Ripasso di alcuni concetti di probabilita' che ci serviranno

September 26, 2024

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

- 1. Le probabilitá 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 puó dare come risultato 19, 30, 48 occorrenze della faccia contrassegnata con 1, il rapporto $0.19, 0.15, 0.16$ fornisce una stima sempre piú affidabile della probabilitá.

- 2. Le probabilitá 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 probabilitá nella meccanica statistica sono basate su criteri soggettivi. Le conseguenze di tali assegnazioni soggettive di probabilitá devono essere verificate in base alle misurazioni e possono dover essere modificate man mano che si rendono disponibili ulteriori informazioni sul risultato.

Variabili continue

Nel caso, particolarmente importante, che l'evento elementare ω sia un numero reale allora l'insieme dei possibile eventi è la retta numerica reale R. E' comodo introdurre la funzione di distribuzione: $P([-\infty, x))$ che indica la probabilità che l' evento che si realizza abbia valore minore di x, e la densità di probabilità

$$p_X(x) = \frac{dP(x)}{dx}$$

$p_X(x)dx$ 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 $P(x)$ 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.

- Media

  Definiamo la media di X come

  $$< x >= \int dxxP(x)$$

  e per una arbitraria potenza k

  $$< x^k >= \int dxx^kP(x)$$

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

  $$< f(x) >= \int dxf(x)P(x)$$

- Fluttuazioni

  Definiamo varianza

  $$\sigma_x^2 =< x^2 > - < x >^2$$

- 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

  $$< Y >= \sum_i a_i < X_i >$$

e per Y^2

$$< Y^2 >= \sum_i \sum_j a_i a_j < X_i X_j >$$

e per la varianza

$$\sigma_Y^2 = \sum_i \sum_j a_i a_j < X_i X_j > - \sum_i \sum_j a_i a_j < X_i >< X_j >=$$

$$\sum_i a_i^2 \sigma_{X_i}^2 + 2 \sum_i \sum_{j<i} a_i a_j < (X_i - < X_i >)(X_j - < X_j >) >$$

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

$$< (X_1 - < X_1 >)(X_2 - < X_2 >) >=< X_1 X_2 > - < X_1 >< X_2 > - < X_1 >< X_2 > + < X_1 >< X_2 >$$

$$=< X_1 X_2 > - < X_1 >< X_2 >$$

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

$$< X_i X_j >= \int dx_i dx_j x_i x_j P(x_i, x_j) =< X_i >< X_j >$$

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

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 di un sistema di N particelle non interagenti, in cui $X = X(q^{3N}, p^{3N})$, la probabilita' che X assuma il valore x e' scrivibile come

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

[dove $\rho(q^{3N}, p^{3N})$ e' la probabilita' del caso q^{3N}, p^{3N}] Nella equazione precedente, praticamente sommiamo su tutti gli eventi per i quali $X(q^{3N}, p^{3N}) = x$. Se avessimo da calcolare la probabilita' di due quantita'

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

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

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

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

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

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)

$$\rho_i(q_i,p_i) = \int (\prod_{k=2}^{N} dq_k dp_k)\rho(q^{3N},p^{3N})$$

Ci aspettiamo, data la assenza di interazioni che

$$\rho(q^{3N},p^{3N}) = \prod_{i=1}^{N}\rho_i(q_i,p_i)$$

che mostra proprio

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

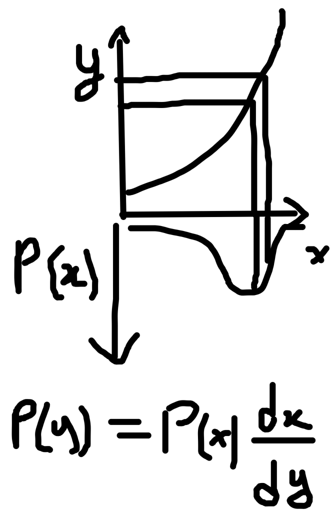
dove ciascuno degli integrali e' pari a uno per la normalizzazione della probabilitá.

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).

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 semplicitá 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

$$P(y_1, y_2, ..., y_N) = \frac{P(x_1, x_2, ..., x_n)}{detJ}$$

dove J e' lo Jacobiano della trasformazione.

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 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 probabilitá 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

$$< t_i >= (p * 1 + q * 0) \qquad < t_i^2 >= N(p * 1^2 + q * 0^2) \qquad \sigma_{t_i}^2 = p - p^2 = p(1 - p) = pq$$

da cui

$$< S_N >= Np \qquad\qquad < S_N^2 > - < S_N >^2 = N\sigma^2 = Npq$$

La probabilitá 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'

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

mentre $< k^2 >$ e'

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

cosi che'

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

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

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

Distribuzione di Poisson

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

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

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

$$< k >= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

Poiche'

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

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

$$0 =< k > -\lambda \quad \rightarrow \quad < k >= \lambda$$

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

$$0 =< k(k-1) > -2 < k > \lambda + \lambda^2 \qquad 0 =< k^2 > -\lambda - \lambda^2$$

cioe'

$$< k^2 > - < k >^2 = \lambda$$

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

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

$$< k >= Np \quad \sigma^2 = Npq \approx Np$$

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 unitá di tempo. Nell' intervallo Δt quindi si realizza un evento con probabilita' $a\Delta t$.

Se volessimo scrivere una equazione per la probabilita' 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 $\Delta t \to 0$

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

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 densitá media ρ. Vogliamo sapere quale é la probabilitá $P(R)dR$ di osservare la prima stella tra R ed $R + dR$ da un punto prefissato.

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

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

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

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

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 trucchetto 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 dovreste mediare $(x-m)^2$ per avere la varianza.

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

Per variabili a media μ e varianza σ^2 finita, la probabilita' che si realizzi un evento nella coda della distribuzione e' piccola. 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

$$< Z_N >= \frac{1}{N} \sum_i < X_i >= \frac{1}{N} \sum_i \mu_i \qquad \sigma_Z^2 = \frac{1}{N^2} \sum_i \sigma_i^2$$

Se applichiamo il teorema precedente (di 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 $N \to \infty$,

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

Questo ci dice che nel limite $N \to \infty$ (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 comoditá 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, $\int dy P_Y(y) e^{izy} =< e^{izy} >$ 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 $< X >$. 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 $X_i' = X_i - < X >$.

In termini di funzione caratteristica

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

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

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

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

$$\tilde{P}_{X'}(z) = 1 + iz < X' > -\frac{1}{2} z^2 < X'^2 > +.... = 1 - \frac{1}{2} z^2 \sigma^2 + O(z^3)$$

e corrispondentemente per Z

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

Nel limite di $N \to \infty$

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

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 $T = \frac{1}{\sigma\sqrt{N}} Y$)

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

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

$$Y = \sum_i X'_i = \sum_i (X_i - < X_i >) = \sum_i X_i - N < X >$$

e

$$Y' = \sum_i X_i \qquad Y = Y' - N < X >$$

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

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

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

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

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

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

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

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

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

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 $z = t/N$,

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

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'

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

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

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

e troviamo cosí

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

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

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