassi in montagna o dei granelli di sabbia, la cui massa segue, con buona approssimazione, la distribuzione lognormale.

Possiamo formalizzare la plausibilità del processo moltiplicativo (66) con il seguente modello probabilistico. Indichiamo con $N_k(x)$ il numero di particelle (sassi) di massa minore di x dopo k eventi di fragmentazione. Sia $M_k(x) = E(N_k(x))$ e $B_k(x|y)$ il numero medio di particelle di massa minore di x generate nella k -ma fragmentazione da particelle di massa y , abbiamo

$$M_k(y) = \int_0^\infty B_k(y|x) dM_{k-1}(x) . \quad (67)$$

Assumendo che il processo di fragmentazione sia indipendente dalla scala¹⁶ cioè che $B_k(x|y)$ dipende solo dal rapporto y/x :

$$B_k(x|y) = C_k\left(\frac{y}{x}\right) ,$$

¹⁶Questa assunzione non è sempre realistica: in molti casi le particelle più piccole si rompono più difficilmente. La validità di questa ipotesi in genere è ristretta a valori del rapporto y/x in un opportuno range.

abbiamo

$$M_k(y) = \int_0^\infty C_k\left(\frac{y}{x}\right) dM_{k-1}(x) .$$

Derivando rispetto ad y l'equazione precedente e notando che $dM_k(x) = \text{cost.} p_k(x) dx$ ove $p_k(x)$ è la densità di probabilità della x dopo k frammentazioni abbiamo

$$p_k(y) = \int_0^\infty g_k\left(\frac{y}{x}\right) p_{k-1}(x) \frac{1}{x} dx , \quad (68)$$

ove g_k è la derivata di C_k .

Non è difficile mostrare che il precedente processo di frammentazione non è altro che un processo moltiplicativo. Consideriamo due variabili indipendenti x_1 e x_2 con distribuzione di probabilità p_1 e p_2 rispettivamente. La distribuzione di probabilità per la variabile $z = x_1 x_2$ è data da

$$p_z(z) = \int \int p_1(x_1) p_2(x_2) \delta(z - x_1 x_2) dx_1 dx_2 ,$$

utilizzando le note proprietà della delta di Dirac si ha

$$p_z(z) = \int p_1(x_1) p_2\left(\frac{z}{x_1}\right) \frac{1}{x_1} dx_1 .$$

La (68) non è altro che la formula per la distribuzione di probabilità del prodotto di due variabili indipendenti, e quindi il processo moltiplicativo (66) è giustificato sotto l'ipotesi di invarianza di scala, cioè $B_k(x|y) = C_k(y/x)$.

10 Funzioni gamma di Eulero e approssimazione di Stirling

Il metodo di Laplace

Consideriamo l'integrale

$$I = \int_a^b e^{Nf(x)} dx$$

ove $N \gg 1$ ed $f(x)$ ha un massimo quadratico in $x_0 \in [a, b]$. È facile convincersi che il contributo ad I proviene dalla regione intorno ad x_0 , quindi approssimando $f(x)$ con lo sviluppo di Taylor:

$$f(x) \simeq f(x_0) - \frac{1}{2} |f''(x_0)| (x - x_0)^2$$

si ha

$$I \simeq e^{Nf(x_0)} \int_a^b e^{-\frac{N}{2}|f''(x_0)|(x-x_0)^2} dx$$

a questo punto poichè $x_0 \in [a, b]$, ed $N \gg 1$ si può approssimare I come

$$I \simeq e^{Nf(x_0)} \int_{-\infty}^{\infty} e^{-\frac{N}{2}|f''(x_0)|(x-x_0)^2} dx$$

in quanto i contributi per $x < a$ e $x > b$ sono esponenzialmente piccoli. Usando la ben nota formula dell' integrale gaussiano si ottiene

$$I \simeq e^{Nf(x_0)} \sqrt{\frac{2\pi}{|f''(x_0)|N}} .$$

La funzione gamma

Nel calcolo delle probabilità (in particolare in meccanica statistica) spesso intervengo le cosiddette funzioni gamma di Eulero

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt .$$

Ci limitiamo al caso con x reale e positivo. È immediato verificare che

$$\Gamma(1) = 1 , \quad \Gamma(x+1) = x\Gamma(x)$$

e quindi, per valori di x interi si ha $\Gamma(n+1) = n!$.

Esercizio.— La funzione gamma interviene nel calcolo del volume delle ipersfere di dimensione D :

$$V_D(R) = \int_{\sum_{j=1}^D x_j^2 \leq R^2} dx_1 \cdots dx_D = C_D R^D$$

dove C_D è il volume dell' ipersfera di dimensione D e raggio unitario. La determinazione di C_D in termini delle funzioni gamma si può ottenere nel seguente modo. Consideriamo l'integrale

$$I_D = \int_{-\infty}^{+\infty} e^{-\sum_{j=1}^D x_j^2} dx_1 \cdots dx_D = \pi^{\frac{D}{2}} .$$

Notiamo che $dV_D(R) = DC_D R^{D-1} dR$, quindi I_D si può scrivere nella forma

$$I_D = DC_D \int_0^\infty R^{D-1} e^{-R^2} dR$$

con il cambio di variabili $x = R^2$ si ha

$$I_D = \frac{D}{2} C_D \int_0^\infty x^{D/2-1} e^{-x} dx = \frac{D}{2} C_D \Gamma\left(\frac{D}{2}\right),$$

da cui

$$C_D = \frac{\pi^{\frac{D}{2}}}{\frac{D}{2} \Gamma\left(\frac{D}{2}\right)} = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2} + 1\right)}.$$

L' approssimazione di Stirling

È facile rendersi conto che $N! = \Gamma(N + 1)$ cresce molto rapidamente con N (ad esempio $5! = 120$, $20! \simeq 2.432 \times 10^{18}$) ed è quindi importante avere un' espressione (anche approssimata) di $N!$ per grandi valori di N . La risposta a questo problema è l' approssimazione di Stirling che si ottiene con il metodo di Laplace.

Scriviamo $N!$ nella forma

$$N! = \Gamma(N + 1) = \int_0^\infty t^N e^{-t} dt,$$

introducendo la variabile $z = t/N$ abbiamo

$$N! = N^{N+1} \int_0^\infty e^{N(\ln z - z)} dz.$$

Per grandi N l' integrale può essere calcolato (approssimativamente) con il metodo di Laplace.

Utilizzando questo metodo di approssimazione si ottiene:

$$N! = \Gamma(N + 1) \simeq N^N e^{-N} \sqrt{2\pi N}$$

questa approssimazione (detta di Stirling) è molto precisa anche per piccoli valori di N , ad esempio per $N = 2, 3, 20, 40$ e 100 per $N!$ si ottiene $1.91, 5.95, 2.42 \times 10^{18}, 8.14 \times 10^{47}$ e 9.32×10^{157} , da confrontare con i valori esatti $2, 6, 2.43 \times 10^{18}, 8.16 \times 10^{47}$ e 9.33×10^{157} rispettivamente.

Un' espressione più accurata, che migliora l' approssimazione di Stirling, è la seguente:

$$N! = \Gamma(N + 1) \simeq N^N e^{-N} \sqrt{2\pi N} \left[1 + \frac{1}{12N} + \frac{1}{144N^2} + O(N^{-3}) \right].$$

11 Bibliografia

Queste note sono basate sui primi tre capitoli del libro:

G. Boffetta e A. Vulpiani *Probabilità in fisica*, Springer Italia, 2012

Un' introduzione alla probabilità particolarmente adatta a studenti di fisica:

B.V. Gnedenko *The theory of probability*, MIR Ed.

Il libro di probabilità da noi preferito tra quelli non troppo avanzati: A. Renyi
Probability Theory, Dover;

in francese *Calcul des probabilités*, Jacques Gabay Ed.

Un libro di meccanica statistica con un bel capitolo introduttivo sulla probabilità:

L.E. Reichl *A Modern Course in Statistical Physics*, Wiley.

Per approfondire gli aspetti concettuali:

A.N. Kolmogorov *Grundbegriffe der Wahrscheinlichkeitsrechnung*, (1933)

Traduzione inglese *Foundations of the Theory of Probability* (1933) Chelsea
Publ. Comp. (1956).

Consultabile gratuitamente su:

<http://www.kolmogorov.com/Foundations.html>