

# Note di Meccanica Statistica II

Marco Cosentino Lagomarsino

1 aprile 2020

## **Premessa. “Regole” di Feynman per apprendere nuovi concetti**

Dato che uno degli scopi principali di questo corso è avviare verso lo studio indipendente, rimastichiamo nel seguito alcuni precetti base dell'apprendimento di una materia, secondo il fisico teorico Richard Feynman (1918-88), premio Nobel (1965) dal carattere originale e attivo in vari campi della fisica.

1) *Scegli l'argomento che vuoi capire e documentati con tutte le fonti a te disponibili.* Annota insieme quello che impari su quell'argomento dalle diverse fonti (libri, internet, lezioni, discussioni). Cerca di integrare le informazioni annotate da diverse fonti, verificando se le trovi coerenti e segnandoti dove trovi problemi.

2) *Pensa di insegnarlo a un bambino.* Dopo aver generato le tue note su un argomento, prova a ripeterlo con l'intento di spiegarlo, come ti trovassi alla lavagna davanti a un gruppo di persone che non sa niente in merito, per esempio dei bambini. Ripeti il concetto usando parole possibilmente semplici, comprensibili a un bambino di 8 anni. Cerca di sottolineare i passaggi logici e gli aspetti importanti e motivanti. Emergeranno i punti e passaggi ancora oscuri, che richiedono un ulteriore sforzo di comprensione.

3) *Torna sui libri quando sei bloccato.* La tecnica del punto 2) permette di tornare sui passaggi logici che non si sono capiti, cercando di superarli. È utile farsi domande e formulare spiegazioni tentative, e cercare dei test per queste spiegazioni. L'obiettivo è essere in grado di spiegare l'argomento da zero, con semplicità, rigore e chiarezza, supportando ogni passaggio del discorso con un'argomentazione.

4) *Organizza e semplifica, anche usando analogie.* I punti 2 e 3 permettono di generare una spiegazione articolata di ogni punto del discorso. È possibile ancora tornare sull'intero discorso per rendere scorrevoli i passaggi che risultano più difficili da digerire. È possibile fornire un'analogia semplice che permette di catturare l'essenza del problema? Perché l'ipotetico pubblico dovrebbe trovare l'argomento interessante? Usa la visione di insieme acquisita per rendere il discorso comprensibile, interessante e scorrevole.

Si noti come questi consigli (anche se è sempre molto utile avere un punto di partenza) scorraggino lo studio basato puramente su materiale predigerito da altri, quali le note che seguono.

# 1 Parte I. Probabilità e variabili aleatorie

Cosa si intenda con "probabilità" è storicamente soggetto di dibattito. Esistono due interpretazioni predominanti, le quali non hanno solo valore filosofico ma si riflettono anche in un diverso utilizzo degli strumenti forniti dalla teorica matematica.

1. L'interpretazione "oggettiva" (o frequentista), definisce la probabilità come frequenza relativa di un evento "favorevole" (cioè un evento di interesse) rispetto a tutti gli eventi possibili. Questa interpretazione è quella più naturale in ambito fisico, poiché essenzialmente empirista.
2. L'interpretazione "soggettiva" (o Bayesiana), associa la probabilità a quanto l'osservatore crede che un evento sia possibile, date le sue conoscenze riguardo ad esso. Questo tipo di approccio comune nei problemi di inferenza.

A prescindere dalle scelte interpretative, esiste una formalizzazione astratta della teoria della probabilità, introdotta da Kolmogorov, che utilizza la teoria della misura. L'approccio di Kolmogorov definisce una terna  $(\Omega, \mathcal{F}, P)$ , dove  $\Omega$  è lo spazio degli stati possibili del sistema in esame,  $\mathcal{F}$  è una  $\sigma$ -algebra sullo spazio degli stati, detta spazio degli eventi possibili, e  $P : \mathcal{F} \rightarrow [0, 1]$  è una misura finita sullo spazio degli eventi, che viene chiamata probabilità. Per esempio consideriamo la roulette, gli stati sono i singoli numeri, gli eventi cose tipo pari, nero, 33, e la probabilità associa pesi agli eventi.

La probabilità deve soddisfare due importanti proprietà. Primo, le probabilità di unioni di eventi disgiunti si sommano,

$$A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B).$$

Questa proprietà vale per unioni numerabili, data la proprietà di sigma-additività della misura. Secondo (la proprietà di "normalizzazione" caratterizzante di una misura di probabilità) la probabilità dell'evento corrispondente all'intero spazio degli stati deve essere uno,

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

I casi più comuni di spazi degli eventi sono insiemi numerabili, per esempio  $\Omega = \mathbb{N}$  o  $\mathbb{Z}$ , oppure i numeri reali (o un intervallo di  $\mathbb{R}$ , e in questo caso  $\mathcal{F}$  è la sigma-algebra di Borel (cioè quella generata dagli insiemi aperti o equivalentemente da quelli chiusi).

## Variabili aleatorie

Una variabile aleatoria (random variable, RV) si può vedere intuitivamente come una variabile i cui valori dipendono dai risultati di un fenomeno stocastico, e che quindi può prendere valori nello spazio degli eventi. Questi valori sono possibili risultati di un esperimento futuro, di un esperimento passato il cui valore è incerto a causa di incertezza nella misura o intrinseca, e possono anche rappresentare concettualmente o i risultati di un processo casuale come il tiro di un dado o la roulette. La definizione formale di una RV però è diversa. Una variabile aleatoria è definita come una funzione  $X : \Omega \rightarrow \mathbb{R}$ , dove lo spazio di arrivo (spazio delle "osservazioni", che in generale può essere un qualsiasi spazio misurabile)  $\mathbb{R}$  è dotato della  $\sigma$ -algebra di Borel, la quale è misurabile, ovvero la controimmagine di  $\mathcal{B}(\mathbb{R})$  è misurabile in  $\Omega$  ( $\forall E \in \mathcal{B}(\mathbb{R}), X^{-1}(E) \in \mathcal{F}$ ).

La variabile aleatoria induce a questo punto in modo naturale una misura "push-forward" nello spazio delle osservazioni

$$P(x \in E) = P(\omega \in \Omega; X(\omega) \in E)$$

Questa è una definizione utile, poiché unifica variabili aleatorie discrete e continue. E porta spazi degli stati generici (fatti per es. di capre, cavoli, capperi, acciughe, orbite di pianeti, etc.) agli insiemi propri dell'analisi matematica. Più in dettaglio, una variabile aleatoria così definita si dice "discreta" se per la sua misura  $\exists K$  tale che  $P(x \in K) = 1$ , con  $K$  sottoinsieme numerabile di  $\mathbb{R}$ , mentre si dice "continua" quando  $P(x = x_0) = 0 \forall x_0 \in \mathbb{R}$ . Inoltre, è possibile avere delle variabili aleatorie "miste", che non soddisfano nessuna di queste due proprietà.

Nonostante questa definizione sia a volte poco intuitiva, ha il pregio di aiutare a definire delle proprietà di convergenza, perchè eredita i criteri di convergenza delle funzioni misurabili. Per esempio (tra le varie possibili) sono interessanti le convergenze che seguono.

- Una convergenza forte, detta "quasi certamente" (che corrisponde alla convergenza "quasi ovunque" dell'analisi matematica):

$$P(X_n \rightarrow X) \xrightarrow{n \rightarrow \infty} 1$$

- Una convergenza debole, detta "in probabilità" (che corrisponde alla convergenza "in misura"):

$$\lim_{n \rightarrow \infty} P(\{|x_n - x| \leq \epsilon\}) = 1$$

Nelle applicazioni pratiche è spesso conveniente utilizzare il concetto intuitivo di variabile aleatoria, ma è molto importante avere presente la definizione formale

per capire aspetti teorici rilevanti, come i teoremi limite, che giocano un ruolo fondante nella meccanica statistica.

## Distribuzioni di probabilità

La distribuzione di probabilità è lo strumento per caratterizzare una random variabile, e dal quale si possono calcolare tutte le quantità interessanti (valori di aspettazione, varianze etc.) relative a una singola RV (se si ha più di una RV le singole distribuzioni non sono tutto ed è necessario caratterizzare le realizzazioni congiunte). Come abbiamo visto, definita una terna  $(\Omega, \mathcal{F}, P)$ , una variabile aleatoria  $X : \Omega \rightarrow \mathbb{R}$  induce una misura sull'insieme di arrivo di tipo push-forward, che possiamo caratterizzare con una funzione nello spazio delle osservazioni.

- Se la variabile aleatoria  $X$  è discreta allora  $P(X = x) = P_x$ , i.e. la funzione che associa agli eventi discreti la loro probabilità, viene definita distribuzione di probabilità (PDF) o anche funzione massa di probabilità (PMF).
- Se la variabile aleatoria  $X$  è continua, allora la PDF viene definita come "densità di probabilità", ossia la funzione  $f : \mathbb{R} \rightarrow \mathbb{R}$  non negativa tale che  $P(x \in A) = \int_A f d\mu = \int_A f(x) dx$ , dove  $\mu$  indica la misura di Lebesgue e l'ultima uguaglianza vale nel caso in cui l'insieme di arrivo della variabile aleatoria sia  $\mathbb{R}$ .

Mentre l'affermazione fatta per variabili discrete non ha bisogno di ulteriori spiegazioni, quella per variabili continue solleva diversi interrogativi su come definire la funzione  $f$ . Lo scopo è esprimere la misura che definisce la variabile aleatoria nei termini della misura caratteristica dello spazio delle osservazioni, la misura di Lebesgue che compare nell'integrale (sulla quale è possibile usare tutti gli strumenti dell'analisi matematica). Quest'ultima però non è normalizzata. D'altra parte il fatto che sia sigma-finita garantisce che l'operazione di "mappa" sia possibile. Questo è stabilito dal seguente teorema di Radon-Nikodym. Dato uno spazio misurabile  $\Omega$  e presa una misura  $\mu$  sigma-finita e tale che  $P$  sia assolutamente continua rispetto a  $\mu$  (finita) su questo spazio, ossia tale che se  $\mu(A) = 0$  allora  $P(A) = 0 \forall A \in \Omega$ , si ha

$$P(A) = \int_A f d\mu \quad (1)$$

con  $f = \frac{\partial P}{\partial \mu}$  detta "derivata di Radon" di  $P$  rispetto a  $\mu$  (ma si noti che non ha le proprietà di una derivata, se non assomigliare allo "jacobiano" che si usa nelle trasformazioni di variabili in un integrale).

Il teorema di Radon-Nikodym si estende anche al caso in cui  $P$  non sia assolutamente continua (i.e. esistono delle masse discrete), e in questo caso le PDF

possono essere distribuzioni nel senso di Schwartz, e quindi giustifica per esempio delle PDF che descrivono variabili aleatorie miste e contengono parti continue e parti singolari, per esempio masse di probabilità fatte di funzioni delta di Dirac, o misure con (Teorema di decomposizione di Lebesgue-Radon-Nikodym).

A partire dalla distribuzione di probabilità si può anche definire la distribuzione di probabilità cumulativa (CDF) come la funzione  $F(x) := P(X < x) = \int_{-\infty}^x f(x')dx'$ , dove il membro destro diventa una somma nel caso discreto, e può essere una combinazione di integrali e somme nel caso misto.

### Valore di aspettazione e momenti della distribuzione

Le distribuzioni di probabilità sono utilizzate per ottenere informazioni sullo spazio degli eventi e sulle variabili aleatorie.

Presa ad esempio una funzione della variabile aleatoria  $\phi(X)$ , definiamo “valore di aspettazione”  $\langle \phi \rangle$  (questa è una notazione fisica, in matematica si usa invece la notazione  $\mathbb{E}(\phi)$ ) la quantità

$$\langle \phi \rangle = \int \phi(x)f(x)dx .$$

Supponiamo ora di considerare  $\phi(x) = x$ , il valore di aspettazione diventa quindi

$$\mu_1 = \langle x \rangle = \int_{\mathbf{R}} xf(x)dx ,$$

che prende il nome di momento primo o media (probabilistica) della distribuzione. Analogamente possiamo definire il momento  $n$ -esimo della distribuzione come

$$\mu_n = \langle x^n \rangle = \int_{\mathbf{R}} x^n f(x)dx .$$

Infine, usando il momento primo e secondo della distribuzione di probabilità possiamo definire la varianza come segue

$$\text{Var}(x) \equiv \sigma^2 := \langle x^2 \rangle - \langle x \rangle^2 = \mu_2 - \mu_1^2 .$$

### Funzione caratteristica e cumulanti

Data una distribuzione di probabilità  $f(x)$ , consideriamo il valore di aspettazione della funzione  $\phi(x) = e^{ikx}$  e definiamo la funzione caratteristica o funzione

generatrice dei momenti della distribuzione  $f(x)$ , in modo analogo a una trasformata di Fourier possiamo usare questa funzione per definire i momenti per derivazione

$$G_x(k) = \langle e^{ikx} \rangle = \int_{\mathbf{R}} e^{ikx} f(x) dx .$$

Possiamo osservare che, data questa definizione, vale

- $G_x(0) = 1$
- $\mu_n = \left(\frac{1}{i}\right)^n \frac{d^n}{dk^n} G_x(k)|_{k=0} \rightarrow G_x(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \mu_n$  ,

cioè come desiderato otteniamo tutti i momenti dalla funzione generatrice.

Data la funzione caratteristica di una distribuzione possiamo anche definire il “cumulante”  $n$ -esimo  $c_n$  come

$$c_n = \frac{d^n}{dk^n} \log G_x(k)|_{k=0}$$

In particolare, i primi quattro sono:

- $c_1 = \mu_1$  valore medio della distribuzione,
- $c_2 = \sigma^2$  varianza della distribuzione,
- $c_3$  parametro di skewness (simmetria) della distribuzione (misura quanto una distribuzione è asimmetrica per riflessione),
- $c_4$  indice di curtosi della distribuzione (misura quanta probabilità va nelle code).

I cumulanti misurano i contributi dei momenti che sono genuinamente dell'ordine del momento e non dipendono da momenti di ordine precedente (inoltre come vedremo hanno la proprietà di sommarsi per somme di random variabili indipendenti).

## Indipendenza

Dati due eventi  $A$  e  $B$ , si definisce la probabilità condizionata “di  $B$  dato  $A$ ” come la probabilità che accada  $B$  assumendo che sia accaduto  $A$ ,

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

Dove la seconda uguaglianza è una notazione introdotta per la probabilità “congiunta” degli eventi  $A$  e  $B$  (la probabilità che accadano entrambi e quindi dell'intersezione).

Due eventi  $A$  e  $B$  si definiscono indipendenti se e solo se vale

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

condizione equivalente a

$$P(A, B) = P(A)P(B) .$$

Dalla definizione di probabilità condizionata deriva immediatamente il teorema di Bayes. Dati due eventi  $A$  e  $B$ , vale la seguente relazione tra le probabilità condizionate,

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

La dimostrazione segue immediatamente dalle condizioni

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

e

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

Il teorema di Bayes ha un'interpretazione interessante, che lo rende il punto di partenza per l'interpretazione soggettiva della probabilità. Supponiamo di avere un modello teorico di partenza che descriva gli eventi, e quindi che  $A$  sia un'ipotesi di ciò che debba accadere. Interpretiamo invece  $B$  come un'osservazione dei dati. Allora  $P(A)$  è la probabilità "a priori" che potesse accadere  $A$  e  $P(B|A)$  (likelihood o verosimiglianza) è una quantità che si può calcolare dal modello (l'aspettativa di osservare ciò che veramente si è osservato date le credenze a priori).  $P(A|B)$  ("posterior") rappresenta come la visione del modello, l'aspettativa, sia cambiata osservando i dati. Si può vedere il posterior  $P(A|B)$  come  $P(A)$  moltiplicata per un coefficiente di "likelihood". In questa interpretazione,  $P(B)$  è una normalizzazione ("evidence"), che misura quanto l'evento osservato fosse previsto.

Questa formula trova molte applicazioni nella risoluzione di problemi di probabilità, come il problema di Monty-Hall, o quello del gemello di Elvis, così formulato:

Elvis aveva un gemello, il quale è morto alla nascita. Qual è la probabilità che fosse un gemello identico?

## Random variabili indipendenti

Discutiamo ora la dipendenza statistica di variabili aleatorie. Due variabili aleatorie  $X_1$  e  $X_2$  si dicono indipendenti quando la PDF congiunta è il prodotto delle PDF di singola variabile

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) ,$$

quindi per il teorema di Bayes,

$$f(x_1|x_2) = f_1(x_1) .$$

Di conseguenza, per le funzioni generatrici vale

$$G(k_1, k_2) = G_1(k_1)G_2(k_2) ,$$

e per quelle connesse vale una proprietà di somma,  $\log G(k_1, k_2) = \log G_1(k_1) + \log G_2(k_2)$ .

Per misurare la dipendenza, è utile definire la “covarianza” di due variabili aleatorie come

$$\text{cov}(X_1, X_2) = \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle ,$$

con terminologia fisica questa è una funzione di correlazione “connessa”,  $\langle X_1 X_2 \rangle_c$ . Due variabili aleatorie indipendenti hanno covarianza nulla. Infatti

$$\langle X_1 X_2 \rangle = \int_{\mathbb{R}} dx_1 dx_2 f(x_1, x_2) x_1 x_2 = \int_{\mathbb{R}} dx_1 f_1(x_1) x_1 \int_{\mathbb{R}} dx_2 f_2(x_2) x_2 = \langle X_1 \rangle \langle X_2 \rangle .$$

**NB:** non è, invece, generalmente vero che due variabili a covarianza nulla sono indipendenti. Un semplice controesempio è il caso di  $x_1$  RV gaussiana a media nulla (o in generale con PDF simmetrica) e  $x_2 = x_1^2$ . Vale  $\langle x_1 x_2 \rangle = \langle x_1^3 \rangle = 0$ , i.e. la covarianza è zero, ma le due variabili sono dipendenti essendo una funzione dell'altra.

### Somma di random variabili indipendenti.

Siano  $X_1$  e  $X_2$  due random variabili indipendenti e  $Y = X_1 + X_2$  la loro somma. Allora possiamo scrivere la distribuzione di probabilità di  $Y$  come convoluzione,

$$f_Y(y) = \int dx_1 dx_2 f(x_1, x_2) \delta(x_1 + x_2 - y) = \int dx_1 f(x_1, y - x_1) = \int dx_1 f_{X_1}(x_1) f_{X_2}(y - x_1)$$

(dove l'ultima uguaglianza vale perché le variabili sono indipendenti) e la funzione generatrice come

$$G_Y(k) = \langle e^{ik(x_1+x_2)} \rangle = G_{X_1}(k)G_{X_2}(k) . \quad (2)$$

Quindi tutti i cumulanti si sommano per random variabili indipendenti poichè vale

$$\log G_Y(k) = \log G_{X_1}(k) + \log G_{X_2}(k) . \quad (3)$$

In particolare  $\langle Y \rangle = \langle X_1 \rangle + \langle X_2 \rangle$ ,  $\sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2$ , etc.

### Trasformazione di random variabili.

Sia  $X$  una random variabile e  $Y = \phi(X)$ , la cui funzione di distribuzione è

$$f_Y(y) = \int dx f_X(x) \delta(y - \phi(x)) . \quad (4)$$

Questa relazione implica che nella funzione generatrice si possano fare le medie con la distribuzione di  $X$ , cioè

$$G_y(k) = \langle e^{ik\phi(x)} \rangle_x . \quad (5)$$

Inoltre se  $\phi(x)$  è monotona crescente vale

$$F_Y(y_0) = P(Y \leq y_0) = P(X \leq x_0) = F_X(x_0) .$$

per le CDF, mentre per le PDF

$$f_X(x) = \frac{d}{dx} F_X(x) = \phi'(x) \frac{d}{dy} F_Y(y) = \phi'(x) f_Y(y)$$

quindi la PDF di una trasformazione di una random variabile è

$$f_Y(y) = \left( \frac{dy}{dx} \right)^{-1} f_X(x) . \quad (6)$$

### Random variabili gaussiane.

Le random variabili gaussiane sono random variabili continue la cui funzione di distribuzione e funzione generatrice sono rispettivamente

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} \quad (7)$$

$$G(k) = \int dx e^{ikx} f(x) = e^{ik\mu_1 - \frac{1}{2}\sigma^2 k^2} , \quad (8)$$

dove la seconda relazione è ottenuta con un integrale gaussiano. Osserviamo che

$$\log G(k) = ik\mu_1 - \frac{1}{2}\sigma^2 k^2 , \quad (9)$$

quindi tutti i cumulanti a eccezione dei primi due sono nulli; ovvero media e varianza determinano completamente la distribuzione.

Generalizzando a  $N$  dimensioni

$$f(x_1, \dots, x_N) = \frac{1}{Z} e^{-\frac{1}{2} x^T A x - b x} \quad (10)$$

- dove  $A$  è una matrice simmetrica definita positiva mentre  $Z$  una costante di normalizzazione pari a

$$Z^{-1} = (2\pi)^{-\frac{N}{2}} (\det A)^{\frac{1}{2}} e^{-\frac{1}{2}b^T A^{-1}b} \quad (11)$$

- i valori di aspettazione si trovano da

$$\langle x_i \rangle = \frac{\partial}{\partial b_i} \log Z = - \sum_{i,j} (A^{-1})_{ij} b_j \quad (12)$$

da cui si può vedere che  $\text{cov}(x_i, x_j) = A_{ij}^{-1}$ , quindi due random variabili *gaussiane* con covarianza zero sono indipendenti.

### Nota tecnica su integrale gaussiano.

Per comodità includiamo un reminder dell'integrale gaussiano. Il punto di partenza è la relazione

$$\int_{\mathbb{R}} dx e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \sqrt{2\pi\sigma^2} .$$

Questa relazione si dimostra in molti modi, per esempio dalla eguaglianza

$$\sqrt{\left( \int_{\mathbb{R}} dx e^{-x^2} \right)} = \sqrt{\int_{\mathbb{R}} dx \int_{\mathbb{R}} dy e^{-(x^2+y^2)}},$$

risolvendo il secondo integrale in coordinate polari.

Da questa relazione si ricava facilmente che

$$\int_{\mathbb{R}} \frac{dx}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x)^2}{2\sigma^2} + \frac{Jx}{\sigma^2}} = e^{\frac{\sigma^2 J^2}{2}} .$$

Questo perché

$$\int_{\mathbb{R}} \frac{dx}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + Jx} = \int_{\mathbb{R}} \frac{dx}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + Jx + \frac{J^2}{2} - \frac{J^2}{2}} = e^{\frac{J^2}{4}},$$

essendo

$$-\frac{x^2}{2} + Jx + \frac{J^2}{2} = \left( \frac{x}{\sqrt{2}} + \frac{J}{\sqrt{2}} \right)^2 .$$

Infine nel caso N-dimensionale possiamo scrivere per  $x, j \in \mathbb{R}^N$  e  $A$  matrice  $N \times N$

$$\int_{\mathbb{R}^N} d^N x e^{-\frac{1}{2}x^t A x + x^t j} = \int_{\mathbb{R}^N} d^N x \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}y^t y + y^t Q^{-1}j},$$

dove la matrice  $Q$  è tale che  $Q^2 = A$ , quindi  $x^t Ax = x^t Q Q x = y^t y$  e  $d^N x \det Q = d^N x \sqrt{\det A} = d^n y$ . Da cui si ottiene che

$$\int_{\mathbb{R}^N} d^N x e^{-\frac{1}{2} x^t A x + x^t j} = \frac{(2\pi)^{N/2}}{\sqrt{\det A}} e^{j^t A^{-1} j}.$$

Questa relazione vale anche se la matrice  $A$  non è simmetrica, e in questo caso conta solo la sua parte simmetrica,  $(A + A^t)/2$ .

### Random variabili Bernoulli e distribuzione binomiale.

Una random variabile di Bernoulli è l'esempio più semplice di variabile stocastica discreta. I valori possibili sono binari, come "testa/croce" oppure "successo/fallimento", mappabili per esempio in  $\{0, 1\}$ . La PDF è definita univocamente da un parametro,  $p = p(1)$ , dato che  $p(0) = 1 - p$ .

Una sequenza di "tiri" (RV) di Bernoulli è detto "esperimento di Bernoulli". Se  $X$  è una RV di Bernoulli, è facile rendersi conto che  $\langle x \rangle = p$  e che  $\text{Var}(x) = p(1 - p)$ .

Possiamo chiederci quale sia la probabilità di avere  $k$  successi in  $n$  prove (per esempio di ottenere  $k$  teste tirando  $n$  monete). Si noti che formulando questa domanda stiamo definendo un'altra variabile aleatoria, derivata da un'osservazione (il numero di successi) su più tiri Bernoulli. Se la variabile associata ad ogni tiro  $X_i$  era a valori in  $\{0, 1\}$ , possiamo dire che la nuova variabile sia  $Y = \sum_{i=1}^n (1 - X_i)$ . La probabilità di avere  $k$  successi (PDF di  $Y$ ) sarà  $p^k (1 - p)^{n-k}$ , da moltiplicare, dato che non ci interessa l'ordine, per il numero di modi in cui si possono ottenere  $k$  successi in  $n$  prove, i.e. il numero di combinazioni di  $k$  oggetti in  $n$  caselle. Questo ragionamento fornisce la "distribuzione binomiale"

$$B_n(k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Come si può mostrare facilmente, la media di una RV binomiale è  $\langle y \rangle = np$ , e la sua varianza  $\text{Var}(y) = np(1 - p)$ .

Un'altra considerazione classica è che  $B_n(k) \rightarrow \text{Poiss}_k(\langle y \rangle)$ , quando  $p \rightarrow 0$  e  $n \rightarrow \infty$  in modo che  $pn$  sia costante, dove

$$\text{Poiss}_k(\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

è la distribuzione di Poisson. La dimostrazione usa l'approssimazione di Stirling del fattoriale.

Inoltre vedremo tra poco che per il teorema del limite centrale quando  $p$  rimane fisso e  $n \rightarrow \infty$  si ottiene la distribuzione Gaussiana  $G(np, np(1-p))$ .

Da osservabili basate su tiri di Bernoulli si possono ricavare e studiare diverse random variabili e corrispondenti PDF interessanti. Le più importanti sono la “binomiale negativa” (che deriva dal numero di insuccessi, ed è detta anche di Pascal, di Polya, o in un caso limite geometrica), e la distribuzione “ipergeometrica” (che deriva da un processo di estrazione da un’urna, senza reimmissione degli elementi estratti).

In fisica statistica, le random variabili di Bernoulli sono alla base dei modelli di spin “tipo Ising” cioè in cui gli spin sono variabili binarie (a valori in  $\{+1, -1\}$ ).

## Random Walk a passi costanti

Consideriamo ora un semplice esempio di random walk, definito come processo a passi costanti fatto di tiri di Bernoulli. Questo modello si può analizzare in dettaglio dato che è costruito usando random variabili molto semplici, e come vedremo i risultati si estendono molto in generale. Diciamo che ogni passo temporale  $\Delta t$  corrisponde a un tiro di Bernoulli, i.e. tiriamo una moneta di peso  $p$  (probabilità di ottenere testa). A seconda dell’esito del lancio, partendo da  $x = 0$  al tempo  $t = 0$ , facciamo un passo a destra quando estraiamo testa e un passo a sinistra quando estraiamo croce. Ogni passo ha lunghezza  $\Delta x$ . Ci chiediamo qual è la posizione media che raggiungiamo al tempo  $t$ , ovvero dopo  $N$  lanci di moneta  $t = N\Delta t$ .

Detti  $i$  e  $N - i$  rispettivamente il numero di passi a destra e a sinistra compiuti, si ha

$$X = (i - (N - i))\Delta x = \Delta x(2i - N) \quad (13)$$

$$\langle X \rangle = \Delta x(2\langle i \rangle - N) = \Delta x(2p - 1)N = (2p - 1)\frac{\Delta x}{\Delta t}t \quad (14)$$

dove  $\frac{\Delta x}{\Delta t} = v_{drift}$  definisce la velocità di drift.

Osserviamo che se  $p = \frac{1}{2}$  il primo momento di  $X$  è zero, ovvero in media non ci spostiamo dal punto di partenza. Guardiamo allora la varianza

$$\langle X^2 \rangle = (\Delta x)^2 \langle (2i - N)^2 \rangle = (\Delta x)^2 (4\langle i^2 \rangle - 4\langle i \rangle N + N^2) = (\Delta x)^2 (N + N^2 - 2N^2 + N^2) = (\Delta x)^2 N$$

quindi la varianza al tempo  $t$ , detta “mean square displacement” è

$$\text{MSD}(t) = \sigma^2(t) = \langle X^2(t) \rangle = \frac{(\Delta x)^2}{\Delta t}t := 2Dt \quad (15)$$

dove  $D = (\Delta x)^2/(2\Delta t)$  è il coefficiente di diffusione (dimensionalmente,  $[D] = L^2/T$ ). La deviazione standard, detta “root mean square displacement”,  $\text{RMSD}(t) \sim t^{1/2}$  definisce la distanza tipica percorsa dal random walk rispetto al punto di partenza, che scala come la radice del tempo (in questo caso si parla di “legge diffusiva”).

## 2 Leggi Limite

Questa sezione tratta un aspetto della teoria della probabilità molto importante per la meccanica statistica, vale a dire la possibilità di definire “attrattori” che unificano il comportamento statistico di sistemi a tanti gradi di libertà a priori molto diversi tra loro.

### Legge dei grandi numeri

Abbiamo finora chiamato “media” un oggetto che a priori ha poco a che fare con il concetto di media empirica con cui siamo familiari dalle elementari.

Siano  $X_1, \dots, X_N$  random variabili indipendenti e identicamente distribuite (“iid”) con PDF  $f(x)$  e sia la loro somma  $S_N = \sum_{i=1}^N X_i$ . Considerando la loro media aritmetica  $S_N/N$ , la legge dei grandi numeri dice che se il primo momento di  $X$   $\mu_1$  esiste (mentre  $\mu_2$  può essere anche infinito), questa media aritmetica converge al primo momento della distribuzione, per cui le due operazioni sono in qualche modo equivalenti,

$$\frac{S_N}{N} \rightarrow \mu_1 ,$$

per  $N \rightarrow \infty$ .

In realtà (sotto le stesse ipotesi) esistono due leggi dei grandi numeri, detta “legge forte” e “legge debole”, a seconda del tipo di convergenza che si prende in considerazione

- La legge debole, si richiede convergenza puntuale, in probabilità, cioè che  $\lim_{N \rightarrow \infty} \text{Prob} \left[ \left| \frac{S_N}{N} - \mu_1 \right| > \epsilon \right] = 0 \forall \epsilon > 0$
- La legge forte richiede convergenza in misura, i.e.  $\text{Prob} \left[ \lim_{N \rightarrow \infty} \frac{S_N}{N} = \mu_1 \right] = 1$  quasi certamente.

Vediamo la dimostrazione della legge debole. Un modo di dimostrarla passa attraverso la seguente “diseguaglianza di Chebyshev”. Sia  $X$  una variabile stocastica,

$\mu$  il suo valore di aspettazione e  $\sigma^2$  la sua varianza. Allora per ogni numero reale positivo  $k$  vale:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \iff P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \quad (16)$$

Assumendo questa disuguaglianza, la legge dei grandi numeri può essere vista come caso particolare, nel limite  $N \rightarrow \infty$ . Infatti, scegliendo  $X = \frac{S_N}{N}$  e  $k = \epsilon$  si ha

$$P\left(\left|\frac{S_N}{N} - \mu\right| > \epsilon\right) \leq \frac{\text{Var}(S_N/N)}{\epsilon^2} = \frac{\frac{1}{N^2} \sum \text{Var}(X_i)}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2} \rightarrow 0$$

Dimostriamo pertanto la disuguaglianza di Chebyshev (16). A tal proposito sia  $X$  una variabile stocastica continua con distribuzione di probabilità  $f(x)$ . Allora

$$\begin{aligned} \sigma^2 &= \int_{\mathbb{R}} dx (x - \mu)^2 f(x) \geq \int_{|x-\mu| \geq k\sigma} dx (x - \mu)^2 f(x) \\ &\geq \int_{|x-\mu| \geq k\sigma} dx k^2 \sigma^2 f(x) = k^2 \sigma^2 P(|X - \mu| \geq k\sigma) \end{aligned}$$

da cui segue la tesi.

## Teorema del limite centrale

Il teorema del limite centrale identifica la distribuzione limite di somme di random variabili iid (e quindi pensando alla fisica statistica, di medie empiriche di variabili microscopiche) come una gaussiana, con una legge di scala specifica (quella “diffusiva”) col numero delle variabili. Oltre a stabilire l’universalità del comportamento che abbiamo già visto per il random walk, è fondamentale in meccanica statistica perché mostra la *possibilità* di avere comportamenti universali anche in casi più complessi (non iid, variabili interagenti) di interesse fisico e non. Infine il teorema trova applicazione in approcci di “campo medio”, dove variabili interagenti si trattano come se fossero iid.

Vediamo l’enunciato del teorema. Siano (come prima)  $X_1, \dots, X_N$  random variabili identicamente distribuite con  $\mu_1$  finito (possiamo traslare  $X_i \rightarrow X_i - \mu_1$  e ottenere il caso con media nulla) e *varianza finita*,  $c_2 = \sigma^2 < \infty$ . Definiamo

$$Z_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i = \frac{1}{\sqrt{N}} S_N.$$

Il teorema del limite centrale afferma che  $Z_N$  è una random variabile gaussiana per  $N \rightarrow \infty$  con PDF

$$f_N(Z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{Z^2}{2\sigma^2}\right), \quad (17)$$

dove  $\sigma$  è la deviazione standard della singola variabile. La distribuzione gaussiana può quindi essere vista come un attrattore nello spazio delle distribuzioni di probabilità a varianza finita.

Per dimostrare il teorema si può mostrare che la funzione generatrice di  $Z_N$  è quella di una RV gaussiana

$$G_{Z_N}(k) = \left(\langle e^{ikX/\sqrt{N}} \rangle_X\right)^N = \left(G_X\left(\frac{k}{\sqrt{N}}\right)\right)^N$$

$$\log G_{Z_N}(k) = N \log G_X\left(\frac{k}{\sqrt{N}}\right) = N \left(-\frac{1}{2}\sigma^2 \frac{k^2}{N} + \frac{1}{6} \frac{c^3}{N^{3/2}} (ik)^3 + \dots\right) \rightarrow -\frac{1}{2}\sigma^2 k^2,$$

dato che il primo termine nello sviluppo è costante per  $N \rightarrow \infty$  mentre gli altri tendono a zero.

## Leggi di Levy

Il teorema del limite centrale si fonda su due ipotesi fondamentali, variabili iid e varianza finita. Il caso di variabili non iid è complesso (e si può dire che in qualche senso “equivale” alla meccanica statistica di equilibrio) e affrontato solo in alcuni casi particolari. Nel caso di variabili iid a varianza infinita, Levy ha classificato i possibili attrattori.

Per capire come ripercorriamo il passo centrale della dimostrazione del teorema del limite centrale, senza l'ipotesi di varianza finita, consideriamo variabili stocastiche  $X$  con densità di probabilità della forma

$$f(x) \sim |x|^{-(\alpha+1)},$$

con  $\alpha \leq 2$ .

Partiamo da un esempio canonico, che è la distribuzione di Cauchy, ottenuta per  $\alpha = 1$

$$f(x) = \frac{c_0}{\pi} \frac{1}{c_0^2 + x^2}. \quad (18)$$

Considerando la sua funzione generatrice,

$$G_X = \langle e^{ikX} \rangle = \int dx f(x) e^{ikx} = e^{-c_0|k|} \simeq 1 - C|k| \quad k \rightarrow 0, \quad (19)$$

dove l'integrale si può effettuare col teorema dei residui. Quindi se definiamo la variabile riscalata

$$Z_N = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{N} S_N \quad (20)$$

$$G_{Z_N}(k) = \left( e^{-c_0|k|/N} \right)^N = e^{-c_0|k|} = G_x(k) \quad (21)$$

questa è una “legge stabile”, i.e. (come nel caso del teorema del limite centrale) è invariante per somma, dato il riscalamento  $1/N$ , e quindi un attrattore. Per tutte le leggi stabili vale un teorema di tipo limite centrale. Levy ha caratterizzato le leggi stabili.

Per caratterizzare le leggi stabili dobbiamo trovare il riscalamento opportuno da applicare alla random variabile somma  $S_N$ , di variabili iid la cui PDF è  $f(x) \sim |x|^{-(1+\alpha)}$ , per  $\alpha$  generico. Nel caso in cui  $\alpha > 2$  la varianza è finita e siamo nel caso del teorema del limite centrale. Nel caso in cui  $\alpha \leq 2$  la varianza diverge, non sono verificate le ipotesi del teorema del limite centrale e quindi l'attrattore della distribuzione di probabilità  $f(x)$  non è gaussiano. Ci chiediamo se è comunque possibile riscalare la variabile casuale  $S_N$  (la somma di  $X_i$ ) in maniera tale da ottenere la stessa proprietà stabile ottenuta per rv Cauchy, e identificare un attrattore a cui la distribuzione di probabilità della somma converge. Sappiamo che il riscalamento è  $1/N$  per  $\alpha = 1$  e  $1/N^{1/2}$  per  $\alpha > 2$ . Una legge che interpola questi comportamenti è  $1/N^{1/\alpha}$ , che si rivela essere il riscalamento giusto.

Possiamo ripetere il ragionamento precedente. In questo caso, per proprietà della trasformata di Fourier di leggi di potenza (per esempio riportato in appendice nel “trattatello” di Marinari e Parisi), vale

$$G_X(k) = \int dx f(x) e^{ikx} \sim e^{-Ck^\alpha}$$

che subito porta alla conclusione che il riscalamento adatto sia  $Z_N = S_N/N^{1/\alpha}$ , dato che

$$G_{Z_N}(k) = \left( e^{-c(|k|/N^{1/\alpha})^\alpha} \right)^N = G_x(k) \quad (22)$$

Per classificare in generale gli attrattori (teorema di Levy), è necessario distinguere in base alla simmetria della considerata, come segue

Nel caso di  $f(x)$  simmetrica,  $f(x) = f(-x)$   $f(x) \sim |x|^{-(1+\alpha)}$  per  $|x| \rightarrow \infty$  (con  $\alpha \leq 2$ ). Definendo come abbiamo detto  $Z_N = \frac{1}{N^{1/\alpha}} S_N$  si ha che  $Z_N$  converge (puntualmente) alla random variabile  $Z$ ,  $Z_N \xrightarrow{n \rightarrow \infty} Z$ , dove  $Z$  distribuzione di probabilità  $f(z)$  in generale caratterizzata dalla sua (anti)trasformata di Fourier,

$$f(z) = L_{\alpha,0}(z) = \frac{1}{2\pi} \int dk e^{-ikz} e^{c|k|^\alpha}, \quad (23)$$

dove l'ultimo integrale rappresenta l'antitrasformata di  $G_Z(k)$ , e definisce implicitamente la legge stabile.

Nel caso di  $f(x)$  asimmetrica, la PDF della variabile  $X$  è data da

$$f(x) = \begin{cases} c_+ |x|^{-(1+\alpha)} & \text{per } x \rightarrow +\infty \\ c_- |x|^{-(1+\alpha)} & \text{per } x \rightarrow -\infty \end{cases} \quad (24)$$

È quindi necessario definire la variabile casuale  $Z_n$  diversamente a seconda del valore che assume il parametro  $\alpha$ , nello specifico definiamo

$$Z_n = \begin{cases} \frac{1}{n^{1/\alpha}} (S_n - \mu_1 n) & \alpha \in (1, 2) \\ \frac{1}{n^{1/\alpha}} S_n & \alpha \in [0, 1] \end{cases} \quad (25)$$

Anche in questo caso vale che  $Z_n \xrightarrow[n \rightarrow \infty]{} Z$ , e la distribuzione di probabilità di  $Z$  è definita dalla trasformata di Fourier

$$f(Z) = L_{\alpha, \beta}(Z) = \frac{1}{2\pi} \int dk e^{-ikZ} e^{c|k|^\alpha (1 + i\beta \operatorname{sgn}(k) \tan \frac{\pi\alpha}{2})}. \quad (26)$$

Oltre alla Cauchy, esiste un'espressione chiusa per il caso  $L_{\frac{1}{2}, 1}(z)$

Osserviamo che rispetto al caso rappresentato del teorema del limite centrale qui non vi è più un singolo attrattore, ma una classe a un (e due) parametro di attrattori  $L_{\alpha, \beta}$  che variano a seconda di come decade la pdf delle random variabili di partenza. Questi due casi sono molto meno generali del precedente, e le classi di universalità trovate nel caso di varianza non finita sono molto più piccole di quella descritta dal limite centrale.

Avendo visto casi diversi di teoremi limite, riprendiamo il loro nesso con la meccanica statistica. In meccanica statistica siamo interessati a ottenere le medie di stati macroscopici, trattati come random variabili, in modo da poter utilizzare principi simili (o uguali) a teoremi limite per osservabili macroscopiche (tipicamente, medie). Comportamenti analoghi a teoremi limite stabiliscono classi di universalità. Se le random variabili sono iid e presentano varianza finita l'attrattore è gaussiano. Nel caso di assenza di varianza c'è un continuo di classi di universalità "più piccole", parametrizzate da  $\alpha$ , che seguono leggi di Levy. In meccanica statistica le variabili elementari sono interagenti, e si perde invariabilmente l'ipotesi di indipendenza, e quindi la situazione è più complessa dal lato matematico, ma, come ci aspettiamo, esistono classi di universalità che è possibile esplorare con i mezzi più euristici della fisica teorica. Il ponte più semplice tra il teorema del limite

centrale standard e la meccanica statistica sono le teorie di campo medio, dove si cerca di rappresentare un sistema di variabili dipendenti con uno di variabili indipendenti (con parametri diversi). Per esempio, consideriamo  $N$  agenti descritti dalle variabili  $x_1, \dots, x_N$ , gli stati sono caratterizzati da probabilità congiunte  $p(x_1, \dots, x_N)$  che in una teoria di campo medio si assumono (con un ansatz) fattorizzate  $p(x_1, \dots, x_N) = \prod_j p(x_j)$ . Sotto questa (radicale ma utile) ipotesi varrà anche il teorema del limite centrale.

## Random Walk, teorema del limite centrale e voli di Levy

Vediamo ora come i teoremi limite permettano di capire come il comportamento dei random walk sia universale. Vediamo sia il caso di varianza finita sia quello di varianza infinita, dove non è possibile applicare il teorema del limite centrale, ma ci sono voli di Levy. Precedentemente, in un modello di random walk, avevamo assunto che la lunghezza dei passi fosse costante, e quindi i passi erano variabili Bernoulli. Ora invece analizziamo il caso generale in cui la lunghezza del passo è una variabile casuale non specificata (e quindi applicabile a modelli dei processi più svariati fisici e non), nei due casi di varianza finita e infinita. Anticipiamo che per sistemi empirici finiti avere varianza infinita non può significare che i singoli passi possano avere lunghezza infinita ma qualsiasi distribuzione avrà come cutoff la taglia del sistema. Quindi il fatto di avere voli di Levy va interpretato come una proprietà di scala dei passi con la taglia del sistema.

Supponendo di essere in uno spazio  $d$  dimensionale, chiamiamo  $X_i$  le RV iid associate alla lunghezza del passo, la cui PDF  $f(x)$  è assunta simmetrica. La somma di  $N$  passi casuali è quindi data da  $S_N = \sum_{i=1}^N x_i$ . In una dimensione, la distanza dall'origine dopo  $N$  passi è quindi  $R_N = R_t = S_N$ . In  $d$  dimensioni avremo un vettore  $\vec{R}_N = \sum \vec{x}_i$ . Assumiamo che  $\vec{X}_i$  siano RV iid simmetriche, in modo da non avere drift, quindi  $\langle \vec{x}_i \rangle = \langle \vec{R}_N \rangle = 0$ .

Consideriamo prima il caso a *varianza finita*,  $\sigma_x^2 < +\infty$ . Possiamo applicare il teorema del limite centrale, e otteniamo che la distribuzione del vettore somma è gaussiana

$$f\left(\frac{\vec{R}_n}{\sqrt{n}}\right) = G(0, \sigma_x^2), \quad (27)$$

e in particolare che lo scaling con  $N$  dà una legge diffusiva

$$\text{MSD}(t) = \langle |R_n|^2 \rangle \sim \sigma^2 dN \sim \sigma^2 dt. \quad (28)$$

Questo significa che qualsiasi sia tipo o legge dei passi, sottraendo un drift e ammettendo che questi abbiano varianza finita (cosa di solito ragionevole in casi

empirici) si avrà una legge diffusiva. In altre parole il random walk è un comportamento universale e “emergente”, nel senso che può scaturire da un gran numero di processi microscopici. È evidente che questo risultato è molto importante dal punto di vista della meccanica statistica.

Nel caso di *varianza infinita*, possiamo trattare la PDF delle variabili microscopiche, i passi, come una funzione che ha decadimento algebrico  $f(x) \sim |x|^{-(1+\alpha)}$ , con  $\alpha < 2$ , e applicare le considerazioni che portano al teorema di Levy. Riscaliamo quindi la variabile  $\vec{R}_N$  con il fattore  $N^{-\frac{1}{\alpha}}$  ottenendo quindi convergenza alle distribuzioni Levy-stabili

$$f\left(\frac{R_N}{N^{\frac{1}{\alpha}}}\right) \xrightarrow{N \rightarrow \infty} L_{\alpha,0} . \quad (29)$$

Notiamo che in questo caso esiste un attrattore diverso (e quindi una “classe di universalità” distinta) per ogni valore dell’ esponente  $\alpha$ .

In questo caso la legge di scala tra il tempo ( $N$ ) e l’estensione lineare ( $\vec{R}_N$ ) della regione di spazio che viene esplorata è

$$\langle |\vec{R}_N| \rangle \sim N^{\frac{1}{\alpha}} \quad (30)$$

(con  $\alpha < 2$ ). Ricordiamo che per il moto browniano valeva una legge analoga con  $\alpha = 2$ , si ha quindi che nel caso in cui si ha varianza infinita  $R_N$  cresce con  $N$  più velocemente che nel random walk browniano. Questo processo è quindi chiamato “superdiffusivo”. In secondo luogo osserviamo che in questi processi con varianza infinita il MSD *non esiste* per  $N \rightarrow \infty$ .

Chiaramente in un caso sperimentale/empirico non è possibile verificare direttamente se il MSD assuma un valore non finito. Questo ci porta dunque ad un valore finito per il MSD in ogni realizzazione. Esistono dunque altri modi per verificare l’assenza del secondo momento per una data random variabile a partire dai dati delle singole realizzazioni. Infatti possiamo considerare il MSD, calcolato dai cammini finiti, come una random variabile notiamo che questa non è “ergodica”, i.e. medie temporali di una traiettoria e medie di ensemble su tante traiettorie non coincidono, e non è “self-averaging”, i.e. la varianza relativa

$$\theta := \text{Var}(R_N) / \langle R_N \rangle^2$$

non va a zero per  $N \rightarrow \infty$ , mentre il teorema del limite centrale garantirebbe che  $\theta \sim 1/N$  asintoticamente. In generale l’assenza di queste due proprietà può essere indice della divergenza di momenti di ordine superiore al primo delle variabili iniziali su cui si media.

## Foraging con voli di Levy

Come esempio di applicazione di random walk con voli di Levy, consideriamo l'analisi e modellizzazione del "foraging behavior" (in italiano, strategie nella ricerca di cibo). Altri esempi noti sono i prezzi nella borsa, e, in sistemi più fisici il gas di Lorentz bidimensionale e i "living polymers".

Con foraging si intende il tipico comportamento che assumono certi animali quando cercano il cibo, una via di mezzo tra una traiettoria balistica (o persistente nella direzione) in direzione del target e un random walk mentre l'animale cerca zone localizzate con cibo. Lo studio di dati empirici (inizialmente per albatros) mostra che il comportamento mostra leggi che non sono di tipo diffusivo, ma interpretabili come voli di Levy (e che quindi possono essere descritti in un modello come random walk con lunghezza degli step distribuita algebricamente).

Studiamo il caso di foraging in due dimensioni (Viswanathan et al. Nature 1996). Come input del modello abbiamo la distribuzione superficiale dei target  $\sigma$  (che assumiamo uniforme), e il raggio del campo visivo  $r_v$ , ovvero la distanza massima entro al quale l'animale vede il target, e può raggiungerlo con un moto balistico. Ci concentriamo sul caso in cui il range  $r_v$  del campo visivo sia molto più piccolo della distanza tipica  $\lambda$  tra i target, questo in due dimensioni equivale a richiedere che

$$\lambda \simeq (2r_v\sigma)^{-1} . \quad (31)$$

Chiamando  $l$  la lunghezza dello step elementare del random walk, diciamo che la distribuzione di probabilità ad essa associata è del tipo

$$p(l) \sim l^{-\mu} .$$

In questo modello, per  $\mu \in [1, 3]$  si hanno voli di Levy, e per  $\mu > 3$  vale il teorema del limite centrale.

Ci chiediamo se esiste un valore di  $\mu$  che ottimizzi il processo di foraging, cioè che massimizzi il numero di target raggiunti a parità di distanza percorsa. Per affrontare questo problema iniziamo con il definire nel modo seguente una "efficienza" del processo come il rapporto tra il numero di siti target visitati e la distanza totale percorsa,

$$\eta = \frac{\#\text{target}}{d_{tot}} = \frac{\#\text{target}}{\langle l \rangle N_{tot}} = \frac{1}{\langle l \rangle n} .$$

dove  $d_{tot}$  è la distanza totale percorsa che può essere espressa come il numero totale di passi fatti ( $N_{tot}$ ) per la lunghezza media degli stessi, mentre nell'ultima uguaglianza abbiamo sostituito il numero medio di step effettuati tra due target

$n = \frac{N}{\#target}$ . Si noti che l'efficienza si può ottimizzare sia rendendo il passo medio più piccolo che diminuendo il numero totale di passi.

Calcoliamo quindi  $\langle l \rangle$  mediante la definizione del primo momento:

$$\langle l \rangle = \int_{-\infty}^{+\infty} lp(l)dl$$

Assumiamo che se la lunghezza del passo è maggiore della distanza tipica tra i target il passo stesso non è lungo  $l$  ma  $\lambda$  (la motivazione ovvia è che passi di questa lunghezza portano a raggiungere i target). Possiamo quindi scrivere la lunghezza  $l$  del singolo step come segue

$$l = \begin{cases} l & \text{se } l < \lambda \\ \lambda & \text{se } l > \lambda \end{cases}$$

La media di  $l$  si può stimare quindi nel seguente modo

$$\langle l \rangle = \frac{\int_{r_v}^{\lambda} ll^{-\mu}dl + \lambda \int_{-\lambda}^{+\infty} l^{-\mu}dl}{\int_{r_v}^{+\infty} l^{-\mu}dl},$$

dove il primo termine al numeratore indica il passo medio se il passo è minore di  $\lambda$ , e il secondo il fatto che se il passo tirato è maggiore si trova il target e ci si ferma. Riscrivendo,

$$\langle l \rangle = \left( \frac{\mu - 1}{2 - \mu} \right) \left( \frac{\lambda^{2-\mu} - r_v^{2-\mu}}{r_v^{1-\mu}} \right) + \left( \frac{\lambda}{r_v} \right)^{2-\mu} \quad (32)$$

Osserviamo già ora che per  $\mu < 2$  si ha una forte divergenza di  $\langle l \rangle$  per  $\lambda \rightarrow \infty$ . Questo ce lo aspettiamo in quanto per  $\mu < 2$  la variabile casuale  $l$  non ha primo momento definito e, in base alla formula precedentemente calcolata, si vede che  $\langle l \rangle$  corrisponde al primo momento.

Per poter concludere il calcolo sull'efficienza rimane da determinare il numero medio di step tra due target consecutivi ( $n$ ). A questo scopo è necessario fare una distinzione fra due tipi di foraging: "distruttivo" e "non distruttivo". Con foraging distruttivo intendiamo un processo in cui il target, una volta raggiunto, non è raggiungibile una seconda volta (per es. è una preda che viene mangiata). Viceversa con foraging non distruttivo intendiamo un processo in cui il target viene mantenuto anche dopo il raggiungimento (per esempio è un campo di erba). Questa distinzione nella tipologia di processo porta ad un diverso numero di passi medi tra target. Si ha

$$n = \begin{cases} \left( \frac{\lambda}{r_v} \right)^{\mu-1} & \text{se il processo è distruttivo} \\ \left( \frac{\lambda}{r_v} \right)^{\frac{\mu-1}{2}} & \text{se il processo è non distruttivo} \end{cases} \quad (33)$$

In conclusione l'efficienza in funzione del parametro  $\mu$  mostra un picco in  $\mu = 2$  nel caso in cui il processo sia non distruttivo, mentre ha un andamento costante nel caso in cui il processo distruttivo. Il picco per foraging non distruttivo giustifica l'osservazione empirica di  $\mu \simeq 2$  (Cauchy) per dati di traiettorie di calabroni, come soluzione che ottimizza il trade-off tra i target visitati e la distanza percorsa.

### Diffusione anomala dovuta a tempi di attesa versus anticorrelazione

Abbiamo visto che la superdiffusione può emergere se le probabilità degli step è a code di potenza con varianza non finita

$$p(l_i) \sim \frac{1}{|l_i|^{1+\alpha}}$$

È possibile avere superdiffusione anche nel caso in cui la distribuzione di lunghezza dei passi elementari abbia varianza finita, se i passi sono correlati, violando quindi l'ipotesi di indipendenza nel CLT (per esempio si pensi a una situazione in cui la probabilità di fare un passo a destra sia proporzionale alla frazione di passi a destra già fatti).

Allo stesso modo ci sono due tipi di scenari possibili per la "subdiffusione", cioè l'esistenza di leggi di scala sublineari che legano MSD a tempo o numero dei passi,

$$MSD(t) \sim t^\alpha \quad R_N \sim N^{\frac{\alpha}{2}} \quad \alpha < 1.$$

Nel primo tipo di scenario, esemplificato dal "Continuous Time Random Walk (CTRW)", i passi sono indipendenti, ma i tempi di attesa  $\tau$  hanno una distribuzione di tipo Levy. Il secondo tipo di scenario si presenta per esempio nei fenomeni viscoelastici, cui step successivi sono anticorrelati.

### Continuous Time Random Walk

Abbiamo detto che CTRW rompe le ipotesi del CLT mantenendo l'indipendenza degli step, ma mostrando una distribuzione di probabilità dei tempi di attesa tra gli step di tipo Levy

$$\psi(\tau) \sim \tau^{-(1+\alpha)}.$$

In questo caso, per violare il CLT non basta che  $\tau$  non abbia varianza, ma è necessario che non abbia media, è necessario quindi chiedere che  $\alpha < 1$ . Infatti se  $\langle \tau \rangle < \infty$  allora  $\langle R(\tau)^2 \rangle \simeq \frac{t}{\langle \tau \rangle}$ .

Nel caso in cui  $\langle \tau \rangle = \infty$  possiamo capire come si arriva alla subdiffusione considerando la variabile temporale  $t_N$  che rappresenta il tempo trascorso dopo  $N$  step. Vale

$$t_N = \sum_{i=1}^N \tau_i \simeq N^{\frac{1}{\alpha}} \gg N$$

e in questo caso  $R_N^2 \sim N \sim t^\alpha (\ll t)$ , e siccome  $\alpha < 1$  questo processo segue una legge scala, ma più lenta di un processo diffusivo.

Un esempio di modello più fisico di questo scenario è un cammino unidimensionale soggetto ad un potenziale “disordinato” (random) che presenta varie buche, spesso chiamato “Trap Model”. Chiamando  $\Delta E_i$  la random variabile che corrisponde alla profondità delle buche, per violare il CLT deve valere che la distribuzione dei  $\Delta E_i$  sia tale che non esista la media di  $e^{C\Delta E_i}$ . Sperimentalmente è possibile realizzare questo scenario utilizzando dei colloidi di dimensioni  $\sim 1\mu m$  molto densi che si dispongono in una conformazione disordinata.

## Viscoelasticità

La viscoelasticità è caratterizzata da passi anticorrelati. Di conseguenza, su certe scale temporali “brevi” un camminatore sente un potenziale di tipo armonico, ma su scale temporali più grandi no. Un modello che vedremo di questo scenario è la catena di Rouse che descrive la dinamica di un polimero come una catena di molle. In questo modello emerge che il moto del singolo monomero ha un comportamento subdiffusivo per scale temporali intermedie, mentre ha un comportamento diffusivo su scale temporali più lunghe.

Ci sono due modi in cui è possibile distinguere tra i due scenari. Il primo consiste nel controllare se alcune osservabili, come l’MSD, sono self-averaging. Nel caso della viscoelasticità l’MSD è self averaging, mentre nel caso di CTRW (come abbiamo visto) no. Un altro modo è osservare il comportamento delle correlazioni tra step successivi a fissata traiettoria. Le correlazioni  $\langle \frac{\Delta \vec{r}(t)}{\Delta t} \frac{\Delta \vec{r}(t+\tau)}{\Delta t} \rangle$ . In uno scenario di subdiffusione dovuta a CTRW queste correlazioni saranno sempre positive e tenderanno a zero all’aumentare del delay  $\tau$ . Nel secondo scenario le correlazioni saranno inizialmente positive per valori piccoli del delay  $\tau$  (quando il passo scelto è minore del passo elementare del sistema), poi diventano negative perchè gli step sono anticorrelati a causa di una risposta elastica, e infine tenderanno a zero perchè l’anticorrelazione è presente solo su alcune scale temporali.

### 3 Entropia e maximum entropy

Come forse è già noto, l'entropia è un concetto fondamentale che nasce dalla fisica statistica (Boltzmann, Gibbs), ed è intimamente collegato alla teoria della probabilità e alla teoria dell'informazione (Shannon). Questa sezione rivede la definizione probabilistica di entropia e discute come il principio di massima entropia (spesso detto "MaxEnt"), che è una generica procedura di inferenza, possa essere usato come alternativa alle procedure standard per definire la teoria degli ensemble della meccanica statistica di equilibrio. L'utilità concettuale di questo approccio è che mostra come la meccanica statistica vada ben oltre le applicazioni a sistemi fisici da cui è nata, e si estenda a sistemi generici di agenti.

#### L'Entropia misura incertezza e informazione

Consideriamo una random variable con PDF  $p_i$ , per semplicità discreta anche se tutto è estendibile in modo semplice al caso continuo. Si definisce funzione entropia (di Gibbs o Shannon) la funzione

$$S(x_i) = - \sum_i p_i \log p_i .$$

Questa osservabile è associata a una random variable, (adimensionale così definita) e misura il "disordine" o dall'altro lato "l'esoticità" dei valori che può assumere. Valgono le seguenti proprietà

1. L'entropia è positiva,  $S \geq 0$ , perchè  $0 \leq p_i \leq 1$ .
2. Se c'è uno stato  $j$  con  $p_j = 1$  allora  $p_i = 0 \forall i \neq j$  e  $S = 0$ .
3. Se tutti gli stati sono equiprobabili  $p_i = p_j \forall i, j$  costanti, o equivalentemente se  $p_i = \frac{1}{\Omega}$  con  $\Omega$  numero degli stati, allora  $S_\Omega = \log \Omega$  (detta "Entropia di Boltzmann"). Quindi l'entropia di una distribuzione uniforme aumenta con il numero di stati  $\Omega$
4. L'entropia di Boltzmann è massima.
5.  $S$  è additiva nei sotto sistemi.

Vediamo come l'entropia di Boltzmann è massima, i.e. l'entropia è massima per una distribuzione omogenea. Per mostrarlo consideriamo la differenza tra l'entropia per una PDF generica e quella di Boltzmann,

$$S_\Omega - S = \log \Omega + \sum_{i=1}^{\Omega} p_i \log p_i$$

scrivendo  $\log \Omega = \sum_{i=1}^{\Omega} p_i \log \Omega$

$$S_{\Omega} - S = \sum_{i=1}^{\Omega} p_i \log(p_i \Omega) = \frac{1}{\Omega} \sum_{i=1}^{\Omega} p_i \Omega \log(p_i \Omega).$$

Chiamando  $x_i = p_i \Omega$  e  $Q(x_i) = p_i \Omega \log(p_i \Omega)$ , possiamo cercare il minimo di  $Q(x_i)$  imponendo il vincolo  $\sum_i x_i = \Omega$

$$\frac{\partial}{\partial x_i} \left[ \sum_j Q(x_j) + \lambda \left( \sum_j x_j - \Omega \right) \right] = 0 \quad \forall i,$$

che fornisce

$$\log(x_i) + 1 + \lambda = 0,$$

da cui

$$x_i = e^{-(1+\lambda)}.$$

Il vincolo diventa quindi

$$\Omega = \sum_i x_i = \Omega x_i = e^{-(1+\lambda)} \implies \lambda = -1,$$

da cui segue che  $x_i = 1$  e  $p_i = \frac{1}{\Omega} \forall i$ . Questa proprietà, in meccanica statistica, può essere utilizzata per definire l'ensemble microcanonico e la sua entropia.

Mostriamo ora l'additività nei sottosistemi. Consideriamo due sottosistemi  $S_1$  e  $S_2$  indipendenti. La PDF sarà quindi fattorizzata,  $p_{ij} = p_i^{(1)} p_j^{(2)}$  e l'entropia si può scrivere come

$$S = - \sum_{ij} p_{ij} \log p_{ij} = - \sum_{ij} p_i^{(1)} p_j^{(2)} (\log p_i^{(1)} + \log p_j^{(2)}) = S_1 + S_2$$

perché  $-\sum_{ij} p_i^{(1)} p_j^{(2)} \log p_i^{(1)} = -\sum_j p_j^{(2)} \sum_i p_i^{(1)} \log p_i^{(1)} = -1 \cdot \sum_i p_i \log p_i$ .

Vediamo ora come l'entropia è collegata all'informazione che si possiede sulla random variabile. A questo scopo ci chiediamo a quante domande sì/no è necessario rispondere per individuare lo stato che la random variabile prende (cioè quanti bit di informazione è necessario fornire). In altre parole è possibile quantificare l'informazione legata a una random variabile immaginando un albero binario che ad ogni nodo separa gli stati in base alla risposta ad una domanda dicotomica riguardo alla variabile. Se  $\Omega$  è l'insieme degli stati è possibile dividerlo in due insiemi disgiunti tali che  $\Omega = \Omega_1 \cup \Omega_2$ , e in modo che i due sottoinsiemi abbiano probabilità totale simile  $\sum_{i \in \Omega_1} p_i = \sum_{j \in \Omega_2} p_j \simeq \frac{1}{2}$ . Diciamo che se lo stato è in  $\Omega_1$  la variabile di scelta è uguale ad 1, ed è nulla in caso contrario. La stessa procedura si può

iterare, e immaginiamo di aver individuato un singolo stato dopo  $k$  iterazioni. Per la variabile di scelta avremo una stringa di zeri e uni di lunghezza  $k$  tale che

$$2^k p_j = 1 ,$$

o equivalentemente

$$k = -\log_2(p_j) .$$

La random variabile  $k$  è il numero di bit necessari ad individuare uno stato. Il suo valor medio sugli stati, a meno di costanti additive, è proprio uguale all'entropia,

$$\langle k \rangle = \sum_i p_i (-\log_2(p_i)) + \text{cost} = S + \text{cost}.$$

## Ensemble micro-canonico e canonico dal principio MaxEnt

Il metodo maximum entropy è usato in generale per inferenza di distribuzioni di probabilità per random variabili di cui sono noti solo alcuni vincoli "macroscopici", che riflettono le informazioni note sul comportamento di quella variabile. L'idea è di scegliere, tra tutte le distribuzioni  $p_i$  compatibili coi vincoli, quella di entropia massima, che corrisponde a quella più agnostica (più degenere) dati i vincoli, in quanto non esclude nessuna configurazione che soddisfa i vincoli. In altre parole una distribuzione MaxEnt non è sensibile ad alcun parametro che non appaia esplicitamente nei vincoli.

Questo asserto può essere legato al cosiddetto "postulato fondamentale" della meccanica statistica, secondo il quale in un sistema isolato e in equilibrio la distribuzione di probabilità dei microstati  $p(C)$  è tale che l'entropia  $S$  sia massima. In particolare nel caso si conservi l'energia  $p(E) = \frac{1}{\Omega(E)}$  e  $p(E') = 0$  per  $E' \neq E$ .

### Ensemble Micro-canonico

Vediamo ora come il principio di MaxEnt definisca l'ensemble microcanonico. Per caratterizzarlo utilizziamo il postulato fondamentale, che può essere espresso in due modi equivalenti. Come dalla definizione classica possiamo porre

$$S|_E = k_B \log \Omega|_E$$

avendo fissato l'energia del sistema fisico al valore  $E$ . Nel caso in cui siamo interessati ad un sistema non fisico, per esempio un problema di ottimizzazione, il ruolo dell'energia viene assunto per esempio da una funzione costo, che fornisce dei

vincoli e, in particolare, stabilisce quanto sia dispendioso spostarsi dalla soluzione ottimale. In modo equivalente, poniamo

$$p_i(E) = \frac{1}{\Omega(E)}$$

ovvero diciamo che la probabilità della configurazione  $i$ -esima per valore fissato di energia è data dal principio di massima entropia, cioè è pari all'inverso del numero di configurazioni con quel valore di energia.

Vediamo alcuni esempi. Trattandosi di sistemi a valore fissato dell'energia, non sono moltissimi gli esempi interessanti di applicazione del micro-canonico.

- Gas ideale. Per questo sistema il “numero di stati” nello spazio delle fasi con energia pari a  $E$  è dato da:

$$\Omega(E) = \frac{1}{h^{3N} N!} \int_{\Sigma(E)} d^{3N} q d^{3N} p$$

dove  $\Sigma(E)$  è l'ipersuperficie dello spazio delle fasi con energia pari a  $E$ . Utilizzando MaxEnt si ritrova l'equipartizione dell'energia e l'equazione di stato per i gas perfetti (si veda per esempio il capitolo 3 del Sethna).

- Sistema di 3 particelle con spin  $1/2$  con energia data da  $E = -\mu H$ ; ci sono 8 possibili configurazioni:

$$E = -3\mu; \quad \{ +, +, + \}; \quad p(3\mu) = 1$$

$$E = -\mu; \quad \left\{ \begin{array}{l} +, +, - \\ +, -, + \\ -, +, + \end{array} \right\}; \quad p_i(3\mu) = \frac{1}{3}, i = 1, 2, 3$$

$$E = \mu; \quad \left\{ \begin{array}{l} -, -, + \\ +, -, - \\ -, +, - \end{array} \right\}; \quad p_i(3\mu) = \frac{1}{3}, i = 1, 2, 3$$

$$E = 3\mu; \quad \{ -, -, - \}; \quad p(3\mu) = 1$$

- Consideriamo un polimero monodimensionale, che viene modellizzato come un random walk di  $N$  passi (monomeri). Siamo interessati all'estensione del polimero. Se abbiamo  $N$  monomeri, l'entropia è legata alle configurazioni della catena, a cui dobbiamo assegnare una distribuzione di probabilità. La direzione di ogni passo è data dal valore di una variabile di Bernoulli con probabilità  $1/2$ . Se indichiamo con  $N_+$  e  $N_-$  rispettivamente il numero di

passi verso destra e verso sinistra abbiamo che la distribuzione di probabilità MaxEnt è quella omogenea,

$$p(N_+, N_-) = \frac{1}{\Omega},$$

con

$$\Omega = \frac{N!}{N_+!N_-!}.$$

## Ensemble canonico

A differenza di quanto avviene per il micro-canonico, in questo caso siamo interessati a fare inferenza della PDF degli stati microscopici avendo come vincolo solamente il valore medio della funzione costo (e assumendo che il suo valore esatto possa fluttuare). Il vincolo, dunque, non è più su  $E$  ma su  $U = \langle E \rangle$ . Applichiamo quindi il principio MaxEnt con i due vincoli

- $\sum_i p_i = 1$
- $\sum_i p_i E_i = U$ .

Per trovare la distribuzione di probabilità per l'insieme canonico imponiamo che l'entropia sia massima:

$$\frac{\partial S}{\partial p_i} + \lambda_1 + \lambda_2 E_i = 0,$$

dove abbiamo usato i moltiplicatori di Lagrange. Ricordando la definizione di entropia troviamo subito che

$$p_i = e^{-(1+\lambda_1+\lambda_2 E_i)}.$$

Imponendo il primo dei due vincoli si trova che

$$e^{-(1+\lambda_1)} \sum_i e^{-\lambda_2 E_i} = 1,$$

da cui

$$e^{-(1+\lambda_1)} = \frac{1}{\sum_i e^{-\lambda_2 E_i}} =: \frac{1}{Z},$$

dove abbiamo chiamato  $Z$  (funzione di partizione) la costante di normalizzazione. Imponendo ora il secondo vincolo, si ha invece

$$U = \frac{\sum_i E_i e^{-\lambda_2 E_i}}{Z},$$

cioè

$$p_i = \frac{1}{Z} e^{-\beta E_i} ,$$

che definisce la distribuzione di probabilità delle configurazioni per l'ensemble canonico, e dove abbiamo posto  $\lambda_2 = \beta$ . Si noti che, per il momento,  $\beta$  è solo un parametro che rappresenta la rigidità del vincolo, e che vorremo indentificare con la temperatura inversa tramite il parallelo con la termodinamica.

Definita la PDF, possiamo calcolare l'energia media, e esplorare il parallelo termodinamico delle espressioni che troviamo.

$$\langle E \rangle = \sum_i p_i E_i = \frac{1}{Z} \sum_i E_i e^{-\beta E_i} = -\frac{1}{Z} \frac{\partial}{\partial \beta} \sum_i e^{-\beta E_i} = -\frac{\partial}{\partial \beta} \log Z .$$

Questa espressione lega l'energia media alla funzione di partizione. Calcolando ora l'entropia,

$$S = -k_B \sum_i p_i \log p_i = k_B \sum_i \left( \frac{1}{Z} e^{-\beta E_i} \right) (\log Z + \beta E_i) = k_B \log Z + \langle E \rangle k_B \beta ,$$

cioè troviamo una parte "entalpica" dovuta all'energia media e una parte, espressa nei termini della funzione di partizione, legata alla PDF della random variabile che lega macrostato ai microstati.

Conoscendo la termodinamica, definiamo la funzione "energia libera"  $F$  dall'espressione  $Z =: e^{-\beta F}$ , che equivale a dire

$$\log Z = -\beta F = -\frac{F}{k_B T} .$$

Questa relazione può essere presa come definizione della temperatura. Sostituendo la relazione definitoria tra funzione di partizione e energia libera nell'espressione dell'entropia, si ottiene

$$S = -\frac{F}{T} + \frac{\langle E \rangle}{T} ,$$

che (identificando l'energia media con l'energia interna) corrisponde alla nota relazione termodinamica

$$F = \langle E \rangle - TS \equiv U - TS .$$

Inoltre ancora una volta possiamo considerare questa espressione come una definizione della temperatura

$$\frac{1}{T} := \frac{\partial S}{\partial \langle E \rangle} .$$

L'equilibrio è definito dalla condizione

$$\frac{\partial F}{\partial U} = 0 ,$$

e quindi (a causa del vincolo) non corrisponde al caso di entropia massima ma a quello di energia libera minima.

Si noti che non sempre il coefficiente  $\beta$  ha bisogno di una interpretazione fisica. Si pensi ad esempio ad un problema di ottimizzazione. Utilizzando MaxEnt possiamo trattarlo nel framework giustificato ensemble di Gibbs per sistemi fisici e non, nei quali però è necessario individuare quale sia la quantità che riveste il ruolo dell'energia, ovvero quale sia la funzione costo.

L'approccio MaxEnt può essere esteso anche al caso in cui ci siano molte osservabili con valori medi vincolati (si veda per es. De Martino e De Martino, Helyion 4 2018, e Banavar *et al.* J Phys Cond Mat 22 2010). In questo caso si avranno altrettanti moltiplicatori di Lagrange e più vincoli da rispettare, ma si ottiene comunque una distribuzione di probabilità esponenziale, in cui ci saranno in generale più temperature. Lo stesso vale anche nel caso in cui si abbia una sola osservabile di cui viene fissato non solo il valore medio, ma anche i momenti superiori; in questo caso si ottiene una distribuzione esponenziale con, all'esponente, un polinomio dell'energia (della variabile di cui si fissano i momenti).

Per concludere, a parte l'utilità pratica, l'interesse teorico del principio di massima entropia è da un lato di aiutare a fondare in modo agnostico la teoria degli ensemble (senza doverla cercare di "derivare" dalla meccanica classica o quantistica, si vedano le trattazioni classiche nei corsi precedenti di Fisica statistica, o i libri per esempio di Huang o di Peliti), e dall'altro giustifica l'uso della meccanica statistica per problemi non fisici. In generale il principio giustifica anche l'utilità di strumenti concettuali "termodinamici" e interpretazioni fisiche in problemi di inferenza.

## 4 Grandi deviazioni

### Motivazione del problema

Il teorema del limite centrale può essere visto come una caratterizzazione delle fluttuazioni "tipiche" di un sistema, che nel caso di iid a varianza finita sono di ordine  $O(N^{1/2})$  attorno alla media  $\mu_1 N$ . La teoria delle "grandi deviazioni" (large deviations) studia distribuzioni di probabilità lontano dal valor medio, cioè

dal punto di vista fisico il comportamento di grandi (e rare) fluttuazioni di un sistema, che possono avere interesse in vari contesti. Come al solito, la teoria della probabilità studia primariamente il caso iid, e l'idea è che in meccanica statistica principi simili possano valere più in generale.

Per motivare il problema in termini di teoria della probabilità consideriamo una random variabile Bernoulli  $X$  che può assumere valori

$$X = \begin{cases} 1 & p \\ 0 & q = (1 - p) \end{cases} .$$

Per un solo tiro la variabile Bernoulli ha valor medio  $\mu_1 = p$  e varianza  $\sigma^2 = p(1 - p)$ . Supponiamo ora di avere una sequenza di  $N$  lanci, ovvero  $N$  variabili aleatorie  $X = (X_1, X_2, \dots, X_N)$  e vogliamo vedere cosa succede alla somma (una RV binomiale)  $S = \sum_i X_i$ . Come vedremo più in dettaglio, possiamo vedere questo come un sistema di spin non interagenti in un campo esterno.

Sappiamo che la probabilità che la somma  $S_N$  sia uguale a  $k$  è specificata dalla distribuzione binomiale

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

Allo stesso tempo, dato che le  $X_i$  sono RV iid a varianza finita, vale il teorema del limite centrale (CLT), che fornisce solo una *approssimazione gaussiana* della distribuzione vera (come sappiamo la gaussiana è un limite per la distribuzione binomiale solamente per  $N \gg 1$ ). L'approssimazione gaussiana della distribuzione binomiale per il teorema del limite centrale sarà:

$$P^G(k, N) = \frac{1}{\sqrt{2\pi p(1-p)}} e^{-\frac{(k-pN)^2}{2p(1-p)N}}$$

Consideriamo ora il caso di "grandi fluttuazioni" in cui  $k \simeq N$ . Sapendo la distribuzione vera, possiamo chiederci come si comporta l'approssimazione del teorema del limite centrale nello stimare la probabilità di questo evento. Per la distribuzione binomiale,

$$P(N, N) = p^N = e^{-N \log\left(\frac{1}{p}\right)} ,$$

mentre per l'approssimazione gaussiana del CLT possiamo scrivere,

$$P^G(N, N) \sim e^{-\frac{(1-p)^2 N^2}{2p(1-p)N}} \sim e^{-\frac{1-p}{2p} N} .$$

Confrontando le due espressioni si vede subito che si commette un errore esponenzialmente grande nello stimare grandi deviazioni con la distribuzione approssimata.

Possiamo quantificare l'errore  $\delta$  che si commette nello stimare il caso estremo con il teorema del limite centrale come lo scarto tra gli esponenti,

$$\delta \simeq e^{\left(\frac{1-p}{2p} - \log \frac{1}{p}\right)N}$$

A seconda della probabilità  $p$  della singola variabile  $X_i$  abbiamo che l'approssimazione gaussiana può risultare sia una sovrastima che una sottostima rispetto alla probabilità ottenuta considerando la distribuzione binomiale. Infatti, vale

$$\begin{aligned} p \rightarrow 0 \quad \log \frac{1}{p} &\ll \frac{1-p}{2p} & P^G(N) &\gg P(N) \\ p \rightarrow 1 \quad \log \frac{1}{p} &> \frac{1-p}{2p} & P^G(N) &\ll P(N) . \end{aligned}$$

Questo esempio mostra chiaramente come il teorema del limite centrale dia solamente un'approssimazione della distribuzione, riuscendo a catturare fluttuazioni tipiche di ordine  $O(N^{1/2})$ , fallendo nel ricostruire il caso di fluttuazioni grandi. I teoremi di grandi deviazioni forniscono un'approssimazione generica per le fluttuazioni estreme, molto utile nei casi in cui la distribuzione vera non è nota (cioè quasi sempre), come nell'esempio delle variabili Bernoulli ora analizzato.

## Grandi deviazioni per somme di variabili indipendenti e analogia con la meccanica statistica

Chiarita la motivazione, enunciamo i risultati principali della teoria. Consideriamo ancora  $S_N = \sum_i^N X_i$ , con  $X_i$  random variabili iid. Allora vale la seguente approssimazione esponenziale della probabilità di fluttuazioni ordine  $O(N)$ ,

$$P(S_N \approx Ns) \sim e^{Ng(s)}, \quad N \rightarrow \infty$$

dove  $g(s)$  è chiamata "rate function" e possiede le seguenti proprietà,

$$g(s) \leq 0, \quad g(\mu_1) = 0 .$$

Ossia  $g(s)$  è una funzione non crescente di  $s$  e si annulla nella media.

Cerchiamo di calcolare la rate function  $g(s)$ . Definiamo la funzione generatrice

$$Z(\beta) := G_X(i\beta) = \int e^{-\beta x} P(X \in dx) = \int e^{-\beta x} f(x) dx ,$$

che è l'analogo di una funzione di partizione in meccanica statistica, e dove  $f(x)$  è la PDF delle RV  $X_i$ . Siamo interessati a valutare

$$\tilde{d}_N(S_N) = \langle e^{-\beta S_N} \rangle = \langle e^{-\beta \sum_i^N X_i} \rangle \equiv \tilde{f}(\beta)^N = Z(\beta)^N$$

dove  $d_N(S_N)$  è la PDF di  $S_N$ , la somma delle variabili  $X_i$ .

Per fare ciò usiamo l'ansatz di grandi deviazioni

$$\tilde{d}_N(S_N) \sim e^{Ng(\frac{S}{N})} = e^{Ng(s)},$$

ottenendo l'espressione

$$\langle e^{-\beta S_N} \rangle = \int ds e^{-\beta Ns} e^{Ng(s)} = \int ds e^{N(g(s) - \beta s)}.$$

Il valore di questo integrale, per  $N \gg 1$  è dominato dal valore dell'integranda valutata nel suo massimo. Questa approssimazione è nota come metodo di "punto sella" (saddle point). Quindi dobbiamo trovare il punto stazionario:

$$s^*(\beta) : \frac{d}{ds}(g(s) - \beta s)|_{s^*} = 0, \quad g'(s^*) = \beta.$$

Troviamo quindi che

$$g(s) = \log z(\beta(s)) + s\beta(s)$$

dove  $\beta(s) = g'(s)$ . Da qui segue che:

$$\beta = \frac{dg(s)}{ds} = \frac{d \log z}{d\beta} \frac{d\beta}{ds} + s \frac{d\beta}{ds} \Rightarrow s(\beta) = -\frac{d \log z}{d\beta}$$

Possiamo notare che ogni volta si cerca la distribuzione di grandi fluttuazioni in sistemi di random variabili (indipendenti, per la teoria), emerge una struttura di distribuzione esponenziale, trasformata di Legendre e funzionali termodinamici simile a quella della meccanica statistica. Si può allora pensare la meccanica statistica come un modo di cercare *rate function* per eventi rari anche in casi in cui le random variabili non sono indipendenti e identicamente distribuite.

### Fluttuazioni di energia in ensemble canonico come problema di grandi deviazioni.

Per capire meglio come questo sia connesso con la meccanica statistica consideriamo come esempio un problema di fluttuazioni di energia in ensemble canonico. La distribuzione di probabilità dell'energia si può ottenere integrando da quella delle configurazioni

$$P(E) = \sum_{E(C) \in [E, E+\delta E]} P(C) = \frac{\Omega(E)e^{-\beta E}}{Z},$$

che possiamo riscrivere come  $P(E) = \exp(-\beta F(E))/Z$ , con  $F(E) := E - TS(E)$ . Definiamo ora una energia interna per grado di libertà  $u = E/N$  e analogamente

un'entropia per grado di libertà  $s = S/N$ . Possiamo scrivere che la distribuzione di  $u$  è

$$p(u) = \frac{1}{Z_N} e^{-N\beta[u - Ts(u)]},$$

che ha la struttura di una distribuzione da grandi deviazioni. Inoltre possiamo dire che

$$Z_N = e^{-N\beta f(T)}$$

con

$$f(T) = \min_u \{u - Ts\}$$

Questa energia libera per grado di libertà  $f(T)$  è la trasformata di Legendre (a meno di un segno) di  $u$ . Si noti, tra l'altro, che la stazionarietà di  $f$  (è valutata nel minimo di  $u - Ts$  rispetto a  $u$ ) implica subito che

$$\frac{\partial f}{\partial u} = 0,$$

da cui

$$1 - T \frac{\partial s}{\partial u} = 0,$$

e infine

$$\frac{1}{T} = \frac{\partial s}{\partial u}.$$

Nel linguaggio delle grandi deviazioni questo significa dire che la "rate function"  $I(u)$  per la nostra PDF  $p(u)$  sia la funzione

$$g(u) = \beta[u - Ts - f].$$

## Stima di grandi deviazioni delle fluttuazioni di variabili binomiali e entropia

Tornando al nostro esempio iniziale basato su una random variabile binomiale, somma di Bernoulli, possiamo chiederci quanto sia buona la stima della probabilità di eventi estremi fornita dalla teoria delle grandi deviazioni. Applicando la teoria, troviamo che la funzione di partizione è

$$Z(\beta) = 1 - p + pe^{-\beta},$$

e quindi

$$s(\beta) = -\frac{d \log Z}{d\beta} = -\frac{d}{d\beta} \log(1 - p + pe^{-\beta}) = \frac{pe^{-\beta}}{1 - p + pe^{-\beta}}.$$

Invertendo, otteniamo

$$\beta(s) = \log \frac{s(1-p)}{p(1-s)},$$

per cui la rate function è

$$g(s) = s \log(p) + (1-s) \log(1-p) - s \log s - (1-s) \log(1-s).$$

Utilizzando quindi la struttura delle grandi deviazioni abbiamo ottenuto l'espressione giusta dell'entropia per la random variabile binomiale (cosa che in questo caso si può ricavare direttamente dalla distribuzione esatta, usando l'approssimazione di Stirling, ma che in generale non è possibile fare), a differenza del CLT, che restituisce una entropia sbagliata per grandi deviazioni.

## Deviazioni dall'approssimazione del teorema del limite centrale

Per concludere, vogliamo chiederci quali sono i limiti di validità del CLT, nel senso di quanto devono essere grandi le fluttuazioni affinché l'approssimazione non sia buona. Per trovare le deviazioni dal limite centrale possiamo sviluppare  $g(s)$  attorno alla media  $\mu_1$ . Sappiamo che  $g(\mu_1) = 0$ , e che  $g'(s) = \beta$ , da cui

$$g(s) = -\frac{1}{2}g''(\mu_1)(s - \mu_1)^2 - \frac{1}{6}g'''(\mu_1)(s - \mu_1)^3 + \dots,$$

dove abbiamo la presenza della derivata terza se la funzione non è simmetrica. Quindi la struttura data dalla distribuzione di probabilità del teorema di grandi deviazioni è

$$P(S_N \simeq Ns) \sim e^{Ng(s)} \simeq e^{-\frac{1}{2}Ng''(\mu_1)(s-\mu_1)^2} = e^{-\frac{(S-N\mu_1)^2}{2\sigma^2N}}$$

dove abbiamo usato il fatto che  $\sigma^2 = (g''(\mu))^{-1}$ . Questa espressione equivale al teorema del limite centrale. Il termine di ordine superiore, invece, fornisce una correzione alla distribuzione gaussiana. Per distribuzioni asimmetriche vale quindi che la prima correzione è

$$\frac{P(S_N \simeq S)}{P^G(S_N \simeq S)} \simeq e^{-\frac{1}{6}g'''(\mu_1)\frac{(S-N\mu_1)^3}{N^2}}.$$

Possiamo quindi dire che l'approssimazione gaussiana della distribuzione data dal teorema del limite centrale è valida per deviazioni per cui questa correzione è trascurabile, quindi quando

$$|S_N - N\mu_1|^3 \ll N^2 \Rightarrow |S_N - N\mu_1| \ll N^{2/3}.$$

Nel caso di una distribuzione simmetrica, il termine con la derivata terza di  $g$  è nullo, e conta quello di quarto ordine. In questo caso l'approssimazione gaussiana della PDF data dal teorema del limite centrale è valida per

$$|S_N - N\mu_1|^4 \ll N^3 \Rightarrow |S_N - N\mu_1| \ll N^{3/4}.$$

La teoria delle grandi deviazioni stabilisce quindi quanto devono essere grandi le fluttuazioni di una variabile macroscopica per vedere code diverse da quelle gaussiane.

## ”Metodi pratici” di calcolo usati spesso in meccanica statistica

Analizziamo ora alcune comuni procedure che permettono di passare da variabili esponenziali a variabili estensive.

### 4.1 Saddle point integrale

Supponiamo che  $I$  sia un integrale della forma:

$$I = \int_{-\infty}^{\infty} e^{N\Phi(x)} dx. \quad (34)$$

In genere non esiste una primitiva esprimibile sotto forma di funzioni elementari, pertanto mostriamo una procedura la procedura del metodo di *saddle point* (o *di punto-sella*) che permette di determinare un'approssimazione del valore di  $I$  raffinata a piacere. Nello specifico, l'idea di base di questo metodo consiste fondamentalmente in uno sviluppo di  $\Phi$  (almeno) al secondo ordine in serie di Taylor attorno al valore  $x_{\max}$ :

$$\Phi(x) = \Phi(x_{\max}) + \frac{1}{2}\Phi''(x_{\max})(x - x_{\max})^2 + \mathcal{O}(x - x_{\max})^3, \quad (35)$$

dove sottolineiamo che

- si assume  $N \gg 1$  ( $N \rightarrow \infty$  nel limite termodinamico);
- la derivata prima si annulla nel punto di massimo, ossia  $\Phi'(x_{\max}) = 0$ ;
- poiché l'espansione è centrata nel massimo di  $\Phi$ , allora  $\Phi''(x_{\max}) < 0$ .

A questo punto, sostituendo l'espansione (35) di  $\Phi$  nell'integrale (34) si ottiene

$$I \approx \int_{-\infty}^{\infty} e^{N(\Phi(x_{\max}) + \frac{1}{2}\Phi''(x_{\max})(x - x_{\max})^2)} dx = \sqrt{\frac{2\pi}{-N\Phi''(x_{\max})}} e^{N(\Phi(x_{\max}))} \quad (36)$$

dove nell'ultima eguaglianza è stato valutato l'integrale gaussiano. Chiaramente, l'approssimazione qui proposta diventa tanto più accurata quanto più cresce  $N$ .

## Metodo di punto sella

Esponiamo qui più in dettaglio il metodo dipunto-sella. Consideriamo l'integrale

$$I(N) = \int_{x_1}^{x_2} dx f(x) e^{Ng(x)} \quad (37)$$

dove supponiamo che  $f, g: \mathbb{R} \rightarrow \mathbb{R}$  (per comodità consideriamo funzioni lisce) e  $N > 0$  reale. Sia  $h_N(x) := f(x)e^{Ng(x)}$  la funzione integranda. Con un approccio naïve, possiamo immaginare che per  $N$  sufficientemente grande l'integranda  $h_N$  mostri un picco stretto e acuto, tale per cui l'integrale  $I(N)$  sia totalmente dominato dall'area sottesa al picco. Siano  $x_0$  e  $x_N$  rispettivamente il massimo assoluto della funzione  $g$  e dell'integranda  $h_N$ , con  $x_0, x_N \in [x_1, x_2]$  (in questo modo rimane ben definita la successione  $\{x_N\} \subset [x_1, x_2]$ ). Supponiamo inoltre che  $f(x_0) \neq 0$ . È chiaro che  $\forall N > 0$  vale l'equazione:

$$h'_N(x_N) = [f'(x_N) + Nf(x_N)g'(x_N)]e^{Ng(x_N)} = 0.$$

Questo implica che  $\forall N > 0$

$$f'(x_N) + Nf(x_N)g'(x_N) = 0,$$

da cui

$$g'(x_N) = -\frac{f'(x_N)}{Nf(x_N)} = \mathcal{O}\left(\frac{1}{N}\right).$$

Quindi,  $g'(x_N) = 0 = g'(x_0)$  per  $N \rightarrow \infty$ , ossia  $x_N \rightarrow x_0$ . Assumiamo pertanto che  $x_0$  sia il massimo assoluto di  $h_N$  nel limite in cui  $N$  sia sufficientemente grande.

A questo punto, cambiamo la variabile di integrazione in  $x = x_0 + y/\sqrt{N}$  ed espandiamo  $Ng(x)$  in potenze di  $y$ :

$$Ng(x) = Ng(x_0) + \frac{1}{2}g''(x_0)y^2 + \frac{g'''(x_0)}{6\sqrt{N}}y^3 + \dots$$

Trattando questa espansione come un'espansione in potenze di  $1/\sqrt{N}$ , otteniamo

$$e^{Ng(x)} = e^{Ng(x_0) + g''(x_0)y^2/2} \left( 1 + \frac{g'''(x_0)}{6\sqrt{N}}y^3 + \mathcal{O}\left(\frac{1}{N}\right) \right). \quad (38)$$

Seguendo lo stesso principio, abbiamo:

$$f(x) = f(x_0) \cdot \left( 1 + \frac{f'(x_0)}{f(x_0)\sqrt{N}}y + \mathcal{O}\left(\frac{1}{N}\right) \right). \quad (39)$$

A questo punto, sostituiamo le espansioni (38) e (39) nella (37):

$$\begin{aligned}
 I(N) &= \int_{y_1}^{y_2} \frac{dy}{\sqrt{N}} f(x_0) \cdot \left( 1 + \frac{f'(x_0)}{f(x_0)\sqrt{N}} y + \mathcal{O}\left(\frac{1}{N}\right) \right) \\
 &\quad \cdot e^{Ng(x_0) + g''(x_0)y^2/2} \left( 1 + \frac{g'''(x_0)}{6\sqrt{N}} y^3 + \mathcal{O}\left(\frac{1}{N}\right) \right) = \\
 &= \frac{f(x_0)e^{Ng(x_0)}}{\sqrt{N}} \int_{y_1}^{y_2} dy e^{g''(x_0)y^2/2} \left( 1 + \frac{f'(x_0)}{f(x_0)\sqrt{N}} y + \frac{g'''(x_0)}{6\sqrt{N}} y^3 + \mathcal{O}\left(\frac{1}{N}\right) \right) \approx \\
 &\approx \frac{f(x_0)e^{Ng(x_0)}}{\sqrt{N}} \int_{-\infty}^{+\infty} dy e^{g''(x_0)y^2/2} \left( 1 + \frac{f'(x_0)}{f(x_0)\sqrt{N}} y + \frac{g'''(x_0)}{6\sqrt{N}} y^3 + \mathcal{O}\left(\frac{1}{N}\right) \right) = \\
 &= \frac{f(x_0)e^{Ng(x_0)}}{\sqrt{N}} \int_{-\infty}^{+\infty} dy e^{g''(x_0)y^2/2} \left( 1 + \mathcal{O}\left(\frac{1}{N}\right) \right) = \\
 &= \sqrt{\frac{2\pi}{-Ng''(x_0)}} f(x_0) e^{Ng(x_0)} \left( 1 + \mathcal{O}\left(\frac{1}{N}\right) \right),
 \end{aligned}$$

dove è importante sottolineare che

- $x_1 = x_0 + y_1/\sqrt{N}$ , da cui chiaramente  $y_1 = -\sqrt{N}(x_0 - x_1) \rightarrow -\infty$  per  $N \rightarrow \infty$ . In modo analogo si verifica che  $y_2 \rightarrow +\infty$ ;
- $x_0$  è un punto di massimo per  $g$ , pertanto  $g''(x_0) < 0$ : stiamo quindi trattando un *integrale gaussiano*.

Sintetizzando il risultato di cui sopra, abbiamo che

$$I(N) \approx \sqrt{\frac{2\pi}{-Ng''(x_0)}} f(x_0) e^{Ng(x_0)}. \quad (40)$$

In prima approssimazione possiamo quindi assumere che l'integrale  $I(N)$  converga (a meno di una costante moltiplicativa) alla corrispondente integranda valutata nel suo punto di massimo (ossia  $I(N) \propto h_N(x_0)$ ).

## Approssimazione di una somma col suo termine massimo

Un analogo di un metodo di punto sella per somme è l'approssimazione della somma con il suo termine massimo. Per capire come questa approssimazione sia giustificata, consideriamo la somma  $S = \sum_{i=1}^N x_i$ , con  $x_i = e^{N\Phi_i} > 0$ . In modo del tutto analogo alla discussione di cui sopra, per grandi  $N$  si ha

$$S \approx x_{\max} = e^{N\Phi_{\max}}. \quad (41)$$

Dal momento che  $x_{\max} \leq S \leq Nx_{\max}$ , allora:

$$\frac{\log(x_{\max})}{N} \leq \frac{\log(S)}{N} \leq \frac{\log(x_{\max})}{N} + \frac{\log(N)}{N}. \quad (42)$$

Ora, considerando che  $N \gg 1$ , ne consegue che  $\frac{\log(N)}{N} \ll 1$ , per cui:

$$\frac{\log(S)}{N} \approx \frac{\log(x_{\max})}{N} = \Phi_{\max} \Rightarrow \log(S) \approx N\Phi_{\max} \Rightarrow S \approx x_{\max} \quad (43)$$

## Approssimazione di Stirling

Un'altra tecnica utile per trattare quantità originate da conteggi combinatori è l'approssimazione di Stirling (valida per grandi  $N$ ),

$$N! \approx \sqrt{2\pi N} \left(\frac{N}{e}\right)^N \quad (44)$$

da cui:

$$\log(N!) \approx N \log(N). \quad (45)$$

Questa approssimazione può essere ricavata da sviluppi di Taylor di  $\log(N!) \simeq \int_1^N \log(x) dx$ , oppure applicando l'approssimazione di saddle point alla funzione gamma di Eulero. Ricordiamo la definizione di questa funzione,

$$\Gamma(z) := \int_0^{+\infty} x^{z-1} e^{-x} dx.$$

Vale

$$N! = \int_0^{+\infty} x^N e^{-x} dx$$

per cui con la sostituzione  $y = x/N$  e un saddle point si ottiene la tesi.

## 5 Parte II. Metodi di Campo Medio

Questa parte delle note tratta varie accezioni e tecniche che vanno sotto il nome di "metodi di campo medio", usando come esempio sistemi magnetici "tipo Ising", cioè da punto di vista probabilistico sistemi di variabili aleatorie Bernoulli o "Booleane". Questi sistemi sono nati in fisica per descrivere ferromagneti, ma sono generici. L'interpretazione "come spin" di variabili bistabili permette di utilizzare analogie fisiche/termodinamiche per ragionare su problemi di vario tipo, come coding, ottimizzazione combinatoria, reti neurali, reti di interazione / regolazione, dinamica di opinioni etc.

### Sistemi magnetici.

#### Spin singolo.

Consideriamo un sistema con hamiltoniana  $H = -h\sigma$ , dove  $h$  è un campo esterno e  $\sigma = \pm 1$  è una variabile "di spin tipo Ising" (quindi in questo caso il sistema è un'unica variabile di spin). La probabilità che questo si trovi nello stato  $\sigma = 1$  è

$$\eta_+ = P(\sigma = 1) = \frac{e^{-\beta h}}{e^{-\beta h} + e^{\beta h}},$$

cui corrisponde il momento  $\mu = \langle \sigma \rangle = \tanh(\beta h)$ . Come possiamo notare, l'hamiltoniana  $H$  è una funzione lineare dello spin  $\sigma$ . Questo perché il sistema può assumere due soli stati energetici a seconda del valore assegnato a  $\sigma$ . Si noti che in generale (dato che  $\sigma^2 = 1$  come tutte le potenze successive), qualunque funzione  $f(\sigma)$  di uno spin ha una dipendenza al più lineare da  $\sigma$ , e quindi si scrive come  $f(\sigma) = A + B\sigma$ .

#### Sistemi di più spin.

Consideriamo ora un sistema di due spin  $(\sigma_1, \sigma_2)$ . La configurazione può assumere quattro valori diversi, ossia  $(1, 1)$ ,  $(1, -1)$ ,  $(-1, 1)$  e  $(-1, -1)$ . Ci aspettiamo dunque che anche una funzione generica  $f(\sigma_1, \sigma_2)$  possa assumere quattro valori diversi. Ancora una volta essa è una forma bilineare scrivere nella forma  $f(\sigma_1, \sigma_2) = A + B\sigma_1 + C\sigma_2 + D\sigma_1\sigma_2$ . Allo stesso modo, una generica funzione di  $N$  spin ha la forma

$$f(\sigma_1, \dots, \sigma_N) = A^{(0)} + A_1^{(1)}\sigma_1 + \dots + A_N^{(1)}\sigma_N + \dots + A^{(n)}(\sigma_1\sigma_2 \dots \sigma_N),$$

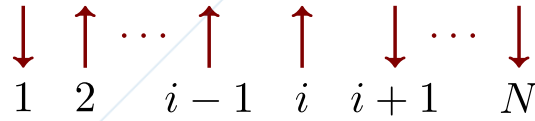


Figura 1: Catena monodimensionale di spin di tipo Ising.

dove ogni contributo  $\mathbf{A}^{(i)}$  contiene prodotti di  $i$  spin (cioè “interazioni” a  $i$  corpi tra gli spin).

Consideriamo ora un sistema caratterizzato dalla dipendenza di  $N$  spin diversi che interagiscono solo a coppie (interazioni binarie). L’hamiltoniana più generale ha la forma

$$H = - \sum_{ij} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (46)$$

dove  $J_{ij}$  è l’interazione tra gli spin  $\sigma_i$  e  $\sigma_j$  e  $\mathbf{h}$  è sempre il campo esterno (che in questo caso può essere diverso per ogni spin). Se indichiamo il vettore degli spin con  $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$ , avremo

$$P(\boldsymbol{\sigma}) \sim e^{-\beta H(\boldsymbol{\sigma})} \quad (47)$$

$$Z = \sum_{\boldsymbol{\sigma}} e^{-\beta H(\boldsymbol{\sigma})} \quad (48)$$

dove  $Z$  è la funzione di partizione.

## Mean-field come fattorizzazione della misura di probabilità

Consideriamo una catena monodimensionale come quella in Figura in cui ogni spin  $\sigma_i$  interagisce con gli spin  $\sigma_{i-1}$  e  $\sigma_{i+1}$  (assumiamo come condizione di periodicità che lo spin  $\sigma_N$  interagisca con  $\sigma_{N-1}$  e  $\sigma_1$ ). Chiamiamo  $J$  l’interazione tra gli spin e sia  $\mathbf{h}$  il campo esterno. L’hamiltoniana sarà quindi

$$H = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - \sum_i h_i \sigma_i$$

e, in accordo con l’Eq. (48) possiamo scrivere la funzione di partizione come

$$Z = \sum_{\boldsymbol{\sigma}} e^{\beta(J\sigma_1\sigma_2+h_1\sigma_1)} e^{\beta(J\sigma_2\sigma_3+h_2\sigma_3)} \dots e^{\beta(J\sigma_N\sigma_1+h_N\sigma_N)}.$$

È subito chiaro che anche per un modello semplice come quello appena proposto la funzione  $Z$  non è fattorizzabile nelle singole probabilità  $P_i$  e questo comporta

difficoltà dal punto di vista dei conti. Nel caso di un modello di Ising monodimensionale, o in due dimensioni senza campo esterno, il metodo della “matrice di trasferimento” permette di trovare soluzioni esatte.

In generale, in quasi tutti i casi interessanti per applicazioni contemporanee, abbiamo bisogno di introdurre approssimazioni che semplifichino il modello e lo rendano gestibile nella computazione. L’ipotesi di *campo medio* introduce come ansatz proprio il fatto che la probabilità  $P(\boldsymbol{\sigma})$  sia fattorizzabile, ossia che si possa scrivere nella forma

$$P_{\text{MF}}(\boldsymbol{\sigma}) = \prod_i^N P_i(\sigma_i), \quad \text{dove } P_i(\sigma_i) = \begin{cases} \eta_+, & \sigma_i = 1 \\ \eta_-, & \sigma_i = -1 \end{cases}. \quad (49)$$

Si noti come l’ipotesi di fattorizzazione sia la stessa per sistemi che hanno dimensionalità diverse, e sarebbe la stessa per topologie di interazione con connettività arbitrarie, quindi in qualche modo stiamo facendo predizioni che “trascurano” aspetti legati alla dimensionalità del sistema, che può entrare solo come fattore legato al numero delle interazioni (che varia con la dimensionalità).

Sotto l’ipotesi di campo medio, la magnetizzazione per sito prende la semplice forma  $m_i = \eta_+ - \eta_-$ , dove chiaramente  $\eta_+ + \eta_- = 1$ . Tramite semplice algebra si ricava la relazione inversa

$$\eta_+ = \frac{1 + m_i}{2}$$

$$\eta_- = \frac{1 - m_i}{2}.$$

Si noti come la magnetizzazione di uno spin si possa vedere semplicemente come una riparametrizzazione del suo peso (la sua PDF) come variabile Bernoulliana.

Nell’ipotesi di campo medio troviamo inoltre che

$$\langle f_i(\sigma_i) \rangle = \eta_+ f_i(1) + \eta_- f_i(-1)$$

$$\langle f_i(\sigma_i) f_j(\sigma_j) \rangle = \langle f_i(\sigma_i) \rangle \langle f_j(\sigma_j) \rangle,$$

dove si osserva che le correlazioni sono fattorizzate, in virtù dell’assunzione di indipendenza (Eq. 49).

Conoscendo la magnetizzazione  $m_i = \langle \sigma_i \rangle$ , si ottengono tutte le informazioni termodinamiche sul sistema. Infatti, è possibile calcolare il valore di aspettazione dell’hamiltoniana, che corrisponde a

$$\langle H \rangle = J \sum_{\langle ij \rangle} m_i m_j - h \sum_i m_i$$

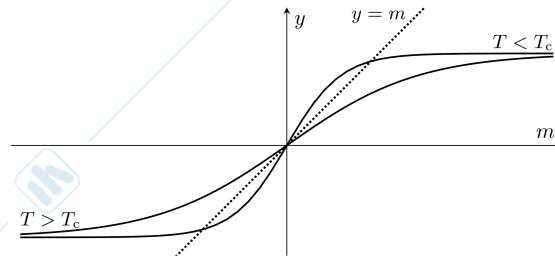


Figura 2: Soluzione grafica dell'equazione di stato di campo medio, in cui sono rappresentate le curve  $y = m$  e  $y = \tanh(4DJk_B^{-1}T^{-1}m)$ . Si osservi che la soluzione  $m = 0$  è sempre presente, mentre le due soluzioni  $m \neq 0$  esistono solo se  $4DJk_B^{-1}T^{-1} > 1$ , ossia solo se  $T < T_c$ .

e l'entropia del sistema

$$\frac{S(P)}{k_B} = -\langle \log P \rangle = -\sum_i \langle \log P_i(\sigma_i) \rangle = -\sum_i P_i \log P_i,$$

da cui

$$\begin{aligned} S(m) &= -\frac{1+m}{2} \log\left(\frac{1+m}{2}\right) - \frac{1-m}{2} \log\left(\frac{1-m}{2}\right) = \\ &= \frac{1}{2} \log\left(\frac{1-m^2}{4}\right) + \frac{m}{2} \log\left(\frac{1-m}{1+m}\right). \end{aligned}$$

## Termodinamica del modello di Ising in campo medio.

Calcolata l'entropia, possiamo accedere a tutta la termodinamica del sistema (cioè - per sistemi generali - il comportamento tipico di sistemi grandi a livelli di rumore fissati) sotto l'ipotesi di campo medio. Denominata con  $D$  la dimensione del sistema, il numero dei vicini di ogni spin è  $2D$ . Nota l'entropia, l'energia libera si scrive

$$F = -J2DNm^2 - hmN + \frac{k_B T}{2} N \left[ \log\left(\frac{1-m^2}{4}\right) + m \log\left(\frac{1+m}{1-m}\right) \right] = NF_1,$$

dove  $F_1$  corrisponde all'energia libera per singola variabile di spin. Se imponiamo che la derivata prima di  $F_1$  rispetto alla magnetizzazione sia nulla, otteniamo

$$\frac{\partial F_1}{\partial m} = -4DJm - hN + \frac{k_B T}{2} \log\left(\frac{1+m}{1-m}\right) = 0.$$

Assumendo che il campo esterno sia nullo ( $h = 0$ ) si ricava l'equazione per la magnetizzazione

$$m = \tanh\left(\frac{4DJ}{k_B T} m\right). \quad (50)$$

Studiando graficamente questa equazione si vede il cambiamento qualitativo tra due soluzioni di magnetizzazione opposta e una di magnetizzazione nulla ("transizione di fase") dato dalla condizione  $4DJ/k_B T = 1$ . Se  $k_B T > 4DJ$ ,  $m = 0$  ("fase paramagnetica"), altrimenti  $m = \pm m_F$  ("fase ferromagnetica"), dove  $m_F(4DJ/k_B T)$  è l'intersezione tra la bisettrice e la tangente iperbolica. Possiamo chiamare temperatura critica il valore di transizione  $T_c = 4DJ/k_b$ .

Per capire il comportamento del sistema, possiamo tornare indietro, considerando l'espressione dell'energia libera per spin, e studiarla attorno alla temperatura critica,

$$\begin{aligned} F_1(m) &= -J2Dm^2 - hm - k_B T S(m) \simeq -J2Dm^2 - hm - k_B T \left[ \frac{m^2}{2} + \frac{m^4}{12} + \dots \right] \simeq \\ &\simeq k_B T \frac{m^4}{12} + \left[ \frac{k_B T}{2} - 2DJ \right] m^2 - hm, \end{aligned}$$

dove il termine del secondo ordine è nullo alla temperatura critica. In altre parole la transizione di fase è caratterizzata da una singolarità nella energia libera.

Per caratterizzare il comportamento attorno a questa singolarità (i cosiddetti "esponenti critici") sviluppiamo usando il teorema della funzione implicita. Possiamo scrivere

$$0 = \frac{\partial F(m, T, h)}{\partial m} = G(m, T, h)$$

da cui, tramite questa funzione implicita, possiamo ricavare la magnetizzazione  $m$ . Sviluppiamo vicino a  $T_C$ :

$$m(T_C + \delta T) = m(T_C) + \frac{\partial m}{\partial T} \delta T + \dots,$$

dove  $m(T_C) = 0$  e il termine di derivata prima non può essere invertito usando il teorema di Dini. Bisogna utilizzare i termini di ordine superiore

$$0 = \delta m \delta T + 2T_C (\delta m)^3,$$

da cui

$$(\delta m)^2 \propto -\frac{\delta T}{T_C} \Rightarrow \delta m \propto \left(\frac{\delta T}{T_C}\right)^{1/2}.$$

In altre parole, l'esponente critico  $\beta = 1/2$  caratterizza in modo univoco il comportamento di questo modello vicino alla temperatura critica.

La natura singolare delle transizioni di fase e la possibilità di caratterizzare il comportamento di transizioni del secondo ordine attraverso sviluppi di questo tipo hanno dato luogo alla "ipotesi di universalità", secondo la quale le transizioni di fase possono essere classificate in un numero abbastanza piccolo di classi caratterizzate da dimensionalità del sistema, simmetria del parametro d'ordine e valori degli esponenti critici. Dalla conoscenza degli esponenti critici possiamo quindi identificare la "classe di universalità" di un modello. L'invarianza di scala attorno alle transizioni di fase e l'analisi del comportamento di modelli sotto riscaldamento ("gruppo di rinormalizzazione") rinforzano questa ipotesi. La potenza concettuale di questi strumenti è quella di giustificare l'uso di modelli molto semplici per descrivere fenomeni complessi. Per informazioni generali su esponenti critici e gruppo di rinormalizzazione rimandiamo il lettore ai libri di Peliti (Cap. 5-6) e Sethna (Cap. 12) o Kadanoff.

## Ferromagnete di Curie-Weiss , "modello di campo medio"

Prima di proseguire con altre accezioni del concetto di campo medio come metodo (soffisticato) di approssimazione, vediamo ora quello che viene detto "modello di campo medio" (di Ising), anche detto modello mean-field di Curie e Weiss (che sono i primi ad averlo ideato). Si tratta di un modello in cui ogni spin interagisce con tutti gli altri "fully connected", per cui l'hamiltoniana ha la forma

$$H = -\frac{J}{2N} \sum_{ij} \sigma_i \sigma_j - h \sum_i \sigma_i , \quad (51)$$

dove notiamo che il fattore  $J/2N$  è necessario affinché l'energia sia estensiva ( $O(N)$ ), dato che ci sono  $O(N^2)$  interazioni (e quindi devono essere  $O(1/N)$ ).

Perché si parla di campo medio in questo contesto? Come vedremo il comportamento efficace di questo modello dal punto di vista termodinamico coincide con quello dell'approssimazione di campo medio. Il fatto che questo succeda è del tutto non banale e vuol dire che una grande quantità di gradi di libertà interagenti (cioè di variabili correlate), in qualche senso si comportano come se fossero descritti da distribuzioni fattorizzate, cioè come se ci fosse una "indipendenza emergente" di tanti gradi di libertà interagenti.

Per capirlo scriviamo la funzione di partizione, data una configurazione  $\{\sigma\} = (\sigma_1, \dots, \sigma_N)$

$$Z = \sum_{\{\sigma\}} e^{-\beta H(\{\sigma\})} = \sum_{\{\sigma\}} e^{\frac{\beta J}{2N} (\sum_{i=1}^N \sigma_i)^2 + \beta h (\sum_{i=1}^N \sigma_i)} .$$

È semplice verificare che questa può essere scritta come

$$Z = \prod_{i=1}^N \left( \sum_{\sigma_i = \pm 1} e^{-\beta H_i(\sigma_i)} \right) ,$$

dove  $H_i(\sigma_i) = \frac{J}{2N} \sigma_i \sum_j \sigma_j - h \sigma_i$ .

Per calcolare questa quantità possiamo utilizzare proprietà dell'integrale gaussiano. Infatti vale

$$e^{\frac{x^2}{2a}} = \sqrt{\frac{a}{2\pi}} \int_{\mathbb{R}} dm e^{-\frac{1}{2}am^2 + mx} ,$$

e nel nostro caso  $a = \beta J N$  e  $x = \beta J \sum_i \sigma_i$ . Quindi otteniamo la relazione

$$e^{\frac{\beta J}{2N} \sum_{ij} \sigma_i \sigma_j} = \sqrt{\frac{\beta J N}{2\pi}} \int_{\mathbb{R}} dm e^{-\frac{1}{2}N\beta J m^2 + \beta J m \sum_i \sigma_i} .$$

In definitiva questo ci permette di riscrivere la funzione di partizione  $Z$  in termini del campo ausiliario  $m$ , convertendo una teoria a gradi di libertà discreti (spin) in una teoria di campo (per il campo  $m$ ) continuo. Questa trasformazione, detta di "Hubbard e Stratonovich" è una possibile giustificazione delle cosiddette "teorie statistiche di campo" che utilizzano metodi di teoria quantistica dei campi per risolvere problemi di meccanica statistica.

Riscrivendo  $Z$  con questa trasformazione otteniamo

$$Z = \prod_{i=1}^N \sum_{\sigma_i = \pm 1} \sqrt{\frac{\beta J N}{2\pi}} \int_{\mathbb{R}} dm e^{-\frac{1}{2}N\beta J m^2 + \beta J m \sum_i \sigma_i + \beta h \sum_i \sigma_i} ,$$

che possiamo immediatamente riscrivere come

$$Z = \sqrt{\frac{\beta J N}{2\pi}} \int_{\mathbb{R}} dm e^{-\frac{1}{2}N\beta J m^2} \left[ \prod_{i=1}^N \left( \sum_{\sigma_i = \pm 1} e^{+\beta J m \sigma_i + \beta h \sigma_i} \right) \right] .$$

Notiamo ora che la quantità tra quadre si può sommare sugli spin ottenendo un coseno iperbolico, uguale per ogni grado di libertà, per cui il prodotto non è che la stessa quantità elevata alla  $N$ -esima potenza. Riscriviamo quindi la quantità

tra quadre come  $[2 \cosh(\beta Jm + \beta h)]^N$ . Possiamo quindi scrivere la funzione di partizione come

$$Z = \sqrt{\frac{\beta JN}{2\pi}} \int_{\mathbb{R}} dm e^{-N\beta f(\beta, m)}, \quad (52)$$

dove abbiamo isolato la parte che (via saddle point) fornisce la funzione “energia libera”  $f$

$$\beta f = \frac{1}{2} \beta Jm^2 - \log(2 \cosh(\beta Jm + \beta h)). \quad (53)$$

Si noti che l’energia libera strettamente non è la stessa ottenuta dall’approssimazione di campo medio del modello di Ising, ma è molto simile. In particolare ha lo stesso minimo. Infatti se calcoliamo il punto stazionario  $\partial f / \partial m = 0$  otteniamo subito la solita equazione di autoconsistenza

$$m = \tanh(\beta Jm + \beta h),$$

che definisce un valore  $m^*$  per la magnetizzazione attesa. Inoltre  $f$  ha lo stesso e lo stesso sviluppo di Taylor attorno a  $m = 0$  del caso di campo medio, per cui vale tutta la termodinamica discussa sopra. Si noti che per derivare la termodinamica abbiamo usato il punto sella  $Z \approx \exp(-\beta N f(\beta))$  con  $f(\beta, h) = \min_m f(\beta, h, m)$ .

La termodinamica di questo modello è quindi identica a quella dell’approssimazione di campo medio. Vediamo più in dettaglio perché. Iniziamo a considerare la distribuzione marginale di magnetizzazione su un singolo spin,  $\langle \delta_{\sigma_{i,1}} \rangle = \eta_+$  si può stimare (lasciato per esercizio) da saddle point (Eq. (52)) come  $\eta_+ = (1 + m^*)/2$ , per cui possiamo identificare per  $N$  grande  $m^*$  con il valore di aspettazione  $\langle m \rangle$ . Analogamente si calcola la correlazione  $\langle \delta_{\sigma_{i,1}} \delta_{\sigma_{j,1}} \rangle = \left(\frac{1+m^*}{2}\right)^2$ . Per cui vale (per  $N \rightarrow \infty$ ) la proprietà di fattorizzazione

$$\langle \delta_{\sigma_{i,1}} \delta_{\sigma_{j,1}} \rangle = \langle \delta_{\sigma_{i,1}} \rangle \langle \delta_{\sigma_{j,1}} \rangle,$$

e gradi di libertà diversi possono considerarsi scorrelati, grazie alla sovrabbondanza di interazioni “a range infinito”. Come vedremo è sufficiente che le interazioni siano “abbastanza abbondanti” o equivalentemente che la dimensionalità del sistema sia abbastanza alta affinché questa proprietà valga (in pratica sopra 4 dimensioni l’approssimazione di campo medio è esatta).

Più generale la probabilità è asintoticamente fattorizzata, per  $N \rightarrow \infty$

$$P(\{\sigma\}) \approx \prod_i P_i(\sigma_i),$$

e possiamo verificarlo sempre per saddle point

$$Z \sim \int_{\mathbb{R}} dm e^{-\frac{1}{2}N\beta Jm^2} \left[ \prod_{i=1}^N \left( \sum_{\sigma_i=\pm 1} e^{+\beta Jm\sigma_i + \beta h\sigma_i} \right) \right] \\ \approx \int_{\mathbb{R}} dm e^{-\frac{1}{2}N\beta Jm^2} \left[ \prod_{i=1}^N \left( \sum_{\sigma_i=\pm 1} e^{+\beta Jm^* \sigma_i + \beta h\sigma_i} \right) \right].$$

In definitiva, vediamo come il modello di Curie e Weiss è detto di campo medio. In questo caso la misura di Gibbs è asintoticamente fattorizzata, e il quindi modello soddisfa l'ipotesi di campo medio. In altre parole la teoria di campo medio è esatta per il modello di Curie e Weiss.

## Approssimazione di campo medio come "campo molecolare"

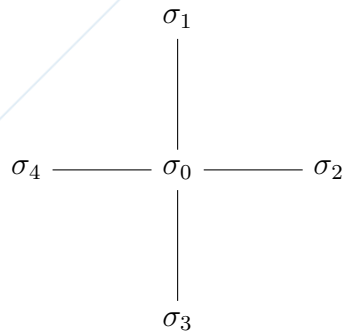
Discutiamo ora un concetto diverso di campo medio che porta ad equazioni simili all'ipotesi di fattorizzazione della misura di probabilità, e che, come vedremo, può essere usata in vari modi per trovare approssimazioni anche molto precise e sofisticate della termodinamica di un sistema di variabili bistabili. Consideriamo uno spin specifico  $\sigma_i$  e vogliamo trovare il "campo medio" (da cui deriva la terminologia) o "campo efficace", cioè l'effetto della magnetizzazione degli altri spin

$$\langle \sigma_i \rangle = \\ \frac{1}{Z} \sum_{\{\sigma\}} \sigma_i e^{-\beta H(\{\sigma\})} = \frac{1}{Z} \sum_{\{\hat{\sigma}\}} e^{-\beta H(\{\hat{\sigma}\})} \sum_{\sigma_i=\pm 1} \sigma_i \frac{e^{-\beta H(\{\hat{\sigma}\}, \sigma_i)}}{Z(\sigma_i)} = \\ \left\langle \tanh\left[\beta\left(h_i + \sum_k J_{ik}\sigma_k\right)\right] \right\rangle$$

Sappiamo che ci sono  $2D$  primi vicini per semplicità assumiamo che il sistema sia bidimensionale  $2D = 4$ , che non ci sia campo esterno,  $h = 0$  e che l'interazione sia omogenea  $J_{ik} = J \forall i, k$ . Chiamiamo  $J\beta = \beta_R$ .

Facendo riferimento alla situazione rappresentata in Fig. ??, possiamo esprimere la magnetizzazione come

$$m = \langle \sigma_0 \rangle = \langle \tanh(\beta_R(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)) \rangle.$$



Essendo la media che compare a destra una funzione di quattro variabili booleane può essere espressa certamente come:

$$\begin{aligned} \tanh(\beta_R(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)) = & A(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4) + \\ & + D(\sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_1\sigma_4 + \sigma_2\sigma_3 + \sigma_2\sigma_4 + \sigma_3\sigma_4) + \\ & + B(\sigma_1\sigma_2\sigma_3 + \sigma_1\sigma_3\sigma_4 + \sigma_1\sigma_2\sigma_4 + \sigma_2\sigma_3\sigma_4) + C(\sigma_1\sigma_2\sigma_3\sigma_4) , \end{aligned}$$

dato che quella scritta è la più generale espressione di 4 variabili booleane. Notiamo però che la funzione  $\tanh$  è dispari, quindi i coefficienti  $D$  e  $C$  sono nulli. Rimane:

$$\begin{aligned} \langle \tanh(\beta_R(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)) \rangle = & \langle A(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4) \rangle + \\ & + \langle B(\sigma_1\sigma_2\sigma_3 + \sigma_1\sigma_3\sigma_4 + \sigma_1\sigma_2\sigma_4 + \sigma_2\sigma_3\sigma_4) \rangle = \\ = & 4mA + 4B \langle \sigma_1\sigma_2\sigma_3 \rangle \end{aligned}$$

I coefficienti  $A$  e  $B$  possono essere calcolati come

$$A = \frac{\tanh(4\beta_R) + 2 \tanh(2\beta_R)}{8}$$

$$B = \frac{\tanh(4\beta_R) - 2 \tanh(2\beta_R)}{8}$$

Utilizziamo solo ora l'ipotesi di campo medio per valutare  $\langle \sigma_1\sigma_2\sigma_3 \rangle$  e troviamo:

$$m = 4mA + 4m^3B ,$$

e notiamo subito che il risultato è diverso rispetto a quello trovato in precedenza; in particolare esso risulta più preciso. Infatti, con questo procedimento troviamo che  $\beta_R^C = 0.357$ . Calcoli più precisi e simulazioni forniscono  $\beta_R^C \simeq 0.44$ , mentre con la tecnica di campo medio usata in precedenza troviamo  $\beta_R^C = 1/4 = 0.25$ .

La ragione del risultato migliore di questa approssimazione è un trattamento più preciso della topologia delle interazioni del sistema. In altre parole, prima di usare

campo medio, abbiamo studiato la configurazione geometrica del sistema, e in particolare i campi locali provenienti dai vicini che agiscono sullo spin, incorporando informazioni sulla “connettività” del grafo, il che permette di trascurare termini dello sviluppo perturbativo della tangente iperbolica presenti nella approssimazione standard, che la nostra analisi della topologia (quattro vicini) ci porta ad arrestare al termine in  $m^3$ .