

Introduzione alla probabilità

September 26, 2022

La teoria della probabilità insegna a predire la probabilità di un evento, scelto tra una classe di possibili risultati. In altre parole, in una sequenza (grande) di eventi possibili, quale frazione di essi è consistente con il caso a cui siamo interessati. Da un punto di vista pratico, vorremmo sapere come assegnare i valori di probabilità ai vari risultati. Esistono due approcci possibili

- 1. Le probabilità oggettive si ottengono sperimentalmente dalla frequenza relativa del verificarsi dell'esito in molte prove della variabile casuale. Se il processo casuale viene ripetuto N volte e l'evento A si verifica N_A volte allora

$$P(A) = \frac{N_A}{N}$$

Per esempio, una serie di $N = 100, 200, 300$ lanci di dadi può dare come risultato 19, 30, 48 occorrenze della faccia contrassegnata con 1, il rapporto 0.19, 0.15, 0.16 fornisce una stima sempre più affidabile della probabilità.

- 2. Le probabilità soggettive forniscono una stima teorica basata sulle incertezze legate alla mancanza di una conoscenza precisa degli esiti. Ad esempio, la valutazione $P(1) = 1/6$ si basa sulla consapevolezza che ci sono sei possibili esiti di un dado e che, in assenza di qualsiasi ragione precedente per credere che il dado sia truccato, tutti e sei i casi sono ugualmente probabili. Tutte le assegnazioni di probabilità nella meccanica statistica sono basate su criteri soggettivi. Le conseguenze di tali assegnazioni soggettive di probabilità devono essere verificate in base alle misurazioni e possono dover essere modificate man mano che si rendono disponibili ulteriori informazioni sul risultato.

Da PROB_2021.pdf (Vulpiani)

L' iniziatore della sistematizzazione sia tecnica che concettuale della teoria delle probabilità e' stato E. Borel. Il programma di formalizzazione può essere considerato concluso nel 1933 con la pubblicazione del libro di A.N. Kolmogorov *Concetti fondamentali di teoria delle probabilità*

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 . Una famiglia di sottoinsiemi e' costituita da tutti gli elementi che si possono creare combinando gli eventi elementari ω .

Nel caso di un dado, $\omega = 1, 2, 3, 4, 5, 6$. \mathcal{F} continene $\{1\}, \{2\}, \dots, \{1 \cap 2\}, \{1 \cup 2\}, \{\bar{2}\}$ etc

Chiamiamo Ω lo spazio degli eventi ed eventi gli elementi di \mathcal{F} :

Gli assiomi dicono che

1. \mathcal{F} è una algebra di insiemi, cioè $\Omega \in \mathcal{F}$, ed \mathcal{F} è chiuso rispetto alla operazione di unione (\cup), intersezione (\cap) e complemento ($\bar{}$); cioè se $A \in \mathcal{F}$, e $B \in \mathcal{F}$, allora anche $A \cap B, A \cup B$ e $\bar{A} \equiv \Omega - A$ sono contenuti in \mathcal{F} . Per ricordarlo notate che (\cup) indica "o" (si realizza A oppure B) mentre (\cap) indica "e" (si realizza sia A che B)
2. Ad ogni elemento A di \mathcal{F} si associa un numero reale non negativo (probabilità di A) $P(A)$.
3. $P(\Omega) = 1$
4. Se due insiemi A e B sono disgiunti (cioè $A \cap B = \emptyset$) allora $P(A \cup B) = P(A) + P(B)$.

E' intuitivo mostrare che $P(\bar{A}) = 1 - P(A)$, $P(\emptyset) = 0$, $0 \leq P(A) \leq 1$.

Quindi abbiamo bisogno di tre quantita' per sviluppare una teoria della probabilita'. Gli eventi elementari, le combinazioni tra gli eventi e l' associazione a ciascuna di queste combinazioni di un valore compreso tra zero ed uno. La terna (ω, \mathcal{F}, P) è detta spazio di probabilità.

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. Nel seguito quindi guardiamo come gli assiomi riflettono la frequenza dei risultati. Assumiamo di poter effettuare un numero praticamente illimitato di ripetizioni, e studiamo il risultato delle realizzazioni.

- Assioma 1: L'assioma 1 specifica gli "oggetti" per i quali ha senso definire la probabilità. Ad esempio, se il caso studiato 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.
- Assioma 2: Le proprietà della probabilità di un evento $P(A)$ devono essere tali che:
 - a) si è praticamente certi che se l'esperimento 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 dell'esperimento.
- Assioma 3: Poiché $0 \leq M/N \leq 1$ e per l'evento Ω si ha sempre $M = N$ e' naturale l'assioma 3.
- Assioma 4: 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 4.

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 $[a, b)$.

E comodo introdurre la funzione di distribuzione:

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

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

$$p_X(x) = \frac{d\mathcal{F}(x)}{dx}$$

o alternativamente

$$\mathcal{F}(x) = \int_{-\infty}^x p_X(x') dx'$$

e dunque $p_X(x)dx$ e' la probabilita' che l'evento sia tra x e $x + dx$.

A voler essere rigorosi la definizione di densità di probabilità ha senso solo se $\mathcal{F}(x)$ 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.

Definizione di misura

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

1. se A_1, A_2 sono insiemi disgiunti e misurabili allora anche la loro unione $A_1 \cup A_2$ è misurabile e $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$
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 $\mu(B - A) = \mu(B) - \mu(A)$
3. un certo insieme E ha misura 1: $\mu(E) = 1$.
4. se due insiemi misurabili sono congruenti hanno la stessa misura.

Altre definizioni importanti

- Probabilità condizionata:

Con $P(A|B)$ indichiamo la probabilità condizionata, cioè che si realizzi l' evento A sapendo che si è realizzato l' evento B .

Per esempio la probabilita' che esca nel lancio di un dado il numero 3 sapendo che e' uscito un numero dispari

Se guardiamo in termini di frequenze, il numero di casi che concorrono ad $A|B$ (M_{AB}) sono quelli dell' intersezione di A con B ($A \cap B$) mentre il numero di casi di B e' M_B . Quindi

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

e se dividiamo e moltiplichiamo per il numero totale di casi M

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

La probabilita' $P(A \cap B)$ che esca il 3 ($A \cap B$) e' $1/6$, la probabilita' $P(B)$ che esca un numero dispari e' $1/2$. Quindi $P(A|B) = 1/3$

- **Eventi Indipendenti:**

Se la probabilita' condizionata di A dato B ($P(A|B)$) si riferisce ad eventi indipendenti (cioe' in cui il risultato di A non dipende in alcun modo da B), allora $P(A|B) = P(A)$.

Per esempio, se lancio due dati colorati in modo diverso (rosso e verde), la probabilita' che esca 3 nel dado rosso non dipende dalla faccia del dado verde

Ne consegue che per eventi indipendenti

$$P(A \cap B) = P(A)P(B)$$

Per esempio, la probabilita' che esca 3 nel dado rosso e 3 nel dado verde e' $P(A) = 1/6$, $P(B) = 1/6$ e $P(A \cap B) = 1/36$ (l'insieme Ω ha 36 casi e $A \cap B$ e' uno solo di questi)

- **Formula di Bayes**

La probabilita' condizionata di A dato B e la probabilita' condizionata di B dato A sono rispettivamente

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

e poiche' $P(A \cap B) = P(B \cap A)$ ne consegue, mettendo insieme le due definizioni che

$$P(A|B)P(B) = P(B|A)P(A)$$

che prende il nome di formula di Bayes.

- **Completezza della probabilita'**

Tasselliamo in modo completo lo spazio con insiemi indipendenti, cioe'

$$B_i \cap B_j = \emptyset \quad \sum_j B_j = \Omega$$

L'insieme A può essere scritto come

$$A = A \cap \Omega = A \cap \sum_j B_j = \sum_j A \cap B_j$$

Se passiamo al calcolo della probabilità, abbiamo

$$P(A) = \sum_j P(A \cap B_j) = \sum_j P(A|B_j)P(B_j)$$

dove abbiamo sfruttato il fatto che $A \cap B_j$ sono eventi indipendenti (visto che $B_i \cap B_j = \emptyset$).

- Proprietà della moltiplicazione

Se consideriamo

$$P(A_1 \cap A_2 \cap A_3 \dots \cap A_n)$$

questa, utilizzando Bayes si può scrivere come

$$P(A_1 \cap A_2 \cap A_3 \dots \cap A_n) = P(A_n | A_1 \cap A_2 \cap A_3 \dots \cap A_{n-1}) P(A_1 \cap A_2 \cap A_3 \dots \cap A_{n-1})$$

ed iterando

$$= P(A_n | A_1 \cap A_2 \cap A_3 \dots \cap A_{n-1}) P(A_{n-1} | A_1 \cap A_2 \cap A_3 \dots \cap A_{n-2}) \dots P(A_2 | A_1) P(A_1)$$

Per convincersi, risolviamo il caso più semplice di $n = 3$

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 \cap [A_1 \cap A_2]) = P(A_3 | A_1 \cap A_2) P(A_1 \cap A_2) \\ &= P(A_3 | A_1 \cap A_2) P(A_2 \cap A_1) = P(A_3 | A_1 \cap A_2) P(A_2 | A_1) P(A_1) \end{aligned}$$

Problemi x capire Bayes

Date di nascita

Stimiamo la probabilità che se abbiamo N persone (con N grande, ma piccolo rispetto a 365), almeno due di esse festeggino il compleanno lo stesso giorno. Definiamo con A l'evento che almeno due persone siano nate lo stesso giorno. Se calcoliamo l'evento complementare \bar{A} (nessuno nato lo stesso giorno)

Indicizziamo tutti i partecipanti per calcolare \bar{A} . Il primo è nato in un giorno arbitrario. Il secondo, affinché non sia nato nello stesso giorno del precedente, deve essere nato in uno dei 365-1 giorni rimanenti, il terzo in 365-2 etc. Quindi

$$P(\bar{A}) = \left(\frac{365-1}{365}\right) \left(\frac{365-2}{365}\right) \left(\frac{365-3}{365}\right) \dots$$

$$\prod_{k=1}^{N-1} \left(1 - \frac{k}{365}\right)$$

Se ora N e' grande

$$\ln P(\bar{A}) = \sum_{k=1}^{N-1} \ln \left(1 - \frac{k}{365}\right) \approx - \sum_{k=1}^{N-1} \frac{k}{365} = - \frac{1}{365} \frac{N(N-1)}{2}$$

[ricordando che $(1+2+\dots+N-1)$ sommata alla stessa somma riscritta in ordine decrescente $(N-1)+(N-2)+\dots+1$ e' uguale a 2 volte $(N-1)$ volte N]

Quindi

$$P(\bar{A}) = e^{-\frac{N(N-1)}{2} \frac{1}{365}} \quad P(A) = 1 - P(\bar{A})$$

Nel caso in cui $N = 60$, $P(A) = 0.99$.

Il problema delle 3 porte

Abbiamo 3 stanze (1,2 e 3) e dietro una delle 3 porte c'e', con uguale probabilita', una macchina, messa a caso da un venditore. Per un evento speciale, il venditore chiede al compratore di scegliere una porta. Se la macchina e' li, il compratore avra' un grosso sconto. Assumiamo che il compratore sceglie la porta numero 1. Prima di verificare se dietro la porta numero 1 c'e' la macchina, il venditore apre la porta numero 2, che sa essere vuota, e chiede al compratore se vuole cambiare scelta. Dobbiamo capire se conviene cambiare porta o meno.

Vediamo gli eventi

- Dove sta la macchina (A_1, A_2, A_3)
- Prima scelta del compratore (B_1, B_2, B_3)
- Porta aperta dal venditore (C_1, C_2, C_3)

Le due probabilita' che dobbiamo valutare per capire se conviene cambiare la scelta sono

$$P(A_1|B_1C_2) \quad \text{e} \quad P(A_3|B_1C_2)$$

a partire dal fatto che

$$P(A_i) = \frac{1}{3} \quad \text{e} \quad P(B_i) = \frac{1}{3}$$

visto che la scelta di dove mettere la macchina e la scelta della porta da parte del venditore sono uguali.

Poi sappiamo che il venditore, apre la porta 2 con le seguenti probabilita' condizionate, rispettivamente associate ai casi in cui la macchina e' dietro la porta 1, 2 o 3 (sempre a parita' del fatto che il compratore abbia scelto la porta 1).

$$P(C_2|B_1A_1) = \frac{1}{2} \quad e \quad P(C_2|B_1A_2) = 0 \quad e \quad P(C_2|B_1A_3) = 1$$

Abbiamo applicando Bayes (a B_1 fisso)

$$P(A_1|B_1C_2)P(C_2|B_1) = P(C_2|A_1B_1)P(A_1|B_1)$$

inoltre poiché la scelta di dove mettere la macchina e' precedente alla scelta del compratore $P(A_1|B_1) = P(A_1)$.

Se facciamo lo stesso per $P(A_3|B_1C_2)$ (a B_1 fisso)

$$P(A_3|B_1C_2)P(C_2|B_1) = P(C_2|A_3B_1)P(A_3|B_1)$$

e anche qui $P(A_3|B_1) = P(A_3)$

Notate che siccome B_1 e' fissato, potevamo anche scrivere direttamente

$$P(A_1|C_2)P(C_2) = P(C_2|A_1)P(A_1)$$

e

$$P(A_3|C_2)P(C_2) = P(C_2|A_3)P(A_3)$$

Se ora confrontiamo le due probabilita', notiamo che essendo $P(C_2|B_1)$ uguale in entrambe ed essendo $P(A_1|B_1) = P(A_3|B_1) = 1/3$, la differenza tra le due probabilita' e' tutta in $P(C_2|A_1B_1)$ e $P(C_2|A_3B_1)$. Abbiamo visto che di queste due probabilita', la prima vale 1/2 e la seconda 1. Quindi facendo il rapporto

$$\frac{P(A_1|C_2)}{P(A_3|C_2)} = \frac{P(C_2|A_1)}{P(C_2|A_3)} = \frac{1}{2}$$

o equivalentemente

$$P(A_3|B_1C_2) > P(A_1|B_1C_2)$$

e dunque conviene cambiare porta.

Un ultimo esempio

Supponiamo che una malattia abbia un tasso di incidenza del $1/1000$. Questo vuol dire che su 1000 individui in media 1 e' malato. Viene creato un test che e' capace di rilevare la malattia. Il test fornisce risultato positivo se effettivamente il paziente e' malato, ma fornisce anche per errore (falso positivo) un risultato positivo $p_{fp} = 0.05$ volte quando il paziente non e' malato. Quale e' la probabilita' che chi viene sottoposto al test e risulta positivo, sia veramente malato?

Vediamo subito che, chiamando S e M l'essere sano o malato, e chiamando con p e n l'esito (positivo o negativo) del test i dati del testo si riscrivono come

$$P(M) = \frac{1}{1000} \quad P(S) = 1 - P(M) \quad P(p|S) = 0.05 \quad P(p|M) = 1$$

e, sfruttando Bayes, la quantita' a cui siamo interessati ($P(M|p)$) sara' scrivibile come

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

La probabilita' che l'esito e' positivo e' la somma della probabilita' che il paziente sia malato piu' la probabilita' di osservare un falso positivo

$$P(p) = P(M) + 0.05P(S) = \frac{1}{1000} + 0.05 \frac{999}{1000}$$

Quindi alla fine

$$P(M|p) = \frac{P(M)}{P(p)} = \frac{1}{1 + 0.05 \frac{P(S)}{P(M)}} = \frac{1}{1 + 0.05 \times 999} \approx 2\%$$

Il test quindi e' inutile. Senza il test, una persona scelta a caso e' malata con probabilita' 0.001. Dopo aver fatto il test (inutile) ha una probabilita' di essere malata 20 volte maggiore !

Approssimando $P(S) \approx 1$ (malattia rara), e se volessimo una formula in termini di $P(M)$ e p_{fp}

$$P(M|p) = \frac{P(M)}{P(p)} = \frac{1}{1 + \frac{p_{fp}}{P(M)}}$$

Ne consegue che tanto piu' rara e' la malattia, tanto piu' bassa deve essere la probabilita' di falsi positivi.

Variabili continue — Da fare DOPO !!!

Abbiamo visto che nel caso della variabili continue si fa uso della densita' di probabilita'. Nel seguito avremo bisogno di capire come si trasformano le probabilita', quando la variabile di interesse e' una funzione degli eventi. Un caso tipico e' per esempio l' energia di una particella, che e' funzione di q, p .

Nel caso in cui $X = X(q, p)$, la probabilita' che X assuma il valore x e' scrivibile come

$$P_X(x) = \int dq^{3N} dp^{3N} \rho(p, q) \delta(X(q, p) - x)$$

dove praticamente sommiamo su tutti gli eventi per i quali $X(q, p) = x$. Se avessimo da calcolare la probabilita' di due quantita'

$$P_{XY}(x, y) = \int dq^{3N} dp^{3N} \rho(p, q) \delta(X(q, p) - x) \delta(Y(q, p) - y)$$

Nel caso in cui noi conoscessimo la $P_{XY}(x, y)$ ma fossimo interessati solo alla $P_X(x)$, possiamo operare una *marginalizzazione*

$$P_X(x) = \int dy P_{XY}(x, y) = \int dq^{3N} dp^{3N} \rho(p, q) \delta(X(q, p) - x) \int dy \delta(Y(q, p) - y) = \int dq^{3N} dp^{3N} \rho(p, q) \delta(X(q, p) - x)$$

dove abbiamo sfruttato che $\int dy \delta(y - y_0) = 1$.

Questo naturalmente si generalizza a molte variabili.

Come esempio semplice di marginalizzazione consideriamo un gas ideale, in cui none essendoci interazioni possiamo aspettarci (lo sara' piu' chiaro in seguito) che la probabilita' della posizione e velocita' di un atomo non dipenda dalla posizione e velocita' degli altri atomi.

La probabilita' di trovare la particella 1 in q_1, p_1 sara' scrivibile (marginalizzando)

$$\rho_i(q_i, p_i) = \int \left(\prod_{k=2}^N dq_k dp_k \right) \rho(q, p)$$

Ci aspettiamo, data la assenza di interazioni che

$$\rho(q, p) = \prod_{i=1}^N \rho_i(q_i, p_i)$$

che mostra proprio

$$\rho_i(q_1, p_1) = \int \left(\prod_{k=2}^N dq_k dp_k \right) \prod_{i=1}^N \rho_i(q_i, p_i) = \rho_i(q_1, p_1) \prod_{k=2}^N \int dq_k dp_k \rho_k(q_k, p_k)$$

dove ciascuno degli integrali e' pari a uno per la normalizzazione della probabilita'.

In caso di variabili continue, se le variabili sono indipendenti

$$P_{XY}(x, y) = P_X(x)P_Y(y)$$

(anche esso generalizzabile banalmente a molte variabili).

La probabilita' condizionata si scrive invece

$$P_{XY}(x|y) = \frac{P_{XY}(x, y)}{P_Y(y)}$$

che si generalizza

$$P(x_1, x_2 \dots x_k | x_{k+1} \dots x_N) = \frac{P(x_1, x_2, \dots x_N)}{P(x_{k+1}, \dots x_N)}$$

- Media

Definiamo la media di X come

$$\langle x \rangle = \int dx x P(x)$$

e per una arbitraria potenza k

$$\langle x^k \rangle = \int dx x^k P(x)$$

Il valore medio di una funzione $f(x)$

$$\langle f(x) \rangle = \int dx f(x) P(x)$$

- Fluttuazioni

Definiamo varianza

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$$

- Somma di variabili random

Se abbiamo una variabili Y che e' somma di variabili aleatorie X_i , per esempio

$$Y = \sum_i a_i X_i$$

Avremo cosi che

$$\langle Y \rangle = \sum_i a_i \langle X_i \rangle$$

e per Y^2

$$\langle Y^2 \rangle = \sum_i \sum_j a_i a_j \langle X_i X_j \rangle$$

e per la varianza

$$\begin{aligned} \sigma_Y^2 &= \sum_i \sum_j a_i a_j \langle X_i X_j \rangle - \sum_i \sum_j a_i a_j \langle X_i \rangle \langle X_j \rangle = \\ &= \sum_i a_i^2 \sigma_{X_i}^2 + 2 \sum_i \sum_{j < i} a_i a_j \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle \end{aligned}$$

(questo ultimo termine prende il nome di correlazione connessa). L' ultimo passaggio si capisce facilmente considerando per esempio il termine $i = 1, j = 2$.

$$\begin{aligned} \langle (X_1 - \langle X_1 \rangle)(X_2 - \langle X_2 \rangle) \rangle &= \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle + \langle X_1 \rangle \langle X_2 \rangle \\ &= \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle \end{aligned}$$

Se le variabili sono indipendenti, per $i \neq j$

$$\langle X_i X_j \rangle = \int dx_i dx_j x_i x_j P(x_i, x_j) = \langle X_i \rangle \langle X_j \rangle$$

e

$$\sigma_Y^2 = \sum_i a_i^2 \sigma_{X_i}^2$$

E' importante ricordare che, date variabili indipendenti, se si studia la loro somma, la media e' la somma delle medie e la varianza e' la somma delle varianze

- Cambio di variabili

Se partiamo da un certo set di variabili aleatorie $X_1 \dots X_N$ e siamo interessati ad M variabili Y_i , funzioni di X_k

$$Y_i = f_i(X_1 \dots X_N)$$

ed assumendo che conosciamo la $P_X(x_1, \dots, x_N)$, possiamo trovare la $P_Y(y_1 \dots y_M)$ come

$$P_Y(y_1, y_2 \dots y_M) = \int dx_1 dx_2 \dots dx_N P_X(x_1, x_2 \dots x_N) \delta(Y_1 - y_1) \delta(Y_2 - y_2) \dots \delta(Y_M - y_M)$$

Per verificare questa formula partiamo da una funzione $g(y_1 \dots y_M)$. Il suo valore medio e'

$$\langle g(y_1 \dots y_M) \rangle = \int dy_1 dy_2 \dots dy_M P(y_1 \dots y_M) g(y_1 \dots y_M)$$

Ma poiche' le y sono funzioni di x io posso anche calcolare la stessa quantita' direttamente utilizzando le variabili x_n .

$$\langle g(y_1 \dots y_M) \rangle = \int dx_1 dx_2 \dots dx_N P(x_1, x_2, \dots x_N) g(f_1(x_1 \dots x_N), f_2(x_1 \dots x_N), \dots, f_M(x_1 \dots x_N))$$

Ricordando che l' integrale di una delta e' uno

$$\int dy_i \delta(y_i - f_i(x_1 \dots x_N)) = 1$$

possiamo sostituirci M di queste relazioni nell' integrale in x_i

$$\langle g(y_1 \dots y_M) \rangle =$$

$$\int dx_1 dx_2 \dots dx_N P(x_1, \dots x_N) g(f_1(x_1, \dots, x_N), f_2(x_1, \dots, x_N), \dots, f_M(x_1, \dots, x_N)) \prod_{i=1}^M \int dy_i \delta(y_i - f_i(x_1, \dots, x_N))$$

e spostando gli integrali su y

$$\langle g(y_1 \dots y_M) \rangle =$$

$$\prod_i^M \int dy_i \int dx_1 dx_2 \dots dx_N P(x_1, \dots x_N) g(f_1(x_1, \dots, x_N), f_2(x_1, \dots, x_N), \dots, f_M(x_1, \dots, x_N)) \delta(y_i - f_i(x_1, \dots, x_N))$$

e sostituendo grazie alle delte y_i con $f_i(x_1, x_2, \dots, x_N)$

$$= \prod_i^M \int dy_i \int dx_1 dx_2 \dots dx_N P(x_1, x_2, \dots, x_N) g(y_1, y_2 \dots y_M) \delta(y_i - f_i(x_1 \dots x_N))$$

Se leggiamo questa espressione come calcolo di un valore medio, dobbiamo concludere che

$$P(y_1, y_2 \dots y_M) = \int dx_1 dx_2 \dots dx_N P(x_1, x_2, \dots x_N) \delta(y_1 - f_1(x_1 \dots x_N)) \dots \delta(y_M - f_M(x_1 \dots x_N))$$

Esempio esplicito cambio di variabile

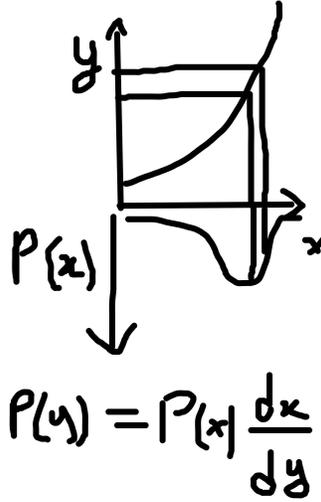


Figure 1: Cartoon utile a capire che occorre modificare la probabilita' del rapporto tra le lunghezze dx e dy

Consideriamo il passaggio da X a $Y = f(X)$ ed assumiamo per semplicita' che $f(X)$ sia una funzione con derivata sempre dello stesso segno, in modo che esista l' inversa $X = f^{-1}(Y)$

Abbiamo visto che

$$P_Y(y) = \int_{-\infty}^{+\infty} dx P_X(x) \delta(y - f(x))$$

Se facciamo un cambio di variabile $z = f(x)$ abbiamo $dz = \frac{df}{dx} dx$ e dunque

$$P_Y(y) = \int_{-\infty}^{+\infty} \frac{dz}{|\frac{df}{dx}|} P_X(f^{-1}(y)) \delta(y - z)$$

Per spiegare il valore assoluto, notate che quando cambiamo variabile di integrazione, dobbiamo anche cambiare gli estremi di integrazione (se $\frac{df}{dx} < 0$). Questo produce un segno meno che viene eliminato sostituendo $\frac{df}{dx}$ con $|\frac{df}{dx}|$.

Se effettuiamo l' integrazione su z sfruttando la delta abbiamo

$$P_Y(y) = \frac{1}{|\frac{df(x)}{dx}|} P_X(f^{-1}(y)) = |\frac{dx}{df(x)}| P_X(x) \quad \text{o equivalentemente} \quad |P_Y(y) dy| = |P_X(x) dx|$$

Questa formula si generalizza al caso di piu' variabili, come

$$P(y_1, y_2, \dots, y_N) = \frac{P(x_1, x_2, \dots, x_n)}{\det J}$$

dove J e' lo Jacobiano della trasformazione.

Un esempio semplice di cambio di variabile (DOPO)

Supponiamo di voler calcolare la probabilita' di osservare una energia E nell' insieme microcanonico. Dalle formule viste troviamo che

$$P(E) = \int dpdq \rho_M(p, q) \delta(H(p, q) - E)$$

Nel caso microcanonico, chiamando \bar{E} l' energia dell 'insieme

$$\rho_M(p, q) = \frac{\delta(H(p, q) - \bar{E})}{\Gamma(\bar{E})}$$

e sostituendo

$$P(E) = \int dpdq \frac{\delta(H(p, q) - \bar{E})}{\Gamma(\bar{E})} \delta(H(p, q) - E) = \int dpdq \frac{\delta(H(p, q) - \bar{E})}{\Gamma(\bar{E})} \delta(\bar{E} - E) = \delta(\bar{E} - E)$$

Come c'era da aspettarsi, se $E = \bar{E}$ la probabilita' e' uno, se $E \neq \bar{E}$ la probabilita' e' nulla.

Un utile esempio di cambio di variabile

Supponete di aver bisogno per la vostra applicazione di una sequenza di numeri y_i distribuiti secondo una distribuzione esponenziale, cioe' con

$$P(y) = \frac{1}{\lambda} e^{-\frac{y}{\lambda}}$$

e che il vostro calcolatore abbia solo la funzione rand() che fornisce numeri random tra 0 e 1. Identificando x_i con i numeri random forniti dal generatore ed utilizzando quanto imparato abbiamo che

$$P(y)dy = P(x)dx \quad \rightarrow \quad \frac{1}{\lambda} e^{-\frac{y}{\lambda}} dy = dx \quad \rightarrow \quad -de^{-\frac{y}{\lambda}} = dx$$

da cui

$$e^{-\frac{y}{\lambda}} = -x \quad \rightarrow \quad y = -\lambda \ln x$$

E' quindi sufficiente estrarre un numero x con $\text{rnd}()$ e poi calcolarne il logaritmo naturale e moltiplicare il risultato per $-\lambda$.

Distribuzioni Notevoli

Distribuzione Binomiale

La distribuzione binomiale descrive una variabile random S_N , somma di N altre variabili uguali, che possono assumere i valori uno o zero, con probabilità p a $q = 1 - p$.

$$S_N = \sum_{i=1}^N t_i$$

S_N puo' dunque assumere tutti i valori interi tra 0 ed N .

Possiamo immediatamente calcolare media e varianza ricordandoci che

$$\langle t_i \rangle = (p * 1 + q * 0) \quad \langle t_i^2 \rangle = N(p * 1^2 + q * 0^2) \quad \sigma_{t_i}^2 = p - p^2 = p(1 - p) = pq$$

da cui

$$\langle S_N \rangle = Np \quad \langle S_N^2 \rangle - \langle S_N \rangle^2 = N\sigma^2 = Npq$$

La probabilità di assumere il valore k e' data da

$$P_{S_N}(k) = \binom{N}{k} p^k q^{N-k}$$

E' facile vedere che questa distribuzione e' normalizzata in modo proprio

$$\sum_{k=0}^N P_{S_N}(k) = \sum_{k=0}^N \binom{N}{k} p^k q^{N-k} = (p + q)^N = 1$$

Il valore medio e'

$$\langle k \rangle = \sum_{k=0}^N k P_{S_N}(k) = \sum_{k=0}^N k \binom{N}{k} p^k q^{N-k} = p \frac{d}{dp} \sum_{k=0}^N \binom{N}{k} p^k q^{N-k} = p \frac{d}{dp} (p+q)^N = pN(p+q)^{N-1} = Np$$

mentre $\langle k^2 \rangle$ e'

$$\langle k^2 \rangle = \sum_{k=0}^N k^2 \binom{N}{k} p^k q^{N-k} = p^2 \frac{d^2}{dp^2} \sum_{k=0}^N \binom{N}{k} p^k q^{N-k} + \langle k \rangle = p^2 \frac{d^2}{dp^2} (p+q)^N + \langle k \rangle = p^2 N(N-1) + Np$$

così che

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 = p^2 N(N-1) + Np - N^2 p^2 = Np(1-p) = Npq$$

Come applicazione della distribuzione binomiale prendiamo un contenitore di volume V diviso in celle di volume ΔV ed inseriamo a caso N atomi. La probabilità che un atomo sia in una specifica cella è $p = \Delta V/V$. La probabilità che ci siano k atomi nella cella è dunque

$$P(k) = \binom{N}{k} p^k q^{N-k}$$

Distribuzione di Poisson

Se adesso supponiamo che $\Delta V/V \ll 1$ (cioè il caso in cui p sia piccola), ci aspettiamo anche che ogni cella contenga pochi atomi. In questo caso la probabilità $P(k)$ sarà significativamente diversa da zero solo per piccoli k . Utilizzando il fatto che $k \ll N$

$$P(k) = \frac{N!}{(N-k)!k!} p^k q^{N-k} = \frac{N(N-1)\dots N(N-k+1)}{k!} p^k q^{N-k} \approx \frac{(Np)^k}{k!} q^N = \frac{N^k p^k}{k!} e^{N \ln(1-p)}$$

ed espandendo $\ln(1-p) \approx -p$ e chiamando $Np = \lambda$,

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

che costituisce la distribuzione di Poisson. Nel caso del gas nel contenitore

$$\lambda = Np = \frac{N\Delta V}{V} = \rho\Delta V.$$

Quindi λ ha il significato di numero medio di particelle nel volume ΔV .

Se calcoliamo valori medi e varianza della distribuzione di Poisson troviamo

$$\langle k \rangle = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

Poiché

$$\lambda \frac{d}{d\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \left(\sum_{k=0}^{\infty} k \frac{\lambda^{k-1}}{k!} e^{-\lambda} - \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \right) = \langle k \rangle - \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

o, utilizzando il fatto che $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$,

$$0 = \langle k \rangle - \lambda \quad \rightarrow \quad \langle k \rangle = \lambda$$

Per calcolare la varianza facciamo lo stesso gioco. Come abbiamo visto prima

$$\frac{dP_k}{d\lambda} = k \frac{\lambda^{k-1}}{k!} e^{-\lambda} - \frac{\lambda^k}{k!} e^{-\lambda}$$

per cui ora

$$\frac{d^2 P_k}{d\lambda^2} = k(k-1) \frac{\lambda^{k-2}}{k!} e^{-\lambda} - k \frac{\lambda^{k-1}}{k!} e^{-\lambda} - k \frac{\lambda^k}{k!} e^{-\lambda} + \frac{\lambda^k}{k!} e^{-\lambda}$$

e moltiplicando per λ^2 e sommando su tutti i k

$$0 = \langle k(k-1) \rangle - 2 \langle k \rangle \lambda + \lambda^2 \quad 0 = \langle k^2 \rangle - \lambda - \lambda^2$$

cioè

$$\langle k^2 \rangle - \langle k \rangle^2 = \lambda$$

Il carattere speciale della distribuzione di Poisson è che la media è uguale alla varianza.

Nota che avremmo potuto anche ottenere questo risultato facendo il limite della media e della varianza della Binomiale per p piccolo

$$\langle k \rangle = Np \quad \sigma^2 = Npq \approx Np$$

Una maniera alternativa di derivare la distribuzione di Poisson

Studiamo un processo in cui, nel tempo, avvengono degli eventi aleatori (entrata di acquirenti in un negozio, emissione di radiazione) con la sola ipotesi che non si realizzino eventi simultanei. Chiamiamo a il numero medio di eventi per unità di tempo. Nell'intervallo Δt quindi si realizza un evento con probabilità $a\Delta t$.

Se volessimo scrivere una equazione per la probabilità di non osservare alcun evento nel tempo t , $P(0, t)$, potremmo scrivere

$$P(0, t) = P(0, t - \Delta t)(1 - a\Delta t)$$

dove abbiamo separato l'intervallo t in due parti ($t - \Delta t$ e Δt) e scritto la probabilita' che non ci siano eventi sia nel primo che nel secondo intervallo. Sviluppando in serie di Taylor $P(0, t - \Delta t)$ troviamo

$$P(0, t - \Delta t) = \left[P(0, t) - \frac{dP(0, t)}{dt} \Delta t \right]$$

per cui

$$P(0, t - \Delta t)(1 - a\Delta t) = P(0, t) - \frac{dP(0, t)}{dt} \Delta t - a\Delta t P(0, t)$$

dove abbiamo trascurato i termini in Δ^2 . Dunque $P(0, t)$ soddisfa

$$\frac{dP(0, t)}{dt} = -aP(0, t)$$

da cui

$$P(0, t) = e^{-at}$$

Nota che la costante di integrazione e' stata scelta a uno per garantire che $P(0, 0) = 1$.

Se ora scrivessi l'equazione per 1 evento nel tempo t avrei

$$P(1, t) = P(1, t - \Delta t)(1 - a\Delta t) + P(0, t - \Delta t)a\Delta t = P(1, t - \Delta t)(1 - a\Delta t) + e^{-a(t-\Delta t)}a\Delta t$$

per cui

$$\frac{P(1, t) - P(1, t - \Delta t)}{\Delta t} = -aP(1, t - \Delta t) + ae^{-a(t-\Delta t)}$$

e nel limite $\Delta t \rightarrow 0$

$$\frac{dP(1, t)}{dt} = -aP(1, t) + ae^{-at}$$

la cui soluzione e', cosi si puo' facilmente verificare

$$P(1, t) = ate^{-at}$$

Se facciamo lo stesso calcolo per un k arbitrario abbiamo

$$\frac{dP(k, t)}{dt} = -aP(k, t) + aP(k - 1, t)$$

cioe' una equazione di ricorrenza. Per risolverla conviene passare ad una forma integrale. Per fare questo, spostiamo a sinistra tutto quello che riguarda $P(k, t)$, moltiplichiamo ambo i membri per e^{at} e poi integriamo

$$e^{at} \left[\frac{dP(k, t)}{dt} + aP(k, t) \right] = ae^{at} P(k - 1, t)$$

$$\frac{de^{at}P(k, t)}{dt} = ae^{at}P(k-1, t)$$

$$e^{at}P(k, t) = a \int_0^t e^{at'} P(k-1, t') dt'$$

da cui per esempio

$$e^{at}P(2, t) = a \int_0^t e^{at'} at' e^{-at'} = a \int_0^t at' = a^2 \frac{t^2}{2}$$

per cui

$$P(2, t) = \frac{(at)^2}{2} e^{-at}$$

E' facile convincersi che

$$P(k, t) = \frac{(at)^k}{k!} e^{-at}$$

Una altra applicazione: Distanza tra stelle

Supponiamo (anche se e' sbagliato) che le stelle siano non-correlate tra loro e descritte da una densita' media ρ . Vogliamo sapere quale e' la probabilita' $P(R)dR$ di osservare la prima stella tra R ed $R + dR$ da un punto prefissato.

Questa probabilita', sulla base di quanto discusso prima, e'

$$P(R)dR = P(0, R)P(1, dR)$$

dove $P(0, R)$ e' la probabilita' di non trovare stelle nel volume $\frac{4}{3}\pi R^3$ e $P(1, dR)$ e' la probabilita' di trovare una stella nel volume $4\pi R^2 dR$ e dunque

$$P(1, R) = \frac{\lambda_0^0}{0!} e^{-\lambda_0} \frac{\lambda_1^1}{1!} e^{-\lambda_1}$$

dove $\lambda_0 = \rho \frac{4\pi R^3}{3}$ mentre $\lambda_1 = 4\pi R^2 dR \rho$.

Dunque, nel limite di dR che tende a zero,

$$P(R)dR = 4\pi R^2 \rho e^{-\rho \frac{4\pi R^3}{3}} dR$$

Distribuzione Gaussiana

Una distribuzione particolarmente importante e' la distribuzione gaussiana. Per una variabile X con media m e varianza σ^2

$$P_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Ne approfittiamo per ricordare l' integrale gaussiano

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

Possiamo facilmente dimostrare che m e' il valore medio e σ^2 e' la varianza, svolgendo gli opportuni integrali. Per calcolare la varianza, si puo' usare il truccetto della derivata. Infatti

$$\int x^2 e^{-\frac{ax^2}{2}} dx = -2 \frac{d}{da} \int e^{-\frac{ax^2}{2}} dx = -2 \frac{d}{da} \sqrt{\frac{2\pi}{a}} = -2 \sqrt{2\pi} \frac{-1}{2} a^{-3/2} = \sqrt{2\pi} a^{-3/2} = \sqrt{\frac{2\pi}{a}} \frac{1}{a}$$

Nel caso in cui $m = 0$, vediamo subito che $a = 1/\sigma^2$

$$\int x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{2\pi\sigma^2} \sigma^2 = \sigma^2$$

Se m fosse diversa da zero, l' integrale sarebbe identico dopo un cambio di variabile $x - m$. Notate che in questo caso dovrete mediare $(x - m)^2$ per avere la varianza.

Per calcolare i momenti di ordine superiore, basta focalizzarsi per simmetria solo su quelli pari.

Per calcolarli, utilizziamo la funzione generatrice dei momenti.

funzione generatrice dei momenti (funzione caratteristica)

$$\tilde{f}(k) = \int e^{ikx} P_X(x) dx = \langle e^{ikx} \rangle$$

e sviluppando

$$e^{ikx} = \sum_{n=0}^{\infty} \frac{(ikx)^n}{n!}$$

da cui

$$\tilde{f}(k) = \int \sum_{n=0}^{\infty} \frac{(ikx)^n}{n!} P_X(x) dx = \sum_{n=0}^{\infty} \langle x^n \rangle \frac{(ik)^n}{n!}$$

Utilizzando le derivate alla origine, troviamo

$$\langle x^n \rangle = (-i)^n \frac{d^n}{dk^n} \tilde{f}(k)|_{k=0}$$

Nel caso della distribuzione gaussiana, la funzione generatrice dei momenti e' (assumendo media nulla per semplicitá)

$$\tilde{f}(k) = \int e^{ikx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx$$

che con la formula precedentemente discussa (identificando b con ik)

$$\tilde{f}(k) = e^{-\frac{k^2\sigma^2}{2}}$$

la cui espansione e'

$$\tilde{f}(k) = \sum_{m=0}^{\infty} \left(\frac{k^2\sigma^2}{2}\right)^m \frac{1}{m!} (-1)^m$$

che mostra che compaiono solo i momenti pari. Confrontando le due sommatorie per i termine generico k^{2m}

$$\langle x^{2m} \rangle \frac{(ik)^{2m}}{(2m)!} = \left(\frac{k^2\sigma^2}{2}\right)^m \frac{1}{m!} (-1)^m$$

troviamo

$$\langle x^{2m} \rangle = \frac{(2m)!}{m!} \frac{\sigma^{2m}}{2^m}$$

0.1 Cumulanti

Una altra quantitá utile e' il cosiddetto cumulante, definito in analogia con i momenti. Anche qui si parte da una funzione generatrice dei cumulanti, $f_c(x)$ definita come il logaritmo naturale della funzione generatrice dei momenti

$$\tilde{f}_c(k) = \ln \langle e^{ixk} \rangle = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle_c$$

La relazione tra i momenti ed i cumulanti si trova sfruttando l' uguaglianza

$$\langle e^{ixk} \rangle = e^{\tilde{f}_c(k)} \quad \langle e^{ixk} \rangle = \tilde{f}(k)$$

$$\sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \langle x^m \rangle = \exp \left[\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle_c \right]$$

e poiche' l' esponenziale di una somma e' il prodotto degli esponenziali

$$\sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \langle x^m \rangle = \prod_{n=1}^{\infty} \exp \left[\frac{(ik)^n}{n!} \langle x^n \rangle_c \right] = \prod_{n=1}^{\infty} \sum_{p_n=0}^{\infty} \frac{1}{p_n!} \left[\frac{(ik)^n}{n!} \langle x^n \rangle_c \right]^{p_n}$$

dove abbiamo espanso l' esponenziale. Troviamo cosi'

$$\sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \langle x^m \rangle = \prod_{n=1}^{\infty} \sum_{p_n=0}^{\infty} \frac{1}{p_n!} (ik)^{np_n} \frac{\langle x^n \rangle_c^{p_n}}{n!^{p_n}}$$

Quindi, se vogliamo confrontare potenze uguali di ik , per trovare la relazione tra momenti e cumulanti, dobbiamo trovare tutti i casi a destra che danno la stessa potenza.

Vediamo cosa succede. Per $n = 1$, $p_n = 0, 1, 2, 3, \dots$ che danno potenze np_n del tipo $0, 1, 2, 3, \dots$. Per $n = 2$, $p_n = 0, 1, 2, 3, \dots$ abbiamo termini np_n del tipo $0, 2, 4, 6, 8, \dots$. Per $n = 3$, $p_n = 0, 1, 2, 3, \dots$ abbiamo termini np_n del tipo $0, 3, 6, 9, 12, \dots$. Poiche' dobbiamo fare il prodotto di tutti gli n , la potenza finale di $(ik)^m$ sara' data dalla somma di tutti i termini tali che $\sum_{p_n=0}^m np_n = m$.

Per $m = 1$, avremo $1 + 0 + 0 + 0$. Per $m = 2$, avremo $2 + 0 + 0 + 0$ e $0 + 2 + 0 + 0$. Per $m = 3$ avremo $3 + 0 + 0 + 0$ e $1 + 2 + 0 + 0$... e $0 + 0 + 3 + 0$... e cosi' via.

Si trova cosi'

$$\begin{aligned} \langle x \rangle &= \langle x \rangle_c & (1) \\ \langle x^2 \rangle &= \langle x^2 \rangle_c + \langle x \rangle_c^2 \\ \langle x^3 \rangle &= \langle x^3 \rangle_c + 3\langle x^2 \rangle_c \langle x \rangle_c + \langle x \rangle_c^3 \end{aligned}$$

Invertendo troviamo

$$\begin{aligned}
 \langle x \rangle_c &= \langle x \rangle & (2) \\
 \langle x^2 \rangle_c &= \langle x^2 \rangle - \langle x \rangle^2 \\
 \langle x^3 \rangle_c &= \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \\
 \langle x^4 \rangle_c &= \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 - 6 \langle x \rangle^4
 \end{aligned}$$

che prendono il nome di media, varianza, skewness, curtosis

Con un po' di algebra si puo' fare vedere che

$$\langle x_c^n \rangle = \langle x^n \rangle - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \langle x^m \rangle \langle x^{n-m} \rangle$$

Quindi, $P(x)$, i momenti, e i cumulanti contengono la stessa quantita' di informazioni.

Una proprieta' utile dei cumulanti e' data dal seguente esempio. Se abbiamo una variabile Z somma di due variabili indipendenti X e Y allora i cumulanti di Z sono la somma dei cumulanti di X e Y .

Infatti

$$P_Z(z) = \int dx dy P(x) P(y) \delta(z - (x + y))$$

La $\tilde{f}(k)$ e'

$$\begin{aligned}
 \tilde{f}_Z(k) &= \int dx dy P(x) P(y) \delta(z - (x + y)) e^{ikz} dz = \int dx dy P(x) P(y) e^{ik(x+y)} = \\
 &= \int dx P(x) e^{ikx} \int dy P(y) e^{iky} = \tilde{f}_X(k) \tilde{f}_Y(k)
 \end{aligned}$$

e naturalmente per i cumulanti

$$\tilde{f}_c^Z(k) = \ln \tilde{f}_X(k) + \ln \tilde{f}_Y(k)$$

Questa formula si generalizza per somme di N variabili aleatorie

$$\begin{aligned}
 Z &= \sum X_i \\
 \tilde{f}_c^Z(k) &= \sum_i \ln \tilde{f}_{X_i}(k)
 \end{aligned}$$

Cumulanti della binomiale (esercizio)

Visto che nel caso della binomiale abbiamo usato i simboli p e q e k e N , chiamiamo t la variabile trasformata. Nel caso della binomiale

$$\tilde{f}(t) = \langle e^{itk} \rangle = \sum_{k=0}^N \frac{N!}{(N-k)!k!} e^{itk} p^k q^{N-k} = \sum_{k=0}^N \frac{N!}{(N-k)!k!} (e^{it}p)^k q^{N-k} = (e^{it}p + q)^N$$

e, quindi

$$\tilde{f}_c(t) = \ln \langle e^{itk} \rangle = N \ln(e^{it}p + q)$$

Se espandiamo $\tilde{f}_c(t)$ intorno a $t = 0$ abbiamo (trascurando la N per semplicità')

$$\begin{aligned} \tilde{f}_c(0) &= \ln(p + q) = 0 \\ \frac{d\tilde{f}_c}{dt} &= \frac{ie^{it}p}{(e^{it}p + q)} \quad \left. \frac{d\tilde{f}_c}{dt} \right|_{t=0} = \frac{ip}{(p + q)} = ip \\ \frac{d^2\tilde{f}_c}{dt^2} &= \frac{-e^{it}p(e^{it}p + q) - ie^{it}pie^{it}p}{(e^{2it}p + q)^2} \quad \left. \frac{d^2\tilde{f}_c}{dt^2} \right|_{t=0} = \frac{-p(p + q) + pp}{(p + q)^2} = -pq \end{aligned} \quad (3)$$

per cui

$$\tilde{f}_c(t) = ipt - \frac{1}{2}pqt^2 + \dots$$

Confrontando con

$$\tilde{f}_c(t) = \ln \langle e^{ikt} \rangle = \sum_{n=1}^{\infty} \frac{(it)^n}{n!} \langle k^n \rangle_c$$

troviamo (rimettendo la N)

$$\langle k \rangle_c = Np \quad \langle k^2 \rangle_c = \sigma^2 = Npq$$

Cumulanti della gaussiana

$$\tilde{f}_X(k) = \int dx e^{ikx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} =$$

definendo $y = x - m$

$$\begin{aligned} &= \int dy e^{ik(y+m)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} = e^{ikm} \int dy e^{iky} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \\ &= e^{ikm} e^{-k^2\sigma^2/2} \end{aligned}$$

per cui la funzione generatrice dei cumulanti e'

$$\tilde{f}_c(k) = ikm - \frac{k^2\sigma^2}{2}$$

Se calcoliamo le derivate di $\tilde{f}_c(k)$ valutate in $k = 0$ troviamo che le uniche derivate non nulle sono la prima e la seconda

$$\left. \frac{d\tilde{f}_c(k)}{dk} \right|_{k=0} = im \quad \left. \frac{d^2\tilde{f}_c(k)}{dk^2} \right|_{k=0} = -\sigma^2$$

per cui

$$\langle x_c \rangle = m \quad \langle x_c^2 \rangle = \sigma^2$$

che mostra, che i cumulanti con $n > 2$ sono nulli.

Distribuzione gaussiana Multivariata

Per N variabili indipendenti x_i si generalizza come

$$P_X(x_1, \dots, x_N) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^N e^{-\frac{\sum_{i=1}^N (x_i - m)^2}{2\sigma^2}}$$

1 Teoremi alla base della legge dei grandi numeri: variabili intensive hanno fluttuazioni trascurabili

Per variabili a media μ e varianza σ^2 finita Chebyshev ci insegna che stare nelle code della distribuzione e' poco probabile. Infatti, ora mostreremo che

$$P(|x - \mu| > \gamma) \leq \frac{\sigma^2}{\gamma^2}$$

Vediamo come dimostrarlo:

$$P(|x - \mu| > \gamma) = \int_{\mu+\gamma}^{\infty} P(x)dx + \int_{-\infty}^{\mu-\gamma} P(x)dx$$

se notiamo che

$$\frac{(x - \mu)}{\gamma} > 1 \quad \rightarrow \quad \left(\frac{(x - \mu)}{\gamma}\right)^2 > 1$$

possiamo inserire nell' integrale queste quantita' sempre maggiori di uno e cosi' maggiorare l' espressione. Abbiamo cosi'

$$\int_{\mu+\gamma}^{\infty} P(x) \left(\frac{(x - \mu)}{\gamma}\right)^2 dx + \int_{-\infty}^{\mu-\gamma} P(x) \left(\frac{(x - \mu)}{\gamma}\right)^2 dx > P(|x - \mu| > \gamma)$$

Se aggiungiamo anche la parte mancante dell' integrale non possiamo che migliorare la disuguaglianza

$$\int_{-\infty}^{\infty} P(x) \left(\frac{(x - \mu)}{\gamma}\right)^2 dx > P(|x - \mu| > \gamma)$$

o

$$P(|x - \mu| > \gamma) < \frac{\sigma^2}{\gamma^2}$$

Un teorema analogo vale anche per i momenti k -esimi delle distribuzioni, come originariamente mostrato da Markov

$$P(|x - \mu| > \gamma) \leq \frac{|x - \mu|^k}{\gamma^k}$$

Ora torniamo ad una variabile che e' somma di variabili indipendenti con media e varianza finita

$$Z_N = \frac{1}{N} \sum_i X_i$$

con

$$\langle Z_N \rangle = \frac{1}{N} \sum_i \langle X_i \rangle = \frac{1}{N} \sum_i \mu_i \quad \sigma_Z^2 = \frac{1}{N^2} \sum_i \sigma_i^2$$

Se applichiamo Chebyshev,

$$P\left(\left|z - \frac{\sum_i \mu_i}{N}\right| > \gamma\right) < \frac{\sum_i \sigma_i^2}{N^2 \gamma^2}$$

Ne consegue che per $N \rightarrow \infty$,

$$\lim_{N \rightarrow \infty} P\left(\left|z - \frac{\sum_i \mu_i}{N}\right| > \gamma\right) < 0$$

Questo ci dice che nel limite $N \rightarrow \infty$ (limite termodinamico) le variabili intensive (come Z) di fatto hanno fluttuazioni nulle. Notate che Z per esempio e' l' energia potenziale del sistema per particella. Questo anche spiega perche' i valori medi delle variabili intensive non dipendono dall' ensemble scelto per calcolarle.

Grandi numeri

Guardiamo adesso ad una variabile aleatoria Y che é somma di N variabili aleatorie X_i indipendenti con valore medio nullo e varianza finita. Per comodità assumiamo che tutte le variabili X_i siano descritte dalla stessa $P_X(x_i)$

Per fare un po' di esperienza, guardiamo il caso di due variabili

$$Y = X_1 + X_2$$

$$P_Y(y) = \int dx_1 dx_2 P(x_1)P(x_2) \delta(y - x_1 - x_2)$$

e integrando su x_1

$$P_Y(y) = \int P_{X_1}(y - x_2)P_{X_2}(x_2)dx_2$$

quindi una convoluzione. Se trasformiamo (di fatto secondo Fourier la y)

$$\int dy P_Y(y) e^{izy} = \int dy \int e^{izy} P_{X_1}(y-x_2)P_{X_2}(x_2)dx_2 = \int dy \int e^{iz(y-x_2)} e^{izx_2} P_{X_1}(y-x_2)P_{X_2}(x_2)dx_2$$

$$= \int dy e^{iz(y-x_2)} P_{X_1}(y-x_2) \int dx_2 e^{izx_2} P_{X_2}(x_2) = \int d(y-x_2) e^{iz(y-x_2)} P_{X_1}(y-x_2) \int dx_2 e^{izx_2} P_{X_2}(x_2)$$

abbiamo

$$\tilde{P}_Y(z) = \tilde{P}_{X_1}(z) \tilde{P}_{X_2}(z)$$

Notiamo che $\tilde{P}_Y(z)$ altro non e' che il valore medio di e^{izy} . A parte costanti di normalizzazione, $\int dy P_Y(y) e^{izy} = \langle e^{izy} \rangle$ ricordiamo prende il nome di funzione caratteristica, o funzione generatrice dei momenti.

Siamo adesso pronti per il teorema del limite centrale che dimostra che la somma di un grande numero di variabili **indipendenti** e con varianza finita e' sempre descritta da una distribuzione gaussiana.

Per comodita' assumiamo che le variabili X_i sono tutte identiche ed indipendenti e con varianza σ^2 e media $\langle X \rangle$. Per mostrare il teorema partiamo dal definire Y e T come

$$Y = \sum_{i=1}^N X'_i$$

e

$$T = \frac{1}{\sigma\sqrt{N}} Y = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N X'_i$$

dove $X'_i = X_i - \langle X \rangle$.

In termini di funzione caratteristica

$$\tilde{P}_Y(z) = \prod_{i=1}^N \tilde{P}_{X'_i}(z) = P_{X'}(z)^N$$

Con un cambio di variabile del tipo $Q = aX$,

$$P_Q(q) = P_X(x) \frac{dx}{dq} = \frac{1}{a} P_X(x)$$

per cui in trasformata

$$\tilde{P}_Q(z) = \int P_Q(q) dq e^{iqz} = \int a dx \frac{P_X(x)}{a} e^{iaxz} = \tilde{P}_X(az)$$

ne consegue che per T (con $a = \frac{1}{\sigma\sqrt{N}}$)

$$\tilde{P}_T(z) = \tilde{P}_Y\left(\frac{1}{\sigma\sqrt{N}}z\right) = \left[\tilde{P}_{X'}\left(\frac{1}{\sigma\sqrt{N}}z\right)\right]^N$$

Sfruttando il fatto che le X' sono a media nulla

$$\tilde{P}_{X'}(z) = 1 + iz \langle X' \rangle - \frac{1}{2}z^2 \langle X'^2 \rangle + \dots = 1 - \frac{1}{2}z^2\sigma^2 + O(z^3)$$

e corrispondentemente per Z

$$\tilde{P}_T(z) = \left[1 - \frac{1}{2}\left(\frac{1}{\sigma\sqrt{N}}\right)^2 z^2\sigma^2 + O\left(\left(\frac{1}{\sigma\sqrt{N}}\right)^3 z^3\right)\right]^N = \left[1 - \frac{z^2}{2N} + O\left(\frac{z^3}{N^{3/2}}\right)\right]^N$$

Nel limite di $N \rightarrow \infty$

$$\tilde{P}_T(z) = e^{-\frac{z^2}{2}}$$

la cui anti-trasformata da

$$P_T(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

e adesso se passiamo da T a Y (ricordando che $T = \frac{1}{\sigma\sqrt{N}}Y$)

$$P_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2N}}e^{-\frac{y^2}{2\sigma^2N}}$$

Se ora volessimo includere anche la media (cioé non usare X' bensí X) avremmo

$$Y = \sum_i X'_i = \sum_i (X_i - \langle X_i \rangle) = \sum_i X_i - N \langle X \rangle$$

e

$$Y' = \sum_i X_i \quad Y = Y' - N \langle X \rangle$$

$$P_Y(y') = \frac{1}{\sqrt{2\pi\sigma^2N}}e^{-\frac{(y' - N\langle X \rangle)^2}{2\sigma^2N}}$$

E' importante ricordare che il teorema non richiede che le varianze delle X_i siano tutte uguali, ma richiede comunque che esse siano finite. In aggiunta, la distribuzione gaussiana e' una approssimazione ottima per descrivere il comportamento vicino alla media, ma naturalmente non puo' catturare in modo proprio gli eventi estremi, avendo trascurato i termini in $O(z^3/N^{3/2})$.

E' interessante notare cosa succede quando $Y = \sum X_i$ e le X_i sono tutte indipendenti e distribuite in modo uguale ma, questa volta, con una varianza infinita. Per esempio nel caso della distribuzione di Cauchy

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

La funzione caratteristica é nota analiticamente

$$\tilde{P}_X(z) = e^{-|z|}$$

e dunque per Y

$$\tilde{P}_Y(z) = e^{-|Nz|}$$

E tornando indietro, per la variabile Y/N abbiamo esattamente la stessa distribuzione del singolo X .

2 Un trucco utile: Saddle-point approximation (Metodo di Laplace)

Consideriamo l' integrale

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

con N grande e $f(x)$ una funzione con un massimo quadratico in x_0 (con $x_0 \in [a, b]$). Se sviluppiamo $f(x)$ intorno a x_0

$$f(x) = f(x_0) + \frac{1}{2} \frac{d^2 f}{dx^2} \Big|_{x_0} (x - x_0)^2$$

(con $\frac{d^2 f}{dx^2} \Big|_{x_0} < 0$) da cui

$$I = e^{Nf(x_0)} \int_a^b e^{N \frac{1}{2} \left| \frac{d^2 f}{dx^2} \Big|_{x_0} (x-x_0)^2} dx$$

Poiche' N e' grande, l' integrando dara' contributi significativi solo intorno a x_0 per cui possiamo estendere i limiti e integrare

$$I = e^{Nf(x_0)} \int_{-\infty}^{\infty} e^{N \frac{1}{2} \left| \frac{d^2 f}{dx^2} \Big|_{x_0} (x-x_0)^2} dx = e^{Nf(x_0)} \sqrt{\frac{2\pi}{N \left| \frac{d^2 f}{dx^2} \Big|_{x_0} \right.}}$$

3 Un trucco utile: Stirling

Iniziamo col mostrare una rappresentazione integrale di $N!$. Partiamo dalla uguaglianza

$$\int_0^{\infty} e^{-at} dt = \frac{1}{a}$$

se deriviamo rispetto ad a N volte abbiamo

$$\int_0^{\infty} (-t)^N e^{-at} dt = (-1)^N N! a^{-(N+1)} \quad \rightarrow \quad a \int_0^{\infty} (at)^N e^{-at} dt = N!$$

una espressione valida per ogni a , incluso $a = 1$

$$\int_0^{\infty} t^N e^{-t} dt = N!$$

Per inciso, questa espressione si puo' generalizzare per N reale definendo una funzione Γ come

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

Riprendiamo dall' espressione trovata

$$N! = \int_0^{\infty} t^N e^{-t} dt = \int_0^{\infty} e^{N \ln t - t} dt =$$

e cambiando variabile $z = t/N$,

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

Utilizzando il punto sella, troviamo che l' esponenziale ha un estremo in

$$\frac{d(\ln z - z)}{dz} = \frac{1}{z} - 1 = 0 \quad \rightarrow \quad z = 1$$

e l' espansione di Taylor quadratica intorno a $z = 1$ da'

$$\ln z - z = -1 + 0z - \frac{1}{2} \frac{1}{z^2} \Big|_{z=1} (z-1)^2 = -1 - \frac{(z-1)^2}{2}$$

per cui (trasformando il limite inferiore di integrazione da -1 a $-\infty$)

$$N N^N \int_0^{\infty} e^{N(\ln z - z)} dz = N N^N \int_0^{\infty} e^{-N - N(z-1)^2/2} dz = e^{-N} N^{N+1} \int_{-\infty}^{\infty} e^{-N(z-1)^2/2} d(z-1) =$$

e troviamo così

$$N! = e^{-N} N^{N+1} \frac{\sqrt{2\pi}}{\sqrt{N}} = e^{-N} N^N \sqrt{2\pi N}$$

Questa formula e' spesso ricordata come

$$\ln N! = N \ln N - N + \ln \sqrt{2\pi N}$$

e per N grandi l' ultimo termine puó essere trascurato.

Una formula ancora piú accurata e'

$$N! = e^{-N} N^N \sqrt{2\pi N} \left[1 + \frac{1}{12N} + \frac{1}{144N^2} \right]$$

3.1 Una espressione integrale per la delta

Se ci ricordiamo che per le trasformate di Fourier – FIX I 2 PI –

$$\tilde{\delta}(z) = \frac{1}{2\pi} \int e^{izx} \delta(x) dx$$

e che

$$\delta(x) = \int e^{-izx} \tilde{\delta}(z) dz$$

possiamo concludere che, risolvendo il primo dei due integrali

$$\tilde{\delta}(z) = 1$$

e dunque

$$\delta(x) = \int e^{-izx} dz$$

Sa la delta non e' centrata sull' origine avremmo

$$\delta(x - x_0) = \int e^{-iz(x-x_0)} dz$$

3.2 Legge dei grandi numeri con i cumulanti

Abbiamo visto che il cumulante della somma e' la somma dei cumulanti Per cui, se $Y = 1/\sqrt{N} \sum x_i$,

$$\langle y^n \rangle_c = N^{-n/2} \sum_i \langle x_i^n \rangle_c$$

Le ipotesi che i cumulanti di Y con $n > 2$ tende a zero con N e' equivalente al teorema del limite centrale.

3.3 Random Walker come applicazione della legge dei grandi numeri

Se chiamiamo x_i una variabile random che puo' avere solo i valori $+l_0$ e $-l_0$ e che descrive la direzione del passo al tempo i -esimo. Se chiamiamo con Y la posizione totale dopo N passi

$$Y = \sum_i^N x_i$$

Il valore medio di x_i e' sempre zero. La varianza di x_i e' l_0^2 (essendo $(1/2 * l_0^2 + 1/2 * (-l_0)^2)$). Possiamo subito scrivere la $P(Y)$ con il teorema del limite centrale

$$P_N(y) = \frac{1}{\sqrt{2\pi l_0^2 N}} e^{-\frac{y^2}{2l_0^2 N}}$$

Se ora assumiamo che un passo corrisponda ad un tempo τ , dopo un tempo t , $N = t/\tau$ possiamo dire

$$P_N(y, t) = \frac{1}{\sqrt{2\pi t/\tau}} e^{-\frac{y^2}{2l_0^2 t/\tau}}$$

e banalmente, definendo un coefficiente di diffusione D

$$\langle y^2(t) \rangle = l_0^2 t/\tau = 2Dt$$