



Università degli Studi di Bari

---

Corso di Laurea in Fisica

# Tecniche Monte Carlo

Ultimo aggiornamento in data 31 ottobre 2013

Leonardo Angelini

Dipartimento Interateneo di Fisica  
via Amendola 173, 70126 Bari, Italy  
[leonardo.angelini@fisica.uniba.it](mailto:leonardo.angelini@fisica.uniba.it)



# Indice

<b>1</b>	<b>Elementi di Teoria della Probabilità</b>	<b>1</b>
1.1	Spazio dei campioni e Spazio degli eventi . . . . .	1
1.2	Probabilità . . . . .	2
1.2.1	Probabilità: il caso continuo . . . . .	3
1.3	Variabili aleatorie . . . . .	4
1.3.1	Variabili aleatorie nel caso discreto . . . . .	4
1.3.2	Variabili aleatorie nel caso continuo . . . . .	6
1.4	Attesa, varianza e covarianza . . . . .	7
1.5	Distribuzioni . . . . .	10
1.5.1	Distribuzione Binomiale . . . . .	10
1.5.2	Distribuzione di Poisson . . . . .	12
1.5.3	Distribuzione Uniforme . . . . .	15
1.5.4	Distribuzione Normale . . . . .	15
1.5.5	Distribuzione Esponenziale . . . . .	17
1.5.6	Distribuzione di Cauchy . . . . .	18
1.5.7	Distribuzioni multivariate . . . . .	19
1.6	Calcoli con le variabili aleatorie . . . . .	20
1.6.1	Moltiplicazione per una costante . . . . .	20
1.6.2	Funzione di una variabile aleatoria . . . . .	21
1.6.3	Somma di variabili aleatorie . . . . .	22
1.6.4	Prodotto di variabili aleatorie . . . . .	23
1.7	Teoremi limite . . . . .	24
1.7.1	Disuguaglianza di Chebyshev . . . . .	24
1.7.2	Legge dei Grandi Numeri . . . . .	25
1.7.3	Teorema del Limite Centrale . . . . .	25
1.7.4	Cenni sulle distribuzioni di Levy . . . . .	27
<b>2</b>	<b>Campionamento casuale e metodo Monte Carlo</b>	<b>29</b>
2.1	Generazione di numeri casuali uniformemente distribuiti . . . . .	29
2.1.1	Numeri veramente casuali e pseudo - casuali . . . . .	29
2.1.2	Test di casualità . . . . .	30
2.1.3	Generatori di numeri pseudo-casuali . . . . .	32
2.2	Campionamento casuale uniforme e metodo Monte Carlo . . . . .	34

2.2.1	Integrazione Monte Carlo . . . . .	34
2.2.2	Integrazione su un dominio triangolare . . . . .	35
2.3	Campionamento d'importanza e metodo Monte Carlo . . . . .	38
2.3.1	Metodo del Rigetto: distribuzione uniforme, distribuzioni non uniformi . . . . .	38
2.3.2	Metodo di Inversione: variabili aleatorie discrete . . . . .	41
2.3.3	Metodo di Inversione: variabili aleatorie continue . . . . .	43
2.3.4	Campionamento gaussiano . . . . .	44
2.3.5	Metodo del Filtraggio . . . . .	45
2.4	Analisi statistica di dati Monte Carlo . . . . .	46
2.4.1	Stimatore ed Errore Statistico . . . . .	46
2.4.2	Tecniche di riduzione della varianza e campionamento d'Importanza . . . . .	48
2.4.3	Assorbimento di particelle da parte di una lastra . . . . .	50
2.4.4	Tecnica di Spanier . . . . .	53
<b>3</b>	<b>Campionamento in Meccanica Statistica</b>	<b>55</b>
3.1	Il concetto di Ensemble . . . . .	55
3.2	Il Campionamento in Meccanica Statistica . . . . .	56
3.3	Lo stimatore . . . . .	58
3.4	Il Campionamento d'Importanza . . . . .	58
3.5	L'algoritmo di Metropolis . . . . .	60
3.5.1	Applicazione ai modelli di spin . . . . .	62
3.6	L'autocorrelazione . . . . .	62

# Presentazione

Questa dispensa è stata redatta per aiutare gli studenti nella preparazione dell'esame di Metodi Matematici Avanzati per la Fisica del vecchio ordinamento quadriennale del corso di laurea in Fisica dell'Università di Bari. La didattica per questo ordinamento si è esaurita con l'Anno Accademico 2003-04. È tuttavia possibile che questo lavoro possa essere utilizzato per il corso di laurea o di laurea specialistica del nuovo ordinamento. Gli argomenti che vengono qui presentati non hanno nulla di originale, la finalità di questa presentazione essendo quella di raccogliere del materiale sparso in una trattazione piuttosto breve. La trattazione stessa manca spesso di rigore, l'accento è spostato piuttosto sulle applicazioni: il corso si suddivide infatti in una parte teorica e in una parte di laboratorio numerico. Ringrazio il prof. Luigi Nitti per i suggerimenti che mi ha dato sulla scelta del materiale e sulla organizzazione del corso e il prof. Nicola Cufaro Petroni per avermi chiarito, di persona o tramite il suo libro [1], molti concetti di Teoria della Probabilità.



# Introduzione

Il Monte Carlo è una tecnica che fa uso di numeri casuali per risolvere numericamente problemi di diverso tipo. Il nome *Monte Carlo* fu coniato da Nicholas Metropolis nel 1949 con riferimento al gioco d'azzardo, ma tecniche di calcolo ed esperimenti basati sul *campionamento statistico* furono sviluppate per tutto il XIX secolo con la finalità di risolvere problemi che non avevano soluzione analitica. Il capostipite di questo tipo di calcoli è probabilmente l'esperimento noto come *Ago di Buffon* progettato dal Conte di Buffon al fine di ottenere il valore di  $\pi$ , ma altro uso frequente era quello di calcolare numericamente, con duro lavoro di penna e carta, integrali non risolvibili altrimenti. Il primo ad aver applicato i metodi di campionamento statistico alla fisica sembra sia stato Enrico Fermi negli anni trenta, che tuttavia, essendo più interessato ai risultati che alle tecniche con cui li aveva ottenuti, non ne fece mai cenno nei suoi lavori.

È intorno al 1948 ai laboratori di Los Alamos nel gruppo (composto da John von Neumann, Stan Ulam e Nicholas Metropolis) che lavorava allo sviluppo delle armi nucleari che queste tecniche vengono usate per descrivere la diffusione di neutroni e prendono piena cittadinanza nella Fisica moderna. Già nel 1949 si teneva la I Conferenza sui metodi Monte Carlo con un centinaio di partecipanti. La caratteristica di questi lavori è che essi utilizzano il computer, e anzi costituiscono una motivazione importante per la progettazione dei primi calcolatori. Da quel momento lo sviluppo di queste tecniche, come è avvenuto per tutti i metodi numerici di calcolo, si lega indissolubilmente allo sviluppo delle tecnologie informatiche.



# Capitolo 1

## Elementi di Teoria della Probabilità

### 1.1 Spazio dei campioni e Spazio degli eventi

Consideriamo un esperimento, reale o immaginario, dal quale si possa ottenere più di un risultato. I risultati possibili, che indichiamo con il simbolo  $\omega_i$ , sono detti *eventi elementari*. L'insieme di tutti gli eventi elementari viene indicato con  $\Omega$  e viene detto *Spazio dei campioni* o *degli eventi elementari*:

$$\Omega = \{\omega_1, \omega_2, \dots\}$$

Lo *Spazio degli eventi*, detto anche *Campo di Borel* è denominato  $\mathcal{B}$  ed è definito come l'insieme dei sottoinsiemi di  $\Omega$  che soddisfa le seguenti proprietà:

1.  $\Omega \in \mathcal{B}$
2.  $A \in \mathcal{B} \Rightarrow \bar{A} \in \mathcal{B}$
3.  $A, B \in \mathcal{B} \Rightarrow A \cup B \in \mathcal{B}$

È chiaro che poichè l'insieme vuoto  $\emptyset = \bar{\Omega}$  anche  $\emptyset \in \mathcal{B}$ . Se  $A, B \in \mathcal{B}$  anche  $A \cap B \in \mathcal{B}$ . Infatti essendo  $\overline{A \cup B} = \bar{A} \cap \bar{B}$  si ha anche  $A \cap B = \overline{\overline{A \cap B}} = \overline{\bar{A} \cup \bar{B}}$  e quest'ultimo appartiene chiaramente a  $\mathcal{B}$ .

#### **Esempio**

Consideriamo l'esperimento consistente nel lancio di un dado. Lo Spazio degli eventi elementari è costituito da

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

dove con i simboli  $1, 2, \dots$  indichiamo i possibili punteggi risultanti dal lancio. Oltre agli eventi elementari sono possibili anche altri eventi del tipo

$A = \{1, 2\}$  corrispondente alla richiesta che il punteggio sia inferiore a 3.

$B = \{1, 3, 5\}$  corrispondente alla richiesta che il punteggio sia dispari.

$C = \{2, 4, 6\}$  corrispondente alla richiesta che il punteggio sia pari.

Notiamo che  $B$  e  $C$  sono eventi mutualmente esclusivi,  $B \cap C = \emptyset$  e  $B \cup C = \Omega$ , mentre  $A \cap B = \{1\}$  e  $A \cap C = \{2\}$

In generale occorre prestare attenzione alla definizione del Campo di Borel in modo che tutti i possibili risultati di un esperimento siano inclusi in  $\mathcal{B}$ . Esistono anche dei casi in cui  $\Omega$  ha dei sottoinsiemi che non appartengono a  $\mathcal{B}$ . Si tratta tuttavia di costruzioni matematiche che sono prive d'interesse ai fini degli argomenti che vogliamo trattare.

## 1.2 Probabilità

Per ciascuno degli eventi di  $\mathcal{B}$  è possibile assegnare una probabilità. Non esiste un modo univoco di farlo. Tuttavia Kolmogorov ha codificato le proprietà alle quali deve soddisfare una completa e consistente definizione di probabilità.

Ad ogni evento  $A \in \mathcal{B}$  associamo un numero reale  $P(A)$ , detto probabilità di  $A$ , in modo che l'assegnazione soddisfi le seguenti proprietà:

1.  $P(A) \geq 0 \forall A \in \mathcal{B}$
2.  $P(\Omega) = 1$
3.  $\forall A, B \in \mathcal{B} : A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Questi requisiti, come si è detto, lasciano ampia libertà di scelta nella definizione di probabilità, per cui sono possibili vari approcci. Ne elenchiamo alcuni:

- **Definizione a priori.** Se tutti gli eventi elementari sono equiprobabili possiamo definire

$$P(A) = \frac{\text{numero di eventi elementari in } A}{\text{numero di eventi elementari in } \Omega} \quad (1.1)$$

Ad esempio, nel caso del lancio di una moneta ritenuta non truccata, possiamo assumere che la probabilità di ciascuno dei due eventi elementari, consistenti in una delle due facce scoperte, sia  $1/2$ . Analogamente per un dado possiamo supporre che ciascuna delle facce si presenti con probabilità  $1/6$ . Per eventi più complessi applicheremo la formula precedente.

- **Definizione a posteriori** In questo caso si suppone di fare degli esperimenti ripetuti e di assegnare a ciascun evento  $A$  la probabilità:

$$P(A) = \lim_{N \rightarrow \infty} \frac{\text{frequenza di uscita di } A \text{ in } N \text{ esperimenti}}{N} \quad (1.2)$$

- **Definizione soggettiva.** Possiamo anche assegnare a ciascun evento una probabilità in base a quanto siamo disposti a scommettere su di esso.

Dalle proprietà 1-3 si possono ricavare proprietà derivate. Ecco qui alcuni esempi

- $\mathbf{P(A)} \leq 1$ . Tale risultato discende dal fatto che  $P(A) + P(\bar{A}) = P(\Omega) = 1$  e che  $P(\bar{A}) \geq 0$ .
- $\mathbf{A} \subseteq \mathbf{B} \Rightarrow \mathbf{P(A)} \leq \mathbf{P(B)}$ . Infatti  $A \subseteq B \Rightarrow B = A \cup (B \cap \bar{A})$  con  $A \cap (B \cap \bar{A}) = \emptyset$ . Di qui, per la proprietà 3), deriva:  $P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$ .
- $\mathbf{P(A \cup B)} = \mathbf{P(A)} + \mathbf{P(B)} - \mathbf{P(A \cap B)}$ , che generalizza la proprietà 3) al caso in cui  $A$  e  $B$  non siano disgiunti. Infatti in generale  $A \cap (B - A \cap B) = \emptyset$ . Per la proprietà 3) abbiamo  $P(A \cup B) = P(A \cup (B - A \cap B)) = P(A) + P(B - A \cap B)$ . Ma  $P(A \cap B) + P(B - A \cap B) = P(B)$ , da cui, ricavando  $P(B - A \cap B)$  e sostituendo nella precedente si ottiene l'asserto.

Qualche esempio di applicazione di queste proprietà nel caso del lancio di un dado è dato da:

- $P(\{1, 3\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
- $P(\{1, 3, 5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$
- $P(\{1, 3\}) \leq P(\{1, 3, 5\})$

### 1.2.1 Probabilità: il caso continuo

Supponiamo che  $\Omega$  sia lo spazio degli eventi elementari consistenti nell'ottenere come risultato di un esperimento un numero reale. In questo caso il Campo di Borel consiste in tutti i punti e in tutti gli intervalli dell'asse reale. Supponiamo di avere inoltre introdotto le probabilità per gli insiemi  $x \leq \lambda$  per ogni  $\lambda$ . Chiamiamo *distribuzione cumulativa di probabilità* la grandezza

$$P(\lambda) = P(\{x \leq \lambda\}) \quad (1.3)$$

Notiamo che:

$$\begin{aligned} \lim_{\lambda \rightarrow +\infty} P(\lambda) &= P(\Omega) = 1 \\ \lim_{\lambda \rightarrow -\infty} P(\lambda) &= P(\emptyset) = 0. \end{aligned}$$

Supponiamo inoltre che  $P(\lambda)$  sia differenziabile e consideriamo

$$\rho(\lambda) = \frac{dP(\lambda)}{d\lambda}. \quad (1.4)$$

Avremo

$$P(\lambda) = \int_{-\infty}^{\lambda} \rho(x) dx$$

Possiamo ora calcolare la probabilità anche per intervalli finiti:

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \rho(x) dx.$$

Notiamo che risultano verificate le eguaglianze

$$P(-\infty \leq x \leq +\infty) = \int_{-\infty}^{+\infty} \rho(x) dx = 1$$

$$P(x) = 0.$$

Evidentemente  $\rho(x)dx$  rappresenta la probabilità dell'evento  $\lambda \in (x, x + dx)$ . La funzione  $\rho(x)$  viene detta *densità di probabilità* e anche, in Fisica, *funzione di distribuzione*.

## 1.3 Variabili aleatorie

Una *variabile aleatoria* o *casuale*  $X(\omega)$  è una funzione

$$X : \Omega \rightarrow R$$

che fa corrispondere ad ogni evento elementare un numero reale. Essendo una corrispondenza tra eventi, ai quali è associata una probabilità, e numeri reali, si può definire una probabilità anche per la variabile aleatoria. Vediamo come si fa, distinguendo il caso di eventi discreti da quello di eventi continui.

### 1.3.1 Variabili aleatorie nel caso discreto

Consideriamo un esperimento i cui risultati  $\{\omega_i : i = 1, \dots, N\}$  siano discreti con  $N$  finito o infinito.

Indichiamo con  $X(\omega)$  la variabile aleatoria che assegna a ciascun risultato  $\omega_i$  il numero reale  $x_i = X(\omega_i)$ .

La probabilità che la variabile  $X(\omega)$  assuma il valore  $x_i$  è

$$p_i = P(\{\omega | X(\omega) = x_i\}).$$

Le probabilità  $p_i$  soddisfano le condizioni

$$0 \leq p_i \leq 1 \quad \forall i = 1, \dots, N$$

$$\sum_{i=1}^N p_i = 1$$

Definiremo *probabilità cumulativa*  $P_n$  la probabilità che la variabile  $X$  assuma valori nell'insieme  $\{x_1, x_2, \dots, x_n\}$ , cioè

$$P_n = \sum_{i=1}^n p_i$$

**Esempio 1:** lancio di una moneta.

Indichiamo con il carattere  $T$  il caso in cui il risultato sia testa e con  $C$  il caso in cui il risultato sia croce. Lo spazio dei campioni è

$$\Omega = \{T, C\}$$

Lo spazio degli eventi è

$$\mathbf{B} = \{\{T\}, \{C\}, \{T, C\}, \{\emptyset\}\}.$$

Le probabilità associate a tali eventi sono, rispettivamente,  $1/2, 1/2, 1, 0$ .

Possiamo definire una variabile aleatoria  $X(\omega)$  nel modo seguente

$$X(\{T\}) = +1 \quad X(\{C\}) = -1$$

Per quanto riguarda le probabilità avremo

$$P(+1) = P(\omega | X(\omega) = +1) = \frac{1}{2}$$

$$P(-1) = P(\omega | X(\omega) = -1) = \frac{1}{2}$$

**Esempio 2:** lancio di  $N$  monete.

In questo caso abbiamo  $N$  variabili aleatorie con valori

$$X_i(\{T\}) = +1 \quad X_i(\{C\}) = -1 \quad \forall i = 1, \dots, N$$

Se le monete non sono truccate avremo  $2^N$  eventi elementari equiprobabili nello spazio  $\Omega$ . Ciascuno di essi è rappresentato da una stringa composta di  $T$  e di  $C$  e ha probabilità  $1/2^N$ .

Definiamo una variabile aleatoria

$$Y_N = X_1 + X_2 + \dots + X_N.$$

All'evento composto da  $n$  teste e  $N - n$  croci corrisponde un valore di  $Y_N$  pari a

$$Y_N = n - (N - n) = 2n - N$$

Il numero di tali eventi è

$$\frac{\text{n. di permutazioni delle monete}}{(\text{n. di permutazioni delle teste}) \times (\text{n. di permutazione delle croci})} = \frac{N!}{n!(N-n)!}$$

e quindi la probabilità (*a priori*) di tale evento è

$$P(n, N) = P(Y_N = 2n - N) = \frac{1}{2^N} \frac{N!}{n!(N-n)!}$$

Nel caso più generale in cui le probabilità di avere testa e croce sono rispettivamente  $p$  e  $q = 1 - p$  si ha la distribuzione *Binomiale*

$$P(n, N) = P(Y_N = 2n - N) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad (1.5)$$

che è la distribuzione di probabilità per eventi dicotomici del tipo *successo - insuccesso*.

### 1.3.2 Variabili aleatorie nel caso continuo

Sostanzialmente questo tipo di variabili aleatorie è già stato presentato quando si è parlato delle probabilità per eventi il cui spazio  $\Omega$  è continuo.

In questi casi è possibile associare a ciascun evento una (o più) variabile aleatoria continua  $X(\omega)$  e definire una probabilità  $\rho_X(x)dx$  dell'evento che consiste nel fatto che  $X(\omega)$  assume valori tra  $x$  e  $x + dx$ , cioè

$$\rho_X(x)dx = P(\omega | x \leq X(\omega) \leq x + dx)$$

Come si è già detto  $\rho_X(x)dx$  viene detta *densità di probabilità* o *funzione di distribuzione*. Essa soddisfa le condizioni

$$\rho_X(x) \geq 0 \quad \int_a^b \rho_X(x)dx = 1$$

dove  $a$  e  $b$  sono gli estremi dell'intervallo dei valori assunti da  $X(\omega)$ .

La probabilità cumulativa che  $X$  assuma valori inferiori a un fissato valore  $\lambda$  è definita da

$$P(\lambda) = P(\omega | X(\omega) \leq \lambda) = \int_a^\lambda \rho_X(x)dx.$$

Ricordiamo inoltre che

$$\rho_X(x) = \frac{dP(x)}{dx} \quad \Rightarrow \quad dP(x) = \rho_X(x)dx.$$

Esempi di funzioni di distribuzione continue sono la distribuzione Uniforme o quella Gaussiana, distribuzioni che studieremo più in dettaglio in seguito.

## 1.4 Attesa, varianza e covarianza

La *media*  $\mu$  su una distribuzione di probabilità di una variabile aleatoria è detta più precisamente *attesa o valore di attesa o di aspettazione*, per non confonderla con il concetto di media (aritmetica) di un insieme di misure. Nel caso di una variabile aleatoria **discreta**  $X(\Omega)$  con assegnate probabilità  $p_i$  è definita da

$$\mu(X) = \sum_i x_i p_i.$$

In caso di una variabile aleatoria **continua**  $X(\Omega)$  distribuita secondo una densità di probabilità  $\rho_X(x)$  avremo invece

$$\mu(X) = \int_a^b x \rho_X(x) dx = \int_a^b x dP(x).$$

Nella pratica un esperimento viene ripetuto  $N$  volte dando luogo ai risultati  $\omega_1, \omega_2, \dots, \omega_N$ . A questi risultati corrispondono i valori  $x_i = X(\omega_i)$  con  $i = 1, \dots, N$  della variabile aleatoria  $X(\omega)$ . Vedremo più avanti discutendo la Legge dei Grandi Numeri, che, nel limite  $N \rightarrow \infty$ , la media aritmetica  $\sum_{i=1}^N x_i/N$  converge<sup>1</sup> a  $\mu(X)$ . Premettiamo tuttavia che la stima dell'errore statistico, fatto nel valutare il valore di attesa usando il valor medio dei risultati degli esperimenti, dipende dal numero degli esperimenti e dal valore di un altro importante parametro, la *varianza* della variabile aleatoria  $X(\omega)$ .

La varianza è definita come il valore di attesa della variabile aleatoria  $(X - \mu)^2$ :

$$\sigma^2(X) = \mu((X - \mu(X))^2) = \begin{cases} \sum_i (x_i - \mu)^2 p_i & \text{nel caso discreto,} \\ \int_a^b (x - \mu)^2 \rho_X(x) dx & \text{nel caso continuo.} \end{cases} \quad (1.6)$$

È facile verificare che:

$$\sigma^2(X) = \mu(X^2) - \mu^2(X) \quad (1.7)$$

La radice quadrata della varianza viene detta *Deviazione Standard* ed è indicata con il simbolo  $\sigma(X)$ .

### Momenti

Definiamo *momento  $k$ -simo* di una variabile aleatoria  $X$  la quantità

$$M_k(X) \equiv \mu(X^k) = \begin{cases} \sum_i x_i^k p_i & \text{nel caso discreto,} \\ \int_a^b x^k \rho_X(x) dx & \text{nel caso continuo.} \end{cases} \quad (1.8)$$

---

<sup>1</sup>Sottolineiamo il fatto che il concetto di convergenza in questo caso è differente da quello usuale dell'Analisi Matematica nella quale una successione  $\{A(\nu) | \nu = 1, 2, \dots\}$  converge a un valore  $B$  per  $\nu \rightarrow \infty$  se, per ogni  $\delta > 0$ , esiste un numero intero  $M$  tale che, per ogni  $k \geq M$ , risulta  $|A(k) - B| \leq \delta$ . In Statistica la convergenza, invece, ha un significato probabilistico. Diremo che  $A(\nu)$  converge stocasticamente a  $B$  per  $\nu \rightarrow \infty$  se, comunque si fissi una probabilità  $P \in ]0, 1[$  e  $\delta > 0$ , esiste un numero intero  $M$  tale che, per ogni  $k \geq M$ , la probabilità che risulti  $|A(k) - B| \leq \delta$  è maggiore di  $P$ .

Avremo chiaramente

$$M_0(X) = 1, \quad M_1(X) = \mu(X), \quad \sigma^2(X) = M_2(X) - M_1^2(X)$$

**Somma di due variabili aleatorie** Siano date due variabili aleatorie  $X_1$  e  $X_2$  continue con valori nell'intervallo  $(a, b)$  e sia  $\rho(x_1, x_2)$  la loro densità di probabilità *congiunta*. Il valor d'attesa di  $Y_2 = X_1 + X_2$  è definito da

$$\mu(Y_2) = \int_a^b dx_1 \int_a^b dx_2 (x_1 + x_2) \rho(x_1, x_2) \quad (1.9)$$

Le densità di probabilità di  $X_1$  e  $X_2$ ,  $\rho_1(x_1)$  e  $\rho_2(x_2)$  rispettivamente, vengono dette densità *marginali* e si ottengono dalla densità congiunta dalle formule

$$\begin{aligned} \rho_1(x_1) &= \int_a^b dx_2 \rho(x_1, x_2) \\ \rho_2(x_2) &= \int_a^b dx_1 \rho(x_1, x_2) \end{aligned}$$

Notiamo che possiamo scrivere  $\mu(Y_2)$  nella forma

$$\begin{aligned} \mu(Y_2) &= \int_a^b dx_1 x_1 \int_a^b dx_2 \rho(x_1, x_2) + \int_a^b dx_2 x_2 \int_a^b dx_1 \rho(x_1, x_2) = \\ &= \int_a^b dx_1 x_1 \rho_1(x_1) + \int_a^b dx_2 x_2 \rho_2(x_2) = \mu(X_1) + \mu(X_2). \end{aligned}$$

Quindi le attese si sommano. Questo invece non accade per le varianze:

$$\begin{aligned} \sigma^2(Y_2) &= \mu((X_1 + X_2 - \mu(X_1) - \mu(X_2))^2) = \\ &= \mu((X_1 - \mu(X_1))^2 + (X_2 - \mu(X_2))^2 + 2(X_1 - \mu(X_1))(X_2 - \mu(X_2))) = \\ &= \sigma^2(X_1) + \sigma^2(X_2) + 2 \text{cov}(X_1, X_2) \end{aligned}$$

dove abbiamo introdotto la covarianza di  $X_1$  e  $X_2$  definita da

$$\begin{aligned} \text{cov}(X_1, X_2) &= \mu((X_1 - \mu(X_1))(X_2 - \mu(X_2))) = \\ &= \int_a^b dx_1 \int_a^b dx_2 (x_1 - \mu_1)(x_2 - \mu_2) \rho(x_1, x_2) \end{aligned}$$

Potremo dire che la varianza della somma di due variabili aleatorie è uguale alla somma delle varianze delle due variabili soltanto se la loro covarianza è nulla. Chiediamoci quando questo può avvenire. Definiamo *densità condizionata* di  $X_1$  dato che  $X_2$  assuma il valore  $x_2$  la quantità

$$\rho(x_1|x_2) = \frac{\rho(x_1, x_2)}{\rho_2(x_2)} \quad (1.10)$$

Notiamo che

$$\rho(x_1|x_2) = \rho_1(x_1) \Rightarrow \rho(x_1, x_2) = \rho_1(x_1)\rho_2(x_2).$$

In questo caso diremo che le due variabili aleatorie  $X_1$  e  $X_2$  sono *indipendenti*; si vede subito che  $\text{cov}(X_1, X_2) = 0$ . Diremo inoltre che

$$\begin{aligned} X_1 \text{ e } X_2 & \text{ sono correlate positivamente se } \text{cov}(X_1, X_2) > 0 \\ X_1 \text{ e } X_2 & \text{ sono correlate negativamente se } \text{cov}(X_1, X_2) < 0 \end{aligned}$$

Notiamo che

indipendenza  $\rightarrow$  covarianza nulla (assenza di correlazione)

Il contrario invece non è vero, cioè vi può essere assenza di correlazione anche per variabili non indipendenti.

Sovente si usa il *coefficiente di correlazione*

$$C(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}$$

che è adimensionale e compreso tra -1 e 1.

Prima di passare a calcolare media e varianza per alcune densità di probabilità di uso frequente, conviene chiarire il concetto di probabilità condizionata. La questione importante sta nel fatto che l'acquisizione di nuova informazione modifica le conoscenze e fa valutare la probabilità in maniera diversa. Per questo

$$\rho(x_1|x_2) \text{ e } \rho(x_1, x_2)$$

sono differenti in quanto nel primo caso sappiamo che  $X_2$  ha assunto un valore preciso  $x_2$ . Infatti, dalla (1.10),

$$\rho(x_1, x_2) = \rho_2(x_2)\rho(x_1|x_2)$$

**Esempio:** Usiamo delle variabili discrete per semplificare. Chiediamoci quale sia la probabilità che ad una estrazione del lotto esca 1 nella 2<sup>a</sup> estrazione sapendo che è uscito 2 nella 1<sup>a</sup>. Tale probabilità è chiaramente

$$p(1|2) = \frac{1}{89}.$$

La probabilità che escano 2 nella 1<sup>a</sup> estrazione e 1 nella 2<sup>a</sup> è invece

$$p(1, 2) = \frac{1}{90} \frac{1}{89},$$

verificando quindi la relazione

$$\frac{1}{89} = p(1|2) = \frac{p(1, 2)}{p(2)} = \frac{\frac{1}{90} \frac{1}{89}}{\frac{1}{90}}$$

## 1.5 Distribuzioni

### 1.5.1 Distribuzione Binomiale

Come abbiamo visto, questa distribuzione si ottiene quando si sottopone a  $N$  prove indipendenti di verifica un esperimento, il cui risultato può essere classificato come successo (per il quale vi è una probabilità  $p$ ) o insuccesso (per il quale vi è una probabilità  $q = 1 - p$ ). Per il singolo risultato abbiamo introdotto una variabile aleatoria  $X_i$ , che può assumere i valori  $+1$  (successo) o  $-1$  (insuccesso). Per le  $N$  prove usiamo invece la variabile aleatoria

$$Y_N = \sum_{i=1}^N X_i.$$

La probabilità di avere  $n$  successi su  $N$  prove è data proprio dalla distribuzione binomiale (1.5)

$$p_n = P(n, N) = P(Y_N = 2n - N) = \frac{N!}{n!(N-n)!} p^n q^{N-n}.$$

Nella figura (1.1) questa distribuzione viene mostrata per  $N = 6$  nei casi  $p = 0.3$  e  $p = 0.5$ . Alternativamente si può definire la variabile aleatoria  $X_i$  in modo che assuma il valore 1 in caso di successo e 0 in caso di insuccesso. In questo caso, denominata

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

la variabile aleatoria relativa alle  $N$  prove, avremo

$$P(n, N) = P(Z_N = n)$$

Al fine di calcolare il valor di attesa di  $Z_N$

$$\mu(Z_N) \equiv \langle n \rangle = \sum_{n=0}^N n p_n,$$

deriviamo rispetto a  $p$  l'identità

$$\sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = 1. \quad (1.11)$$

Si ottiene

$$\sum_{n=0}^N \frac{N!}{n!(N-n)!} [np^{n-1} (1-p)^{N-n} - (N-n)p^n (1-p)^{N-n-1}] = 0,$$

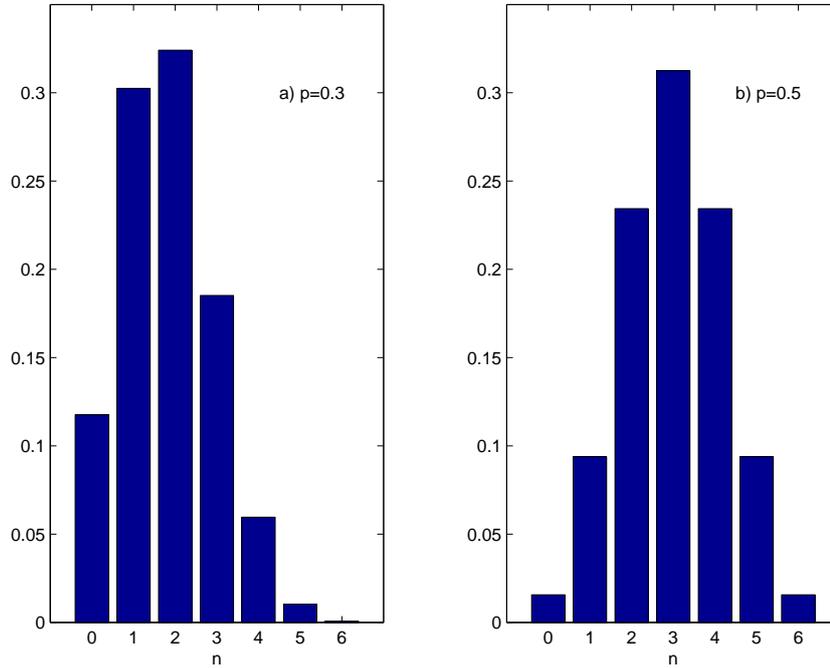


Figura 1.1: La distribuzione Binomiale  $P(n,6)$  per  $p = 0.3$  e  $p = 0.5$

da cui si deduce

$$\frac{1}{p} \langle n \rangle = \frac{\langle N - n \rangle}{1 - p} \Rightarrow \mu(Z_N) \equiv \langle n \rangle = pN$$

Questa espressione consente di ricavare agevolmente anche il valore di attesa di  $Y_N$ :

$$\mu(Y_N) \equiv \langle 2n - N \rangle = 2\langle n \rangle - N = 2pN - N = (2p - 1)N.$$

La varianza può essere calcolata derivando una seconda volta la relazione (1.11):

$$\begin{aligned} \sum_{n=0}^N \frac{N!}{n!(N-n)!} [n(n-1)p^{n-2}(1-p)^{N-n} - n(N-n)p^{n-1}(1-p)^{N-n-1} - \\ - n(N-n)p^{n-1}(1-p)^{N-n-1} + (N-n)(N-n-1)p^n(1-p)^{N-n-2}] = 0 \\ \frac{1}{p^2} \langle n(n-1) \rangle + \frac{1}{(1-p)^2} \langle (N-n)(N-n-1) \rangle - \frac{2}{p(1-p)} \langle n(N-n) \rangle = 0 \end{aligned}$$

E, ricordando che  $\langle n \rangle = pN$ , si trovano facilmente il valor medio di  $n^2$

$$\langle n^2 \rangle = p^2 N^2 + p(1-p)N$$

e la varianza di  $Z_N$ :

$$\sigma^2(Z_N) = \langle n^2 \rangle - \langle n \rangle^2 = p^2 N^2 + p(1-p)N - p^2 N^2 = p(1-p)N$$

Per la variabile  $Y_N$  si ha, invece,

$$\sigma^2(Y_N) = \langle (2n - N)^2 \rangle - \langle 2n - N \rangle^2 = 4\langle n^2 \rangle - 4N\langle n \rangle + N^2 - (2p - 1)^2 N^2 = 4p(1 - p)N$$

Notiamo che sia i valori di attesa di  $Y_N$  e  $Z_N$  che le loro varianze crescono come  $N$ . Tuttavia la variabile rilevante, che è il rapporto tra deviazione standard e attesa, decresce come  $1/\sqrt{N}$ . Per avere una variabile aleatoria il cui valore di attesa non dipende da  $N$  occorre considerare  $Y_N/N$  o  $Z_N/N$ . In questo caso la deviazione standard decresce come  $1/\sqrt{N}$ . Infine, le variabili  $Y_N/\sqrt{N}$  e  $Z_N/\sqrt{N}$  hanno un'attesa che va come  $\sqrt{N}$  e varianza e deviazione standard che sono indipendenti da  $N$ .

Vale la pena di notare che valore d'attesa e varianza si possono calcolare molto più facilmente tenendo conto del fatto che  $Y_N$  (o  $Z_N$ ) è somma di variabili aleatorie indipendenti e quindi l'attesa è la somma delle attese e la varianza è la somma delle varianze. Ad esempio, nel caso di  $Y_N$  si ha, per il singolo esperimento,

$$\begin{aligned}\mu(X_i) &= p \times (+1) + (1 - p) \times (-1) = 2p - 1 \\ \sigma^2(X_i) &= p \times (+1)^2 + (1 - p) \times (-1)^2 - (2p - 1)^2 = 4p(1 - p)\end{aligned}$$

e, per  $N$  esperimenti,

$$\begin{aligned}\mu(Y_N) &= (2p - 1)N \\ \sigma^2(Y_N) &= 4p(1 - p)N.\end{aligned}$$

### 1.5.2 Distribuzione di Poisson

La distribuzione di Poisson viene considerata quando si hanno accadimenti aleatori nel tempo. Supponiamo che esista una costante di tempo caratteristica di un processo  $\lambda^{-1}$ , di modo che  $\lambda$  rappresenti la probabilità per unità di tempo che un certo evento si verifichi. A partire da un fissato istante  $t$ , l'evento si verificherà nel successivo lasso di tempo  $\Delta t$  con probabilità  $\lambda \Delta t$  e non si verificherà con probabilità  $1 - \lambda \Delta t$ .

Denominata con  $P(n, t)$  la probabilità che si verifichino  $n$  eventi nell'intervallo  $(0, t)$ , possiamo scrivere la *Master Equation*:

$$\begin{aligned}P(n, t) &= \lambda \Delta t P(n - 1, t - \Delta t) + (1 - \lambda \Delta t) P(n, t - \Delta t) \quad \text{per } n \geq 1 \\ P(0, t) &= (1 - \lambda \Delta t) P(0, t - \Delta t) \\ P(n, 0) &= \delta_{n,0}\end{aligned}$$

Per risolvere questa equazione introduciamo la *funzione generatrice*

$$\tilde{P}(z, t) = \sum_{n=0}^{\infty} z^n P(n, t).$$

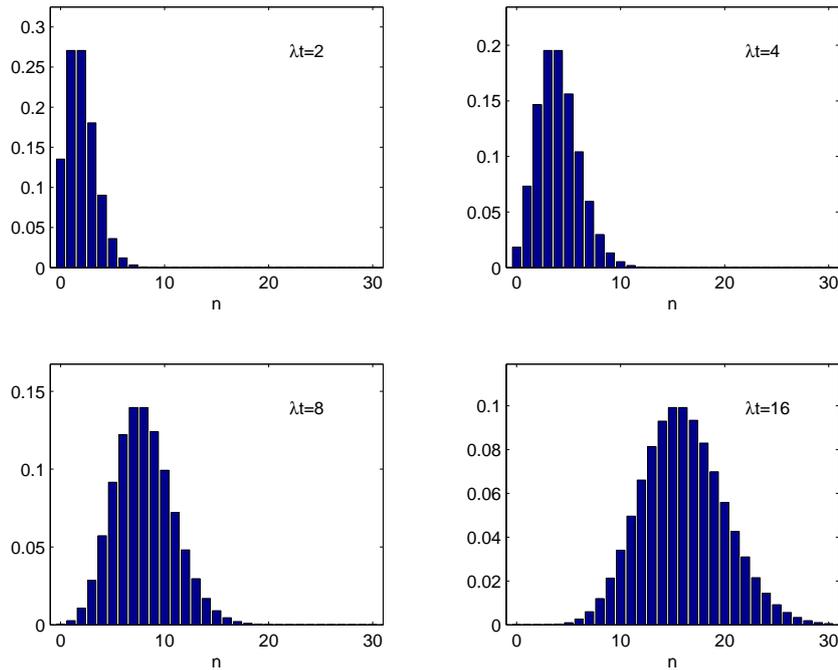


Figura 1.2: La distribuzione di Poisson  $P(n,t)$  per  $\lambda t = 2, 4, 8, 16$

Moltiplicando le equazioni precedenti per  $z^n$  e sommando su  $n$ , si ottiene

$$\begin{aligned}\tilde{P}(z, t) &= \lambda z \Delta t \tilde{P}(z, t - \Delta t) + (1 - \lambda \Delta t) \tilde{P}(z, t - \Delta t) \\ \tilde{P}(z, 0) &= 1\end{aligned}$$

Nel limite  $\Delta t \rightarrow 0$  si ottiene l'equazione differenziale

$$\frac{\partial \tilde{P}(z, t)}{\partial t} = -\lambda(1 - z) \tilde{P}(z, t); \quad \tilde{P}(z, 0) = 1$$

Il suo integrale, per le condizioni iniziali date, è

$$\tilde{P}(z, t) = e^{-\lambda(1-z)t} = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} z^n e^{-\lambda t}$$

$P(n, t)$ , che prende il nome di *distribuzione di Poisson*, è il coefficiente di  $z^n$  nello sviluppo di Taylor di  $\tilde{P}(z, t)$

$$P(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Nella figura (1.2) viene mostrata la dipendenza da  $n$  della distribuzione  $P(n, t)$

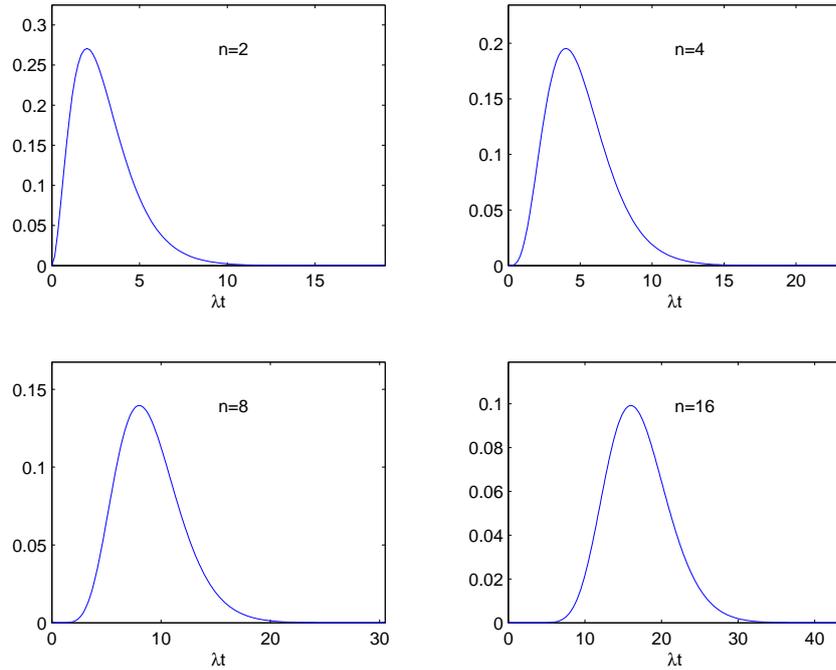


Figura 1.3: La distribuzione di Poisson  $P(n,t)$  per  $n = 2, 4, 8, 16$

per alcuni valori di  $\lambda t$ , mentre nella figura (1.3) se ne mostra la dipendenza da  $\lambda t$  per fissati valori di  $n$ .

Calcoliamo ora il valor d'attesa degli eventi che si verificano nell'intervallo  $(0, t)$ :

$$\begin{aligned} \langle n(t) \rangle &= \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \sum_{n=1}^{\infty} \lambda t \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \\ &= \lambda t \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t \end{aligned}$$

Il secondo momento della distribuzione è invece dato da:

$$\begin{aligned} \langle n^2(t) \rangle &= \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \sum_{n=1}^{\infty} \lambda t n \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \\ &= \lambda t \sum_{n=0}^{\infty} (n+1) \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t (1 + \langle n(t) \rangle) = \lambda t (1 + \lambda t) \end{aligned}$$

Poichè la varianza è

$$\sigma^2(n) = \langle n^2 \rangle - \langle n \rangle^2 = \lambda t,$$

per la distribuzione di Poisson attesa e varianza coincidono.

ESERCIZIO: Calcolare attesa e varianza per la distribuzione di Poisson usando il metodo visto per la distribuzione binomiale, derivando, cioè, rispetto a  $\lambda$  l'identità

$$\sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = 1$$

### 1.5.3 Distribuzione Uniforme

La più semplice delle distribuzioni continue è quella relativa ad una variabile aleatoria  $U(a, b)$  che assume valori  $x \in [a, b]$  con densità di probabilità uniforme

$$\rho_U(x) dx = \frac{1}{b-a} dx \quad (1.12)$$

L'attesa è data in questo caso da

$$\mu(U) = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}$$

e la varianza da

$$\sigma^2(U) = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{1}{b-a} \int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} x^2 dx = \frac{(b-a)^2}{12}$$

Un caso particolare di uso frequente è la variabile aleatoria  $U(0, 1)$  per la quale attesa e varianza sono date da:

$$\mu(U) = \frac{1}{2} \quad \sigma^2(U) = \frac{1}{12}. \quad (1.13)$$

$U(0, 1)$  è importante nel contesto dei numeri pseudo-casuali e degli algoritmi per generarli.

ESERCIZIO: Considerare  $N$  variabili aleatorie indipendenti distribuite uniformemente nell'intervallo  $[0, 1]$ . Sia  $Y_N$  la loro somma. Calcolare attesa e varianza per  $Y_N, Y_N/N, Y_N/\sqrt{N}$ .

### 1.5.4 Distribuzione Normale

La distribuzione gaussiana o normale è di fondamentale importanza in numerosi campi della Fisica e di altre discipline scientifiche. Essa riguarda una variabile aleatoria  $X$  che assume valori nell'intervallo  $(-\infty, +\infty)$  con densità di probabilità

$$\rho_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \quad (1.14)$$

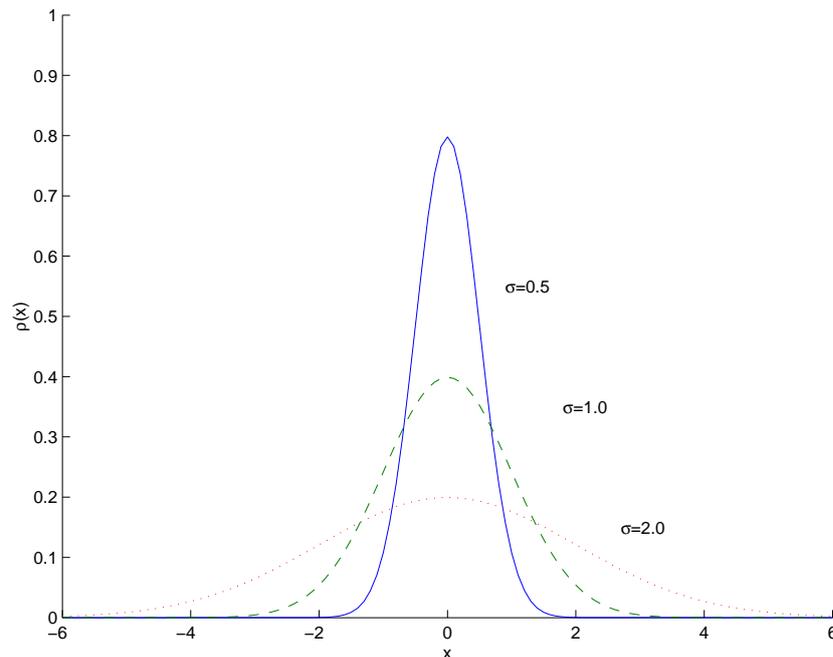


Figura 1.4: La distribuzione di Gauss per  $\mu = 0$  e  $\sigma = 0.5, 1, 2$

$\mu$  e  $\sigma$  sono due parametri che coincidono con il valore d'attesa e la deviazione standard della distribuzione. In figura (1.4) la distribuzione normale è riportata nel caso  $\mu = 0$  per alcuni valori di  $\sigma$ .

La probabilità che la variabile aleatoria assuma valori compresi nell'intervallo  $(\mu - \sigma, \mu + \sigma)$  è data da

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx = 0.68268 \quad (1.15)$$

Quindi più del 68% dei valori estratti casualmente e indipendentemente da una distribuzione gaussiana cadono in questo intervallo, che è detto *intervallo di confidenza di una sigma*.

Vedremo che la somma  $Y_N$  di  $N$  variabili aleatorie gaussiane è anche essa una variabile gaussiana con attesa  $\mu N$  e varianza  $\sigma^2 N$ . La media delle variabili  $Y_N/N$  sarà ancora gaussiana con attesa  $\mu$  e varianza  $\sigma^2/N$ . Vedremo inoltre che il Teorema del Limite Centrale stabilisce che  $Y_N/N$ , nel limite  $N \rightarrow \infty$  tende ad una distribuzione gaussiana, anche se le distribuzioni di partenza non sono gaussiane (purchè  $\sigma$  sia finita). La standard deviation della distribuzione limite sarà proporzionale a  $1/\sqrt{N}$  e tenderà a zero per  $N \rightarrow \infty$ . Si potrà così interpretare l'intervallo di confidenza di una sigma come l'errore statistico associato con la media stimata delle  $N$  variabili aleatorie.

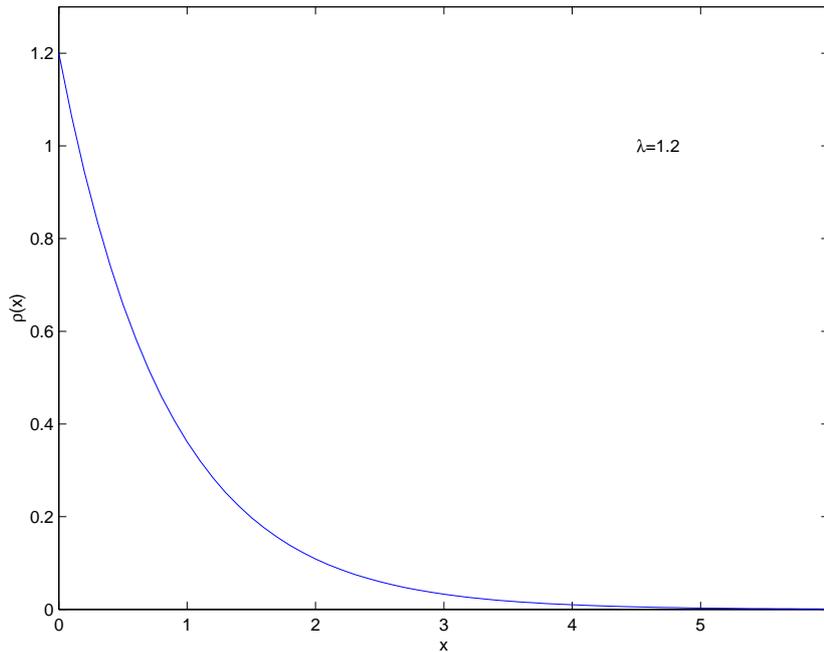


Figura 1.5: La distribuzione esponenziale per  $\lambda = 1.2$

### 1.5.5 Distribuzione Esponenziale

La densità di probabilità esponenziale viene utilizzata in Fisica in vari contesti, come ad esempio per il tempo di decadimento di un nucleo radioattivo. Essa è data da

$$\rho_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{per } x \geq 0, \\ 0 & \text{per } x < 0 \end{cases} \quad (1.16)$$

La variabile aleatoria  $X$  assume valori nell'intervallo  $(0, +\infty)$ .  $\lambda$  è un parametro positivo che caratterizza la distribuzione. La figura (1.5) mostra l'andamento della distribuzione per  $\lambda = 1.2$ . Il valore d'attesa della distribuzione esponenziale è dato da

$$\mu = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

e la varianza è data da

$$\sigma^2(X) = \int_0^{+\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}.$$

La standard deviation è  $\sigma = \frac{1}{\lambda}$ .

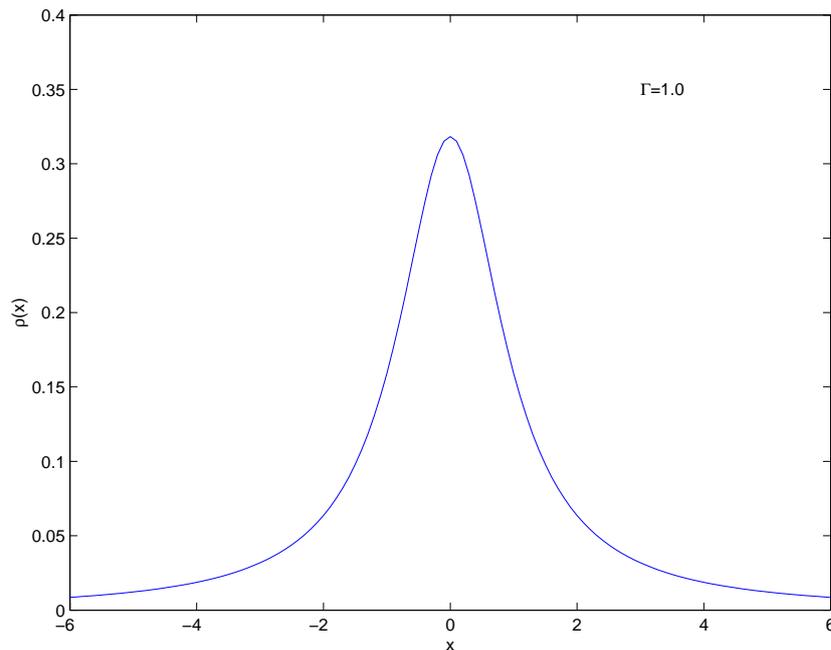


Figura 1.6: La distribuzione di Cauchy per  $\Gamma = 1.0$

La probabilità che la variabile aleatoria  $X$  assuma valori compresi nell'intervallo  $(\mu - \sigma, \mu + \sigma)$  è data da

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = \int_{\mu - \sigma}^{\mu + \sigma} \lambda e^{-\lambda x} dx = 0.8647$$

Quindi più del 86% dei valori ottenuti da un campionamento casuale per una variabile distribuita esponenzialmente sono compresi nell'intervallo di confidenza di una sigma.

Nel caso del decadimento radioattivo  $\mu = 1/\lambda$  viene detto tempo di vita media della sostanza radioattiva.

### 1.5.6 Distribuzione di Cauchy

La distribuzione di Cauchy è data da

$$\rho_X(x) = \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + x^2} \quad (1.17)$$

dove  $\Gamma$  è un parametro di scala. Nella figura (1.6) è rappresentata la distribuzione di Cauchy per  $\Gamma = 1$ . L'attesa della distribuzione è data da

$$\mu = \int_{-\infty}^{+\infty} x \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + x^2} dx = 0.$$

A rigore questo integrale non esiste perchè la funzione non si azzera abbastanza rapidamente all'infinito. Tuttavia possiamo considerarlo in senso lato nullo, poichè si tratta dell'integrale di una funzione dispari su un'intervallo simmetrico rispetto all'origine.

Per calcolare la varianza dovremmo invece calcolare l'integrale

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + x^2} dx \quad (1.18)$$

che è divergente. Non è possibile, quindi, definire la deviazione standard e l'intervallo di confidenza di una sigma.

Ciononostante la distribuzione di Cauchy soddisfa le condizioni di Kolmogorov e rappresenta quindi una valida densità di probabilità. Invece della varianza questa distribuzione può essere caratterizzata dalla *ampiezza totale a metà massimo*, cioè l'intervallo della variabile  $x$  nel quale il valore di  $\rho(x)$  supera la metà del suo valore massimo ( $1/\Gamma\pi$ ). L'ampiezza di tale intervallo è  $2\Gamma$ .

La distribuzione di Cauchy è nota in Fisica Nucleare come distribuzione di Lorentz o di Breit-Wigner e viene utilizzata per descrivere le risonanze nelle sezioni d'urto.

### 1.5.7 Distribuzioni multivariate

Come si è già detto, accade spesso, in particolare in Fisica, che ad un evento elementare siano associate più variabili aleatorie o anche variabili aleatorie che sono dei vettori a  $n$  componenti. In questo caso si parla *variabili aleatorie multivariate o multidimensionali*. Un esempio è dato dal momento  $\mathbf{p}$  di una molecola di un gas che, in Meccanica Statistica Classica, è considerata una variabile aleatoria tridimensionale distribuita secondo la densità

$$\rho(\mathbf{p}) = \frac{1}{A} e^{-\frac{\mathbf{p}^2}{2mk_B T}} \quad (1.19)$$

dove  $T$  è la temperatura assoluta,  $k_B$  la costante di Boltzmann e  $A$  una costante di normalizzazione. Vediamo quindi che, in questo caso, abbiamo una densità tridimensionale gaussiana.

Come altro esempio di distribuzione multivariata possiamo considerare la *distribuzione di Boltzmann* che descrive un gas di  $N$  particelle classiche

$$\rho(\mathbf{p}, \mathbf{q}) = \frac{1}{A} e^{-\frac{\mathcal{H}(\mathbf{p}, \mathbf{q})}{k_B T}}. \quad (1.20)$$

In questo caso le variabili aleatorie

$$(\mathbf{p}, \mathbf{q}) = (\mathbf{p}_1, \dots, \mathbf{p}_N, \mathbf{q}_1, \dots, \mathbf{q}_N)$$

rappresentano lo stato del sistema classico, cioè la sua posizione nello spazio delle fasi a  $6N$  dimensioni, mentre  $\mathcal{H}(\mathbf{p}, \mathbf{q})$  è la sua funzione Hamiltoniana.

La più generale distribuzione gaussiana  $n$ -dimensionale (o multivariata) per la variabile aleatoria  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  è data da

$$\rho_{\mathbf{X}}(\bar{\mathbf{x}}, \mathbf{A}; \mathbf{x}) = \frac{(2\pi)^{-n/2}}{\sqrt{\det \mathbf{A}}} e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})_i(\mathbf{A}^{-1})_{ij}(\mathbf{x}-\bar{\mathbf{x}})_j} \quad (1.21)$$

dove abbiamo usato la convenzione di somma sugli indici ripetuti. Il vettore  $\bar{\mathbf{x}}$  e la matrice  $\mathbf{A}$  sono parametri della distribuzione. Il loro significato è chiaro se si calcolano attesa e covarianza. Si trova

$$\langle \mathbf{X} \rangle = \bar{\mathbf{x}} \quad (1.22)$$

$$\text{Cov}(x_i, x_j) = \langle (\mathbf{x} - \bar{\mathbf{x}})_i (\mathbf{x} - \bar{\mathbf{x}})_j \rangle = A_{ij} \quad (1.23)$$

Per questo la matrice  $\mathbf{A}$  viene detta matrice di covarianza per la distribuzione normale. Per variabili indipendenti

$$A_{ij} = \sigma_i^2 \delta_{ij} \Rightarrow A_{ij}^{-1} = \frac{\delta_{ij}}{\sigma_i^2};$$

cioè tutte le coppie di variabili hanno covarianza nulla e ciascuna di esse ha varianza  $\sigma_i^2$ .

## 1.6 Calcoli con le variabili aleatorie

Le variabili aleatorie possono essere combinate in vario modo: si possono considerare le variabili aleatorie risultanti da operazioni su v. a. o anche dall'applicazione di funzioni a v. a.. Nel seguito supporremo che il risultato di tali operazioni abbia significato.

Il problema che qui ci poniamo è come determinare la densità di probabilità per le variabili risultanti.

### 1.6.1 Moltiplicazione per una costante

Consideriamo una variabile aleatoria  $X$  e la sua densità di probabilità  $\rho_X(x)$ . Poniamo

$$Z = \alpha X \quad (1.24)$$

Avremo

$$\rho_Z(z) = \int \rho_X(x) \delta(z - \alpha x) dx = \int \rho_X(x) \delta(\alpha(\frac{z}{\alpha} - x)) dx = \frac{1}{|\alpha|} \rho_X(\frac{z}{\alpha}) \quad (1.25)$$

Per il valore d'attesa si ha

$$\langle Z \rangle = \mu(Z) = \int z \rho_Z(z) dz = \frac{1}{|\alpha|} \int z \rho_X(\frac{z}{\alpha}) dz = \alpha \langle X \rangle \quad (1.26)$$

Analogamente si trova

$$\sigma^2(Z) = \alpha^2 \sigma^2(X). \quad (1.27)$$

### 1.6.2 Funzione di una variabile aleatoria

Siano  $X$  e  $Z$  due variabili aleatorie e  $f$  una funzione  $f : R \rightarrow R$  e sia

$$Z = f(X).$$

Se conosciamo la densità di probabilità per la variabile  $X$  possiamo ricavare facilmente quella per  $Z$ :

$$\rho_Z(z) = \int \rho_X(x) \delta(z - f(x)) dx = \sum_i \frac{1}{|f'(x_i(z))|} \rho_X(x_i(z)) \quad (1.28)$$

dove  $\{x_i(z)\}$  sono le soluzioni dell'equazione  $z = f(x)$  nella variabile  $x$  a  $z$  fissata. Occorre tener conto del fatto che, se  $f(x)$  non è monotona, possono esistere più soluzioni. Nel caso  $f(x)$  sia monotona la trasformazione nella densità può essere ottenuta anche dalla identità

$$\int \rho_Z(z) dz = 1 = \int \rho_X(x) dx = \int \left| \frac{dx}{dz} \right| \rho_X(x(z)) dz,$$

da cui

$$\rho_Z(z) = \left| \frac{dx}{dz} \right| \rho_X(x(z)). \quad (1.29)$$

In più dimensioni la relazione è analoga. Sia  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vettore di variabili aleatorie e

$$Z_i = Z_i(\mathbf{X}) \quad i = 1, \dots, n$$

un altro vettore di variabili aleatorie funzioni delle precedenti. La nuova densità di probabilità sarà data da

$$\rho_{\mathbf{Z}}(\mathbf{z}) = \left| \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(z_1, z_2, \dots, z_n)} \right| \rho_{\mathbf{X}}(\mathbf{x}(\mathbf{z})). \quad (1.30)$$

**Esempio 1:** Sia  $X$  una variabile aleatoria distribuita uniformemente in  $[0, 1]$ .

$$\rho_X(x) = 1 \quad \text{per } x \in [0, 1]$$

e

$$Z = -\ln X.$$

Troviamo subito che  $Z$  è distribuita secondo la distribuzione esponenziale:

$$\rho_Z(z) = \int_0^1 \delta(z + \ln x) dx = \int_0^1 \left| \frac{dz}{dx} \right|^{-1} \delta(x - e^{-z}) dx = e^{-z}.$$

Per una generica distribuzione si ha:

$$\rho_Z(z) = e^{-z} \rho_X(x(z)).$$

**Esempio 2:** Date due variabili aleatorie  $X_1$  e  $X_2$  indipendenti e distribuite uniformemente in  $[0, 1]$ , che possiamo considerare come componenti di un vettore  $\mathbf{X}$ , consideriamo la trasformazione

$$\mathbf{Z} = (Z_1, Z_2) = \sqrt{-2 \ln X_1} (\cos 2\pi X_2, \sin 2\pi X_2) \quad (1.31)$$

Poichè  $Z_1^2 + Z_2^2 = -2 \ln X_1$ , troviamo facilmente:

$$\frac{\partial(z_1, z_2)}{\partial(x_1, x_2)} = -\frac{2\pi}{x_1} = -\frac{2\pi}{e^{-\frac{z_1^2+z_2^2}{2}}}$$

e infine

$$\rho_{\mathbf{Z}}(\mathbf{z}) = \left| \frac{\partial(x_1, x_2)}{\partial(z_1, z_2)} \right| \rho_{\mathbf{X}}(\mathbf{x}(\mathbf{z})) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_2^2}{2}}.$$

Quindi  $Z_1$  e  $Z_2$  sono anche esse variabili aleatorie indipendenti, ma distribuite secondo la distribuzione normale. Questa trasformazione è alla base dell'algoritmo di Box-Muller per la generazione di numeri casuali distribuiti in modo gaussiano che vedremo più avanti.

### 1.6.3 Somma di variabili aleatorie

Questo argomento è stato già affrontato in parte quando abbiamo studiato varianza e covarianza per coppie di variabili aleatorie nel paragrafo 1.4. Siano  $X_1$  e  $X_2$  due variabile aleatorie con densità di probabilità congiunta  $\rho(x_1, x_2)$ . Consideriamo la nuova variabile

$$Z = X_1 + X_2$$

Avremo

$$\rho_Z(z) = \int \delta(z - x_1 - x_2) \rho(x_1, x_2) dx_1 dx_2 = \int \rho(x_1, z - x_1) dx_1. \quad (1.32)$$

Nel caso di due variabili con distribuzione gaussiana, possiamo notare che l'integrando  $\rho(x_1, z - x_1)$  è un esponenziale avente per esponente una forma quadratica in  $x_1$  e  $z$ . Ne risulta che anche l'integrale sarà una funzione esponenziale con esponente quadratico in  $z$  e quindi anche  $z$  è una variabile aleatoria distribuita normalmente.

**ESERCIZIO:** Dimostrare che, se  $X_1$  e  $X_2$  sono due variabile aleatorie con uguale densità di probabilità gaussiana avente attesa  $\mu$  e varianza  $\sigma^2$ , la loro somma è anch'essa distribuita gaussianamente con attesa  $2\mu$  e varianza  $2\sigma^2$ .

**Esempio:** Trovare la densità di probabilità per la somma di due variabili aleatorie indipendenti distribuite uniformemente in  $[0, 1]$ .

Dette  $X_1$  e  $X_2$  le due variabili uniformemente distribuite abbiamo

$$\rho(x_1, x_2) = \rho(x_1) \rho(x_2)$$

con

$$\rho(x_i) = \begin{cases} 1 & \text{per } 0 \leq x_i \leq 1, \\ 0 & \text{altrove.} \end{cases}$$

Per

$$Z = X_1 + X_2$$

avremo

$$\rho_Z(z) = \int \rho(x_1, z - x_1) dx_1 = \int \rho(x_1) \rho(z - x_1) dx_1.$$

dove

$$\rho(z - x_1) = \begin{cases} 1 & \text{per } 0 \leq z - x_1 \leq 1 \text{ cioè per } z - 1 \leq x_1 \leq z, \\ 0 & \text{altrove.} \end{cases}$$

L'integrale, a fissato  $z$ , va fatto sull'intersezione tra i due intervalli

$$\begin{aligned} z - 1 &\leq x_1 \leq z \\ 0 &\leq x_1 \leq 1 \end{aligned}$$

Risulta quindi che

$$\rho_Z(z) = \begin{cases} \int_0^z dx_1 = z & \text{per } 0 \leq z \leq 1, \\ \int_{z-1}^1 dx_1 = 2 - z & \text{per } 1 \leq z \leq 2. \end{cases}$$

$Z$  è, quindi, distribuita secondo una densità triangolare.

#### 1.6.4 Prodotto di variabili aleatorie

Siano  $X_1$  e  $X_2$  due variabile aleatorie con densità di probabilità congiunta  $\rho(x_1, x_2)$ . Consideriamo la nuova variabile

$$Z = X_1 X_2$$

Avremo

$$\rho_Z(z) = \int \delta(z - x_1 x_2) \rho(x_1, x_2) dx_1 dx_2 = \int \frac{1}{|x_1|} \rho(x_1, \frac{z}{x_1}) dx_1. \quad (1.33)$$

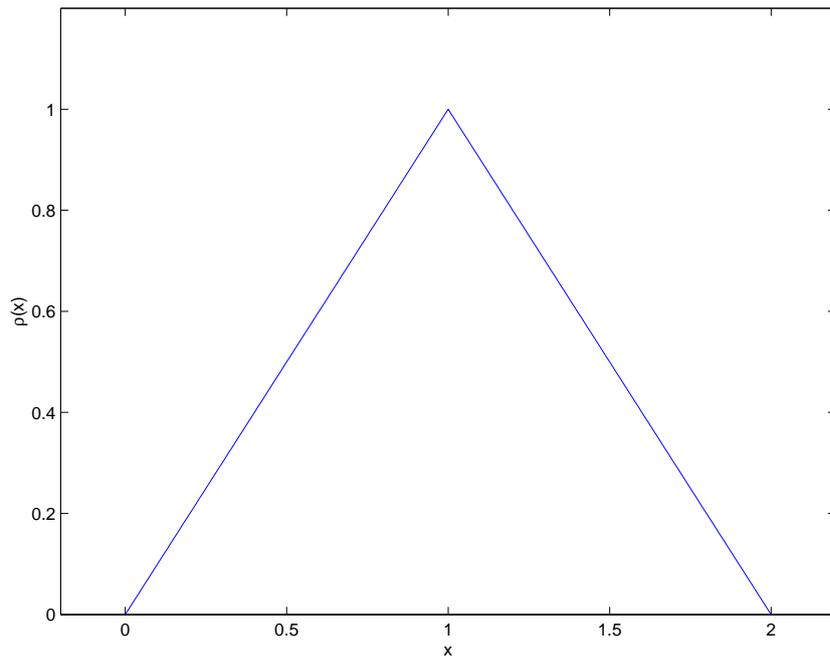


Figura 1.7: La distribuzione relativa alla somma di due variabili distribuite uniformemente

## 1.7 Teoremi limite

### 1.7.1 Disuguaglianza di Chebyshev

Sia  $X$  una variabile aleatoria distribuita con densità di probabilità  $\rho_X(x)$  avente attesa  $\mu$  e varianza finita  $\sigma^2$ . La disuguaglianza di Chebyshev si ricava dalla definizione stessa di varianza. Fissato  $k \geq 1$  si ha infatti

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{+\infty} (x - \mu)^2 \rho_X(x) dx \geq \\
 &\geq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 \rho_X(x) dx + \int_{\mu + k\sigma}^{+\infty} (x - \mu)^2 \rho_X(x) dx \geq \\
 &\geq k^2 \sigma^2 \left[ \int_{-\infty}^{\mu - k\sigma} \rho_X(x) dx + \int_{\mu + k\sigma}^{+\infty} \rho_X(x) dx \right] = \\
 &= k^2 \sigma^2 P\{|X - \mu| \geq k\sigma\}
 \end{aligned}$$

che comporta

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}.$$

Posto  $\epsilon = k\sigma$  ( $\epsilon \geq \sigma$ ) avremo

$$P\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$$

e quindi

$$P\{|X - \mu| \leq \epsilon\} > 1 - \frac{\sigma^2}{\epsilon^2}. \quad (1.34)$$

Questa è la disuguaglianza di Chebyshev. Il suo significato è che è sempre possibile fissare un  $\epsilon \geq \sigma$  abbastanza grande da poter rendere più grande di  $1 - \sigma^2/\epsilon^2$  la probabilità che la variabile  $X$  assuma un valore tra  $\mu - \epsilon$  e  $\mu + \epsilon$ .

### 1.7.2 Legge dei Grandi Numeri

Consideriamo  $N$  realizzazioni indipendenti di una variabile aleatoria  $X$  avente valore d'attesa  $\mu$  e varianza  $\sigma^2$  e chiamiamo  $\bar{Y}_N$  la variabile aleatoria media di tali realizzazioni

$$\bar{Y}_N = \frac{X_1 + X_2 + \dots + X_N}{N}. \quad (1.35)$$

Estendendo quanto abbiamo visto nel caso di due variabili nel paragrafo 1.4, possiamo affermare che  $\bar{Y}_N$  ha attesa  $\mu$  e varianza  $\sigma^2/N$ .

Dalla disuguaglianza di Chebyshev deduciamo che una particolare realizzazione di  $\bar{Y}_N$  cadrà fuori dell'intervallo  $\mu - \epsilon, \mu + \epsilon$  con probabilità inferiore a  $\sigma^2/(N\epsilon^2)$ . In questo modo si può sempre pensare di scegliere un valore di  $N$  abbastanza grande da rendere abbastanza prossima ad uno la probabilità di trovare  $\bar{Y}_N$  vicina al suo valore d'attesa quanto desiderato. Posso quindi affermare che, comunque si fissi  $\epsilon > 0$ ,

$$\lim_{N \rightarrow \infty} P\{|\bar{Y}_N - \mu| \geq \epsilon\} = 0 \quad (1.36)$$

Questa relazione è nota come *Legge dei grandi numeri* (o, almeno, è una delle sue tante versioni). Notiamo che è essenziale per la sua dimostrazione che la varianza di  $X$  sia finita e, inoltre, che non è stata fatta alcuna ipotesi sulla natura della densità di probabilità.

In pratica la Legge dei grandi numeri afferma che *la media dei campionamenti converge, in senso stocastico, al valore d'attesa della popolazione quando la dimensione del campione aumenta*. Per questo motivo  $\bar{Y}_N$  è detta *stima consistente* del valore d'attesa  $\mu$ .

### 1.7.3 Teorema del Limite Centrale

Siano  $X_1, X_2, \dots, X_N$ ,  $N$  variabili aleatorie indipendenti e identicamente distribuite secondo una comune densità di probabilità gaussiana con valore d'attesa 0 e varianza  $\sigma^2$

$$\rho_{X_i}(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x_i^2}{2\sigma^2}\right\}$$

La variabile

$$Y_N = X_1 + X_2 + \cdots + X_N$$

è distribuita normalmente con valore d'attesa 0 e varianza  $N\sigma^2$

$$\rho_{Y_N}(x) = \frac{1}{\sigma\sqrt{2\pi N}} \exp\left\{-\frac{x^2}{2N\sigma^2}\right\}$$

mentre la variabile

$$\bar{Y}_N = \frac{Y_N}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

è distribuita normalmente con valore d'attesa 0 e varianza  $\sigma^2/N$ . Oltre che dalle proprietà delle distribuzioni gaussiane, quest'ultimo risultato può essere ricavato ricordando (paragrafo 1.6.1), che, se  $Z = \alpha X$ , allora  $\rho_Z(z) = \frac{1}{|\alpha|} \rho_X\left(\frac{z}{\alpha}\right)$  e quindi

$$\rho_{\bar{Y}_N}(x) = N\rho_{Y_N}(Nx) = \frac{\sqrt{N}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{Nx^2}{2\sigma^2}\right\}$$

Si comprende facilmente che, per avere una variabile che è proporzionale alla somma e la cui varianza è uguale alla varianza della distribuzione delle  $X_i$ , occorre introdurre, come abbiamo visto per la distribuzione binomiale, la variabile  $Y_N/\sqrt{N}$

$$\rho_{Y_N/\sqrt{N}}(x) = \sqrt{N}\rho_{Y_N}(\sqrt{N}x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

$Y_N/\sqrt{N}$  ha esattamente stesso valore d'attesa nullo e stessa varianza della variabile  $X$  di partenza.

Il *Teorema del Limite Centrale* asserisce che, anche quando la densità di probabilità comune alle  $N$  variabili non è quella gaussiana, ma un'altra densità arbitraria con valore d'attesa 0 e varianza finita, i risultati che abbiamo visto sono ancora validi nel limite  $N \rightarrow \infty$ , cioè la densità limite di  $Y_N/\sqrt{N}$  è ancora una densità gaussiana con valore d'attesa nullo e varianza uguale a quella di partenza.

Il Teorema del Limite Centrale conserva la sua validità anche sotto ipotesi più generali:

1. Non è necessario che le variabili aleatorie siano identicamente distribuite. Infatti se si considerano  $N_1$  variabili aleatorie che hanno una distribuzione comune e  $N_2$  che ne hanno un'altra le distribuzioni delle loro somme tendono per  $N_1, N_2 \rightarrow \infty$  verso due distribuzioni gaussiane. Ma, per la proprietà delle distribuzioni gaussiane anche la somma delle  $N_1 + N_2$  variabili tende per  $N_1, N_2 \rightarrow \infty$  ad una distribuzione gaussiana.
2. Il teorema può essere dimostrato anche se le variabili aleatorie sono debolmente dipendenti.

### 1.7.4 Cenni sulle distribuzioni di Levy

Nel Teorema del Limite Centrale è essenziale l'ipotesi che la distribuzione comune alle variabili considerate abbia una varianza finita. Ci si può chiedere se il teorema sia suscettibile di una generalizzazione al caso di distribuzioni che, come quella di Cauchy, non hanno varianza finita. Una tale generalizzazione è dovuta a Paul Levy.

Abbiamo visto che, se due variabili sono uniformemente distribuite, la loro somma è distribuita secondo una distribuzione diversa, la distribuzione triangolare. Al contrario, se due variabili sono distribuite normalmente, anche la loro somma lo è. Si dice, a tale proposito che la distribuzione gaussiana è *stabile* per operazione di somma. Se questo non accadesse, la distribuzione gaussiana non potrebbe mai essere una distribuzione limite per la somma di variabili aleatorie.

Levy ha individuato una classe generale di distribuzioni che godono della proprietà di stabilità, e delle quali la distribuzione gaussiana è un caso particolare. Non esiste una forma generale che descrive le distribuzioni di Levy; essa esiste invece per la sua trasformata di Fourier:

$$\tilde{L}(k; \alpha) = e^{-D|k|^\alpha}$$

dove  $D$  è un fattore di scala positivo,  $k$  la variabile coniugata di  $x$  e  $\alpha$  un parametro detto *indice di Levy*.

La trasformata di Fourier inversa è la distribuzione di Levy  $L(x; \alpha)$ . Si può dimostrare che, perchè essa sia non negativa, occorre che  $0 < \alpha \leq 2$  e, per essa, si trova il seguente comportamento

$$L(x; \alpha) \sim \begin{cases} D^{-1/\alpha} & \text{per } x = 0, \\ D|x|^{-\alpha-1} & \text{per } |x| \rightarrow \infty, \alpha < 2, \\ D^{-1/2}e^{-x^2/4D} & \text{per } \alpha = 2. \end{cases}$$

Vediamo quindi che la distribuzione gaussiana e la distribuzione di Cauchy sono particolari distribuzioni di Levy corrispondenti rispettivamente a  $\alpha = 2$  e  $\alpha = 1$ .

Il *Teorema di Levy* generalizza il Teorema del Limite Centrale: *Le distribuzioni di Levy sono le sole possibili distribuzioni limite per la somma di variabili aleatorie indipendenti identicamente distribuite.*

Il Teorema di Levy comprende anche le distribuzioni a varianza infinita, mentre il Teorema del Limite Centrale contempla solo le distribuzioni a varianza finita e corrisponde al caso  $\alpha = 2$ .



# Capitolo 2

## Campionamento casuale e metodo Monte Carlo

### 2.1 Generazione di numeri casuali uniformemente distribuiti

Consideriamo un esperimento, lo spazio degli eventi ad esso associato ed una variabile aleatoria relativa a tale spazio con la sua distribuzione di probabilità. Esula da questa trattazione la vasta problematica su come dei processi aleatori possano insorgere in sistemi fisici macroscopici che si crede obbediscano a leggi deterministiche. Da una serie di misure possiamo ricavare una sequenza di numeri casuali, cioè la sequenza dei valori assunti dalla variabile aleatoria. Si ipotizza che questa sequenza sia imprevedibile e, quindi, non riproducibile. Tipici processi fisici che possiamo considerare a tale fine sono i decadimenti radioattivi, il rumore termico in un circuito elettronico, i tempi di arrivo di raggi cosmici, ... . In questo capitolo consideriamo il problema di ottenere da un esperimento una sequenza di numeri casuali uniformemente distribuiti.

#### 2.1.1 Numeri veramente casuali e pseudo - casuali

Il problema più importante che occorre affrontare per essere sicuri di ottenere dei numeri veramente casuali è quello di eliminare possibili distorsioni introdotte dal metodo di misura. Un esempio classico di tecniche destinate a tale fine è quello introdotto da Frigerio e Clark (1975) [2]. Essi considerarono una sorgente di particelle  $\alpha$  e un contatore ad alta risoluzione che misurava il numero di decadimenti avvenuti ogni 20 ms. Per ogni conteggio (in media 24,315) veniva memorizzato uno 0 se tale numero era pari e un 1 se era dispari. Si otteneva così una sequenza di bits, cioè di 0 e di 1. Per eliminare una distorsione introdotta dal fatto che la probabilità  $p_0$  di un conteggio pari poteva non essere esattamente uguale alla probabilità  $p_1$  di un conteggio dispari, si procedeva nel modo seguente. I numeri

venivano considerati in coppie e tutte le coppie di numeri uguali venivano eliminate. Delle coppie di numeri diversi si usava solo il secondo. A questo punto le probabilità di che questo fosse 0 era  $p'_0 = p_1 p_0$ , la stessa probabilità che fosse 1,  $p'_1 = p_0 p_1$ . Da questa successione di bit si era in grado di ricavare una successione di 0 e 1, che, raggruppati secondo un fissato numero di bits, davano origine ad una successione di numeri casuali uniformemente distribuiti in forma binaria, o decimale. Questo algoritmo fu in grado di generare 6000 numeri casuali di 31 bits all'ora. Presso l'Argonne National Laboratory fu registrata una sequenza di 2.5 milioni di essi (<http://www.nea.fr/abs/html/nesc0843.html>).

Si tratta di sequenze di *numeri veramente casuali*, tuttavia il loro uso su calcolatore per generare eventi aleatori complessi è sconsigliato perchè il continuo accesso alla memoria di massa allunga considerevolmente i tempi dei calcoli.

Si preferisce pertanto rinunciare ai numeri veramente casuali ed utilizzare i cosiddetti *numeri pseudo-casuali*, che sono invece numeri generati da algoritmi matematici veloci e che consumano poca memoria. Poichè si tratta di sequenze generate in maniera deterministica, i numeri pseudo-casuali sono esattamente riproducibili e predicibili. *Il requisito diventa allora che tali sequenze siano statisticamente indistinguibili da una sequenza di numeri veramente casuali.* Questo vuol dire che i generatori di numeri pseudo-casuali, se sottoposti a test di casualità, devono dare dei risultati vicini a quelli che ci si aspetta da dei numeri veramente casuali. Prima di passare a descrivere alcuni di questi test, premettiamo che, d'ora in avanti, ove non diversamente specificato, riserveremo il termine di *numeri pseudo-casuali* (o *pseudo-random*) a numeri generati da un algoritmo matematico che si suppongono indipendenti e uniformemente distribuiti nell'intervallo  $(0, 1)$ .

### 2.1.2 Test di casualità

Un test di casualità consiste nel costruire una funzione  $\psi(r_1, r_2, \dots)$ , con  $r_1, r_2, \dots$  variabili aleatorie indipendenti, calcolare questa funzione per una sequenza di numeri pseudo-random, e paragonare il valore ottenuto con quello atteso per una sequenza di numeri veramente casuali uniformemente distribuiti in  $(0, 1)$ .

**Esempio 1:** Consideriamo la funzione

$$\psi(r_1, r_2, \dots, r_N) = \frac{1}{N} \sum_{i=1}^N r_i$$

che dà l'attesa di  $N$  variabili aleatorie che come si è detto, supporremo uniformemente distribuite in  $(0, 1)$ . Notiamo che per  $N$  grande,  $\psi$ , per il Teorema del Limite Centrale, tende ad essere distribuita gaussianamente con media 0.5 e varianza  $\sigma^2 = (12N)^{-1}$ . Se consideriamo l'intervallo  $(0.5 - 2\sigma, 0.5 + 2\sigma)$ , detto intervallo di confidenza di due sigma, la probabilità di trovare  $\psi$  in questo intervallo è uguale a 0.95. Se una sequenza di  $N$  numeri pseudo-casuali ha una media

che cade al di fuori di questo intervallo si dice che essa fallisce il test al livello del 5%.

**ESERCIZIO:** Verificare sul proprio computer che le sequenze generate tramite il generatore di numeri pseudo-casuali di libreria passano il test al livello del 5%.

**Esempio:** Un altro test consiste nello studio di

$$\psi(r_1, r_2, \dots, r_N) \equiv C(k) = \frac{N \sum_{i=1}^N r_i r_{i+k} - \left( \sum_{i=1}^N r_i \right)^2}{N \sum_{i=1}^N r_i^2 - \left( \sum_{i=1}^N r_i \right)^2}$$

che è detta anche funzione di autocorrelazione a due punti. È facile vedere che, per una sequenza di veri numeri casuali, si ha  $C(0) = 1$ , mentre per  $k \neq 0$  si ha

$$C(k) = \frac{\langle r_i r_{i+k} \rangle - \langle r_i \rangle^2}{\langle r_i^2 \rangle - \langle r_i \rangle^2} = \frac{\langle r_i \rangle \langle r_{i+k} \rangle - \langle r_i \rangle^2}{\langle r_i^2 \rangle - \langle r_i \rangle^2} = 0,$$

dato che, essendo le variabili indipendenti,  $\langle r_i \rangle = \langle r_{i+k} \rangle$ . In definitiva, quindi,

$$C(k) = \delta_{k,0}$$

**ESERCIZIO:** Effettuare sul computer il test di autocorrelazione sul generatore di numeri pseudo-casuali di libreria.

Come ben si comprende, vi è un numero indefinitamente grande di modi di costruire funzioni di test. Quindi in linea di principio un generatore di numeri pseudo-casuali non può mai essere completamente validato e può sempre riservare sorprese.

### 2.1.3 Generatori di numeri pseudo-casuali

Il primo generatore di numeri casuali è stato il *generatore di von Neumann*. La generazione avviene nel modo seguente: si parte da un numero intero di  $2m$  cifre  $\nu_1$  e se ne considerano le  $m$  cifre centrali, ottenendo così il numero  $\eta_1$ ; si quadra poi  $\eta_1$  e si ottiene così l'intero di  $2m$  cifre  $\nu_2$ ; si considerano le  $m$  cifre centrali e si ottiene  $\eta_2$ . Proseguendo in questo modo si ottiene una sequenza di interi  $r_1, r_2, \dots$  e, dividendo ciascuno di essi per  $10^m$ , si ottiene una sequenza di numeri reali (razionali)  $\eta_1, \eta_2, \dots$  nell'intervallo  $(0, 1)$ . Questo generatore ha mostrato diversi problemi e riveste più che altro importanza storica.

Il generatore più usato oggi è il *Generatore Congruenziale Lineare* (GCL) o di *Lehmer*. La sequenza viene generata a partire da un numero a piacere  $r_1$ , detto *seme* e si generano i successivi numeri secondo la formula

$$r_{i+1} = (ar_i + b) \bmod m \quad (2.1)$$

dove  $a$ ,  $b$  e  $m$  sono parametri interi della generazione.  $r_{i+1}$  è quindi il resto della divisione nel campo degli interi fra  $ar_i + b$  e  $m$ . Il massimo numero ottenibile è  $m - 1$ , perciò i numeri

$$\eta_i = \frac{r_i}{m}$$

sono dei numeri pseudo-casuali compresi tra 0 e 1 (1 escluso).

Questo generatore è usato da una cinquantina di anni e le sue proprietà sono ben note ed apprezzate. Tuttavia occorre notare che i parametri  $a$ ,  $b$  e  $m$  devono essere scelti con attenzione.

**Esempio 1:** Generare con il metodo di Lehmer una sequenza di numeri pseudo-casuali con  $a = 5$ ,  $b = 1$ ,  $m = 100$  a partire da  $r_1 = 1$ .

Avremo

$$\begin{aligned} r_1 &= 1 \\ r_2 &= (5 \times 1 + 1) \bmod 100 = 6 \\ r_3 &= (5 \times 6 + 1) \bmod 100 = 31 \\ r_4 &= (5 \times 31 + 1) \bmod 100 = 56 \\ r_5 &= (5 \times 56 + 1) \bmod 100 = 81 \\ r_6 &= (5 \times 81 + 1) \bmod 100 = 6 \\ r_7 &= r_3 = 31 \\ r_8 &= r_4 = 56 \\ &\vdots = \vdots \end{aligned}$$

Abbiamo ottenuta una sequenza con periodo 4 ed è un po' difficile in questo caso parlare di una sequenza di numeri pseudo-casuali.

In generale una buona scelta dei parametri consente di avere un periodo dell'ordine di  $m$ , che è il massimo periodo possibile. Per massimizzare a sua volta  $m$  si considera un valore pari a  $2^{t-1} - 1$ , dove  $t$  è il numero di bits usato

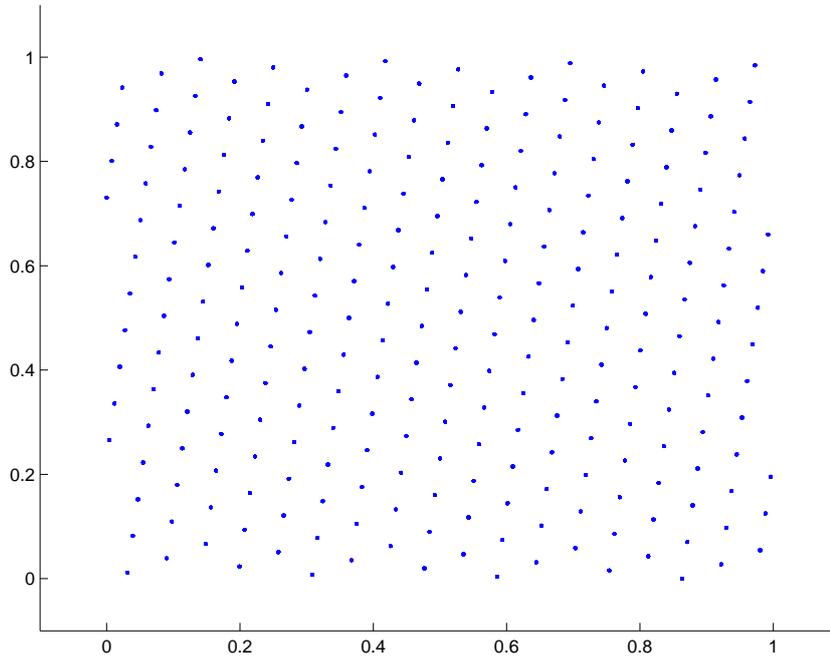


Figura 2.1: Figura di Marsaglia generata a partire da 10000 numeri ottenuti dal Generatore Lineare Congruenziale con parametri  $a = 137$ ,  $b = 187$ ,  $m = 256$

per gli interi dal calcolatore in uso ( $2^t - 1$  sarebbe il massimo intero possibile, ma un bit è riservato al segno). Per processori oggi standard a 32 bits si ha  $M = 2^{31} - 1 = 2147483647$ . In questo caso una buona scelta per  $a$  e  $b$  è  $a = 7^5 = 16801$  e  $b = 0$ .

Un'altra problematica è legata alle ricerche di Marsaglia. Se si guardano le  $n$ -uple di numeri pseudo-casuali ottenuti da un Generatore Lineare Congruenziale

$$(\eta_1, \eta_2, \dots, \eta_n), (\eta_2, \eta_3, \dots, \eta_{n+1}), (\eta_3, \eta_4, \dots, \eta_{n+2}), \dots$$

come coordinate di punti nell'ipercubo di lato unitario a  $n$  dimensioni, i punti risultanti si trovano in un numero relativamente piccolo di iperpiani paralleli. Il numero di iperpiani di Marsaglia non eccede  $(n!2^t)^{1/n}$ .

**ESERCIZIO:** Costruire le figure di Marsaglia in uno spazio bidimensionale usando i valori  $a = 137$ ,  $b = 187$ ,  $m = 256$  (vedi fig. 2.1).

È chiaro che questo fenomeno può dare distorsioni non volute nei calcoli di Monte Carlo.

Un altro problema riguarda la presenza di correlazioni nella sequenza. Quanto influiscano questi problemi dipende dal tipo di calcolo. In linea generale si può affermare che le cose vanno bene se il Monte Carlo usa soltanto una piccola

frazione del periodo completo della sequenza. Per evitare la struttura di Marsaglia sono stati proposti generatori non lineari. Alcune direttrici di ricerca sono rivolte verso le mappe caotiche, che sono sempre algoritmi deterministici, ma privi di predicibilità a lungo termine.

## 2.2 Campionamento casuale uniforme e metodo Monte Carlo

### 2.2.1 Integrazione Monte Carlo

L'esistenza di generatori uniformi di numeri casuali o pseudo-casuali permette di dare una prima risposta al problema di usare tecniche stocastiche per l'integrazione di funzioni.

Supponiamo di voler valutare il seguente integrale

$$I = \int_a^b f(x) dx \quad (2.2)$$

dove  $f(x)$  è una funzione reale di variabile reale, ovviamente integrabile in  $[a, b]$ , con  $b > a$ . Consideriamo la seguente funzione

$$g(x) = \begin{cases} \frac{1}{b-a} & \text{per } a \leq x \leq b, \\ 0 & \text{altrove.} \end{cases}$$

L'integrale può essere riscritto nella forma

$$I = (b - a) \int_a^b f(x) g(x) dx$$

$g(x)$  rappresenta una distribuzione di probabilità uniforme in  $[a, b]$ ; possiamo quindi considerare l'integrale come valore di aspettazione di  $f(x)$ , con  $x$  variabile aleatoria distribuita secondo  $g(x)$ . Possiamo allora utilizzare la Legge dei Grandi Numeri e utilizzare come stimatore consistente dell'attesa  $\langle F \rangle$  di  $f(x)$  la grandezza

$$\bar{F} = \frac{1}{N} \sum_{i=1}^N f(x_i)$$

dove gli  $\{x_i\}$  costituiscono una sequenza di numeri campionati in maniera uniforme nell'intervallo  $[a, b]$  costruita da una sequenza  $\{\eta_i\}$  di numeri campionati in maniera uniforme nell'intervallo  $[0, 1]$ .

$$x_i = (b - a)\eta_i + a.$$

Otteniamo così uno stimatore consistente per l'integrale cercato:

$$I = \langle F \rangle \times (b - a) \simeq \bar{F} \times (b - a)$$

Questa tecnica richiede che  $a$  e  $b$  siano finiti, ma può essere utilizzata, tramite opportuno cambiamento di variabili, anche per intervalli d'integrazione infiniti. Come tutte le altre tecniche Monte Carlo, questo metodo consente di integrare funzioni che hanno punti di discontinuità (in numero finito), e, teoricamente, consente anche di integrare funzioni con singolarità (ovviamente integrabili). Tuttavia il metodo è efficiente solo per funzioni che non variano molto sull'intervallo di integrazione. Infatti, poichè tale intervallo viene campionato in maniera uniforme, si rischia di non esplorare in maniera più dettagliata delle zone, magari piccole, dove la funzione assume valori grandi e che, quindi, danno un grande contributo all'integrale. In questi casi sono preferibili tecniche Monte Carlo più raffinate che vedremo in seguito. Per funzioni non particolarmente patologiche, questa tecnica dà dei risultati abbastanza affidabili, anzi può essere facilmente generalizzata a domini in più dimensioni, con qualche cautela, come vedremo nel prossimo paragrafo.

### 2.2.2 Integrazione su un dominio triangolare

Supponiamo di voler calcolare

$$I = \int_0^1 dx \int_0^x dy f(x, y)$$

usando il metodo precedentemente esposto. Il dominio di integrazione è un triangolo di area  $\frac{1}{2}$ , quindi la distribuzione uniforme

$$g(x, y) = \begin{cases} 2 & \text{se } x \text{ nel triangolo,} \\ 0 & \text{altrove.} \end{cases}$$

L'integrale può essere riscritto nella forma:

$$I = \frac{1}{2} \int_0^1 dx \int_0^x dy f(x, y) g(x, y)$$

A tal fine occorrerà campionare in maniera uniforme il triangolo rettangolo con cateto unitario sull'asse  $x$  e avente per ipotenusa la diagonale mostrato in figura 2.2. Consideriamo quattro metodi diversi.

**Metodo 1:** Si può pensare di applicare in maniera pedissequa il metodo appena visto nel caso unidimensionale:

- a) si estrae un numero pseudo-casuale  $x_i$  tra 0 e 1;
- b) si estrae un numero pseudo-casuale  $y_i$  tra 0 e  $x_i$ ;

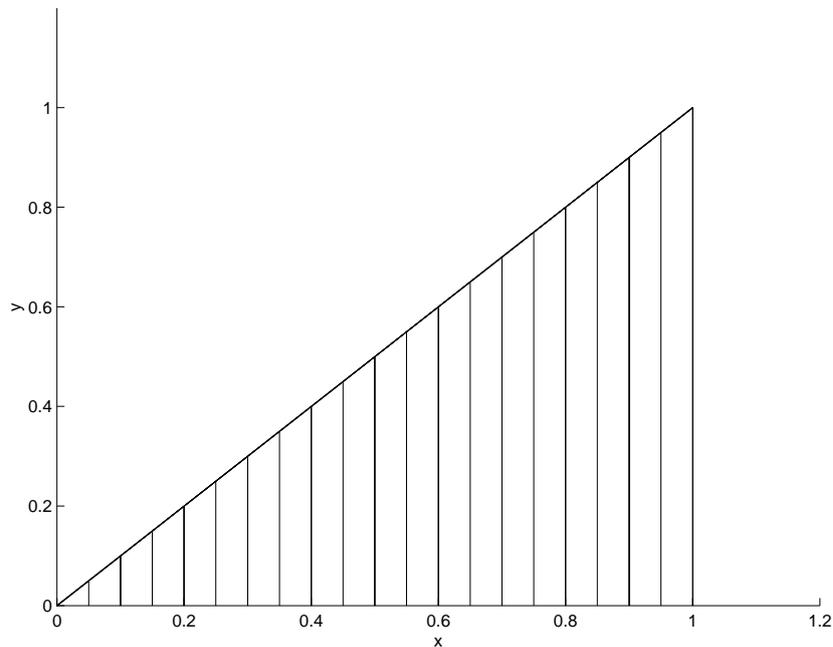


Figura 2.2: Dominio di integrazione triangolare

- c) si ripetono  $N$  volte i passi a) e b) calcolando ogni volta  $f(x_i, y_i)$ ;  
 d) si calcola

$$I = \frac{1}{2} \frac{1}{N} \sum_{i=1}^N f(x_i, y_i)$$

Questo metodo è **sbagliato**. Infatti, è facile capire che verranno generati con maggiore densità punti aventi coordinata  $x_i$  piccola.

**Metodo 2:** Per correggere si possono generare punti nel quadrato di lato 1 ed eliminare quelli contenuti nel triangolo superiore.

- a) si estrae un numero pseudo-casuale  $x_i$  tra 0 e 1;  
 b) si estrae un numero pseudo-casuale  $y_i$  tra 0 e 1;  
 c) se  $y_i > x_i$  la coppia viene scartata;  
 d) si ripetono i passi a), b) e c) e si calcola  $f(x_i, y_i)$  soltanto per i punti accettati.  
 e) Detto  $N$  il numero di valori accettati si calcola

$$I = \frac{1}{2} \frac{1}{N} \sum_{i=1}^N f(x_i, y_i)$$

Questo metodo è corretto perchè i punti vengono generati uniformemente nel triangolo, ma solo una metà circa dei punti generati sono utilizzati.

**Metodo 3:** Questo è un primo metodo per evitare gli scarti. Esso viene detto *metodo del piegamento*.

- a) si estraggono due numeri pseudo-casuali  $r_1$  e  $r_2$  tra 0 e 1;
- b) si pone  $x_i = \max(r_1, r_2)$  e  $y_i = \min(r_1, r_2)$ ;
- c) si ripetono  $N$  volte i passi a) e b) calcolando ogni volta  $f(x_i, y_i)$ ;
- d) si calcola

$$I = \frac{1}{2} \frac{1}{N} \sum_{i=1}^N f(x_i, y_i)$$

Questo metodo corrisponde a generare punti uniformemente distribuiti nel quadrato e a piegarlo poi lungo la diagonale in modo da generare uniformemente nel triangolo. È chiaramente un metodo molto efficace; i punti generati sono tutti utilizzati, anche se è necessario valutare quale delle due coordinate è la  $x$  e quale la  $y$ .

**Metodo 4:** Come abbiamo detto il primo metodo sopravvalutava la zona dei piccoli  $x$ . Quest'altro metodo tende a correggere il precedente introducendo una funzione peso compensativa.

- a) si estrae un numero pseudo-casuale  $x_i$  tra 0 e 1;
- b) si estrae un numero pseudo-casuale  $y_i$  tra 0 e  $x_i$ ;
- c) si ripetono  $N$  volte i passi a) e b) calcolando ogni volta  $f(x_i, y_i)$ ;
- d) si calcola

$$I = \frac{1}{N} \sum_{i=1}^N x_i f(x_i, y_i)$$

I punti generati nelle zone con maggiore densità vengono penalizzati dal fattore  $x_i$ . Notiamo infatti che possiamo riscrivere l'integrale  $I$  nella forma

$$I = \int_0^1 dx x \int_0^x dy g(y) f(x, y)$$

dove il fattore  $x$  tiene conto del fatto che, per una corretta normalizzazione della funzione  $g(y)$ , distribuzione uniforme della coordinata  $y$ , occorre porre

$$g(y) = \begin{cases} \frac{1}{x} & \text{per } 0 \leq y \leq x, \\ 0 & \text{altrove.} \end{cases}$$

È opportuno notare che quanto abbiamo ora esposto può essere utilizzato per l'integrazione su un triangolo non solo quando si utilizza il Monte Carlo rozzo

visto nel paragrafo precedente, ma anche per metodi più raffinati, come quello del rigetto, che vedremo in seguito.

## 2.3 Campionamento d'importanza e metodo Monte Carlo

Le tecniche di campionamento d'importanza hanno lo scopo di convertire una sequenza di numeri casuali  $\{\eta_i\}$  (nella pratica numeri pseudo-casuali) uniformemente distribuiti nell'intervallo  $(0, 1)$  in una sequenza  $x_i$  di numeri casuali distribuiti secondo una densità di probabilità  $f(x)$ . Vi sono numerose tecniche per attuare questa conversione. Le più comuni e di uso più generale sono il metodo del *Rigetto* e quello dell'*Inversione*. Per particolari distribuzioni si possono sfruttare le proprietà di composizione e di trasformazione delle variabili aleatorie. Ad esempio, per generare dei numeri distribuiti secondo una distribuzione triangolare, basterà generare delle coppie di numeri dalla distribuzione uniforme e sommarli (par. 1.6.3).

### 2.3.1 Metodo del Rigetto: distribuzione uniforme, distribuzioni non uniformi

L'idea alla base di questo metodo è molto semplice: dato un set di numeri casuali, ne vogliamo eliminare una parte in maniera che i rimanenti siano distribuiti secondo una certa densità di probabilità.

Sia  $f(x)$  la distribuzione desiderata, definita sull'intervallo  $(a, b)$  e sia  $f_{max}$  il massimo valore che essa assume in tale intervallo, o, per lo meno, un reale tale che

$$f_{max} > f(x) \quad \forall x \in (a, b).$$

Consideriamo le due sequenze di numeri uniformemente distribuiti in  $(0, 1)$ ,  $\{\eta_i\}$  e  $\{\bar{\eta}_i\}$ . Procediamo nel modo seguente:

1. dagli  $\{\eta_i\}$  costruiamo una sequenza  $\{x_i\}$  di numeri distribuiti uniformemente in  $(a, b)$ ,

$$x_i = (b - a)\eta_i + a ;$$

2. dagli  $\{\bar{\eta}_i\}$  costruiamo una sequenza  $\{y_i\}$  di numeri distribuiti uniformemente in  $(0, f_{max})$ ,

$$y_i = f_{max}\bar{\eta}_i ;$$

3. accettiamo gli  $x_i$  soltanto se risulta,

$$y_i < f(x_i) .$$

La sequenza ottenuta  $\{x'_i\}$  risulta distribuita secondo una densità di probabilità pari al prodotto della probabilità di estrazione, che è

$$g(x) dx = \frac{1}{b-a} dx$$

per la probabilità di accettazione

$$\frac{f(x)}{f_{max}},$$

cioè

$$p(x) dx = \frac{1}{b-a} \frac{1}{f_{max}} f(x) dx$$

$p(x)$  differisce da  $f(x)$  soltanto per una costante di normalizzazione.

Questo metodo consente anche di valutare l'integrale (2.2). Infatti

$$I = (b-a)f_{max} \int_a^b p(x) dx \simeq (b-a)f_{max} \frac{N_{acc}}{N},$$

dove  $N_{acc}$  è la quantità di numeri accettati degli  $N$   $\{x_i\}$  iniziali.

**Esempio:** Applichiamo il metodo del Rigetto alla densità di probabilità

$$f(x) = \frac{4}{\pi} \sqrt{1-x^2} \text{ con } 0 \leq x \leq 1.$$

Con un semplice programma al calcolatore che segue le linee di quanto esposto sopra possiamo generare i grafici che vediamo in figura 2.3. A sinistra vediamo l'istogramma normalizzato a 1 dei valori di  $x$  accettati degli  $N$  generati sovrapposto alla curva che rappresenta  $f(x)$ . A destra i 10000 punti di coordinate  $(x_i, y_i)$  sono riportati in colore diverso a seconda se  $x_i$  sia tra i numeri accettati o tra i rigettati.

Il metodo può essere facilmente generalizzato a distribuzioni multivariate. In questo caso si seleziona un punto nel dominio a  $n$  dimensioni (sul quale la distribuzione è definita) generando una  $n$ -upla di numeri casuali  $(\xi_1, \xi_2, \dots, \xi_n)$ . Essa viene accettata se, estratto un altro numero casuale  $\bar{\xi}$  tra 0 e 1, risulta

$$f_{max}\bar{\xi} < f(\xi_1, \xi_2, \dots, \xi_n).$$

Ripetuto  $N$  volte tale procedimento l'integrale potrà essere approssimato come

$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \dots \int_{a_n}^{b_n} dx_n f(x_1, x_2, \dots, x_n) \simeq (b_1-a_1)(b_2-a_2) \dots (b_n-a_n) f_{max} \frac{N_{acc}}{N}$$

con ovvio significato dei simboli.

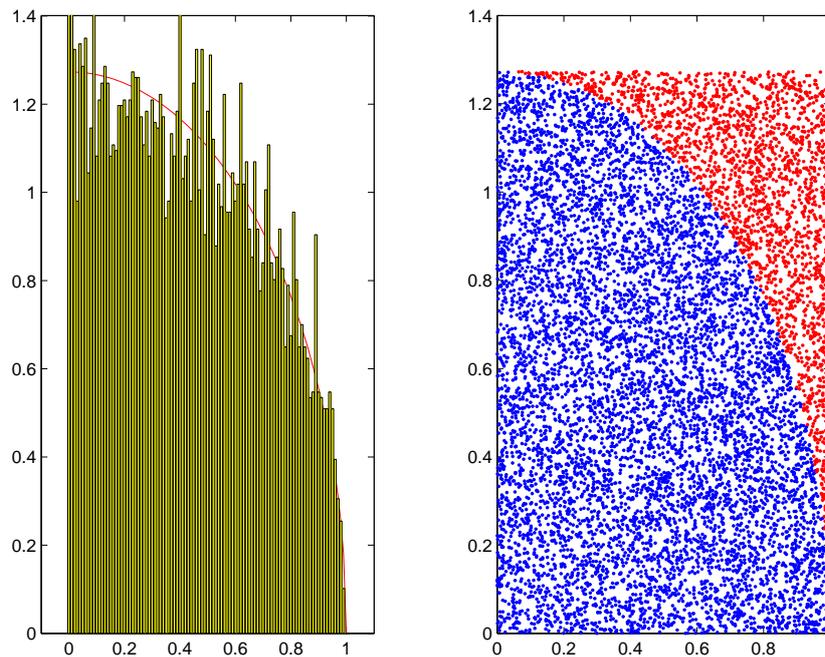


Figura 2.3: Il metodo del Rigelto usato per generare un set di numeri distribuiti secondo la densità di probabilità  $f(x) = \frac{4}{\pi} \sqrt{1-x^2}$ .

ESERCIZIO: Generare delle coppie di valori per le variabili aleatorie  $x$  e  $y$  definite entrambe nell'intervallo  $[0, 1]$  e distribuite secondo la funzione

$$f(x, y) = Ax(1-x)[1 + (2y-1)^2]$$

costruendo i due istogrammi per i valori di  $x$  e di  $y$  accettati e confrontarli con le distribuzioni marginali delle due variabili.

Il metodo del Rigelto può essere usato anche per integrare funzioni non definite positive. In tal caso basta, infatti, applicarlo alla funzione  $f(x) - f_{min}$ , dove  $f_{min}$  è il minimo assunto da  $f(x)$  nel dominio d'integrazione, che è sicuramente non negativa. Ovviamente non è, invece, possibile applicarlo in presenza di singolarità, pur se integrabili, della funzione integranda, per la mancanza del massimo.

In generale l'efficienza del metodo è legata alla percentuale di rigetti. Esso perde di efficienza, quindi, per funzioni dotate di picchi molto pronunciati o di grandi oscillazioni. Un modo per utilizzare il metodo del Rigelto anche in questi casi è quello di modificare leggermente questa tecnica. È evidente che il numero di rigetti dipende dall'area tra le due curve descritte dalle funzioni  $y = f_{max}$  e  $y = f(x)$ . L'efficienza può essere aumentata sostituendo la prima delle due

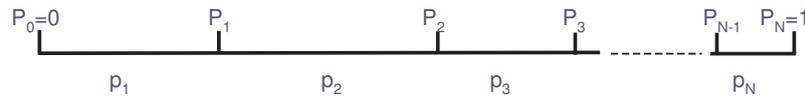


Figura 2.4: Il metodo d'Inversione nel caso di variabili aleatorie discrete. Le probabilità  $p_i$  e le probabilità cumulative  $P_i$ .

funzioni con  $y = g(x)$  tale che

1. sia  $g(x) > f(x) \forall x \in [a, b]$
2. esista un metodo di campionamento efficiente per  $g(x)$

In questo caso la sequenza  $\{x_i\}$  sarà costruita campionando in  $(a, b)$  non uniformemente, ma secondo la distribuzione  $g(x)$ . Ciascun valore  $x_i$  sarà accettato se, estratto un nuovo numero  $y_i$  dalla distribuzione uniforme in  $(0, g(x_i))$

$$y_i = g(x_i)\bar{\eta}_i,$$

risulterà

$$y_i < f(x_i).$$

A questo punto la densità di probabilità dei punti estratti sarà il prodotto della distribuzione di generazione  $g(x)$  per la probabilità di accettazione

$$\frac{f(x)}{g(x)},$$

quindi proprio la densità richiesta.

### 2.3.2 Metodo di Inversione: variabili aleatorie discrete

Consideriamo una variabile aleatoria  $X$  che può assumere i valori discreti  $\{x_i\}$ ,  $i = 1, 2, \dots, N$  con probabilità  $\{p_i\}$ . Supponiamo di riportare in successione su un asse, a partire dall'origine,  $N$  segmenti di lunghezza pari alle probabilità  $p_i$ .

Se estraiamo un numero casuale con distribuzione uniforme in  $(0, 1)$ , esso cadrà nel segmento  $i$ -simo con probabilità pari alla sua lunghezza, cioè proprio  $p_i$ . Al fine di costruire un algoritmo che realizza quanto esposto, consideriamo le probabilità cumulative definite da

$$P_0 = 0, P_i = \sum_{k=1}^i p_k \text{ con } i = 1, N. \quad (2.3)$$

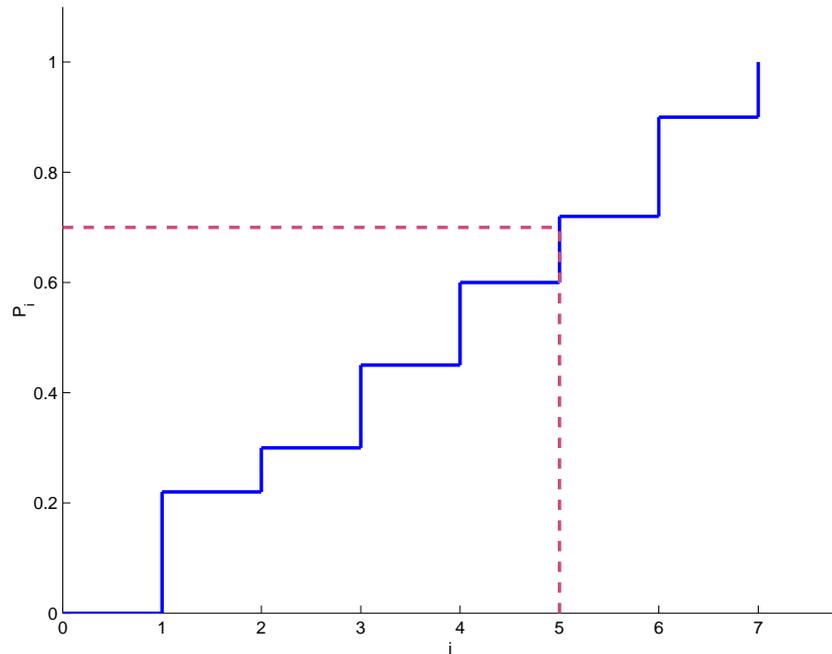


Figura 2.5: Il metodo d'Inversione nel caso di variabili aleatorie discrete.  $P_i$  è la probabilità cumulativa relativa al valore  $x_i$ .

Avremo ovviamente  $P_N = 1$ . Generiamo ora un numero casuale  $\eta$  dalla distribuzione uniforme in  $(0, 1)$  e determiniamo l'intero  $i$  tale che risulti

$$P_{i-1} \leq \eta \leq P_i \quad (2.4)$$

$x_i$  sarà il valore campionato per la variabile aleatoria  $X$ .

Da un punto di vista grafico (vedi fig. 2.5), possiamo riportare le probabilità cumulative in funzione del rispettivo indice  $i$ . Estratto  $\eta$ , tracciamo su questo grafico la retta  $y = \eta$  parallela all'asse delle  $i$ . Essa incontrerà la curva a scaletta in un punto la cui ascissa darà il valore di  $i$  e quindi il valore di  $x_i$  da selezionare.

Notiamo che *la probabilità cumulativa viene selezionata tramite una variabile uniformemente distribuita  $\eta$ ; essa si comporta, cioè, come una variabile aleatoria con distribuzione uniforme*. Si intuisce facilmente come si possa estendere il metodo al campionamento di variabili aleatorie continue.

**ESERCIZIO:** Campionare mediante il metodo d'inversione dalle distribuzioni Binomiale e di Poisson. Confrontare gli istogrammi ottenuti con le distribuzioni esatte.

### 2.3.3 Metodo di Inversione: variabili aleatorie continue

Consideriamo la variabile aleatoria continua  $X$  avente densità di probabilità  $\rho_X(x)$ , che supporremo definita, senza venir meno alla generalità, sull'intero asse reale. Ricordiamo che, se  $Z = f(X)$  è una variabile aleatoria funzione monotona di  $X$ , la sua distribuzione è data da (1.29)

$$\rho_Z(z) = \left| \frac{dz}{dx} \right|^{-1} \rho_X(x(z)).$$

Consideriamo ora il caso in cui  $Z$  assume proprio i valori della distribuzione cumulativa, cioè

$$z(x) = \int_{-\infty}^x \rho_X(t) dt \quad (2.5)$$

che è chiaramente funzione monotona crescente di  $x$  e che assume valori in  $[0, 1]$ . Avremo

$$\rho_Z(z) = \rho_X(x(z)) \frac{1}{\rho_X(x)} = 1 \quad \forall z \in [0, 1].$$

Pertanto la distribuzione cumulativa di probabilità è sempre una variabile aleatoria uniformemente distribuita in  $[0, 1]$ , indipendentemente dalla distribuzione di probabilità di partenza. Nella pratica  $z(x)$  è la primitiva di  $\rho_X(x)$  e stiamo facendo un cambiamento di variabile tale che

$$\rho_X(x) dx = dz.$$

La costruzione di un algoritmo per il campionamento è immediata:

1. Campionare una sequenza di numeri  $\{\eta_i\}$  uniformemente distribuiti in  $[0, 1]$
2. Risolvere (analiticamente o numericamente) per ciascun valore  $\eta_i$  l'equazione

$$z(x_i) = \eta_i$$

determinare cioè

$$x_i = z^{-1}(\eta_i)$$

**Esempio 1:** Applichiamo il metodo dell'Inversione alla distribuzione esponenziale. Consideriamo la densità di probabilità per la variabile aleatoria  $X$

$$\rho_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{per } x \geq 0, \\ 0 & \text{per } x < 0. \end{cases}$$

La distribuzione cumulativa di probabilità è data da

$$z(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

Generato  $\eta$  da una distribuzione uniforme in  $[0, 1]$  dobbiamo risolvere nella variabile  $x$  l'equazione<sup>1</sup>:

$$1 - e^{-\lambda x} = \eta \quad \text{cioè} \quad x = -\frac{1}{\lambda} \ln(1 - \eta).$$

In questo caso il metodo dell'inversione si presenta particolarmente efficiente, perchè l'equazione è facilmente risolvibile. Nei casi in cui non esiste una soluzione analitica, l'equazione può essere risolta numericamente. Non è detto, tuttavia, a questo punto che il metodo d'Inversione sia competitivo con il metodo del Rigetto.

**Esempio 2:** Appliciamo il metodo dell'Inversione alla distribuzione di Cauchy. Consideriamo la variabile aleatoria  $X$  distribuita secondo la densità di probabilità

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

Consideriamo la distribuzione cumulativa di probabilità data da

$$z(x) = \int_{-\infty}^x \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + t^2} dt = \frac{1}{\pi} \int_{-\infty}^{x/\Gamma} \frac{1}{1 + \alpha^2} d\alpha = \frac{1}{\pi} \arctan \frac{x}{\Gamma} + \frac{1}{2}.$$

Data quindi la sequenza  $\{\eta_i\}$  di numeri casuali distribuiti uniformemente in  $[0, 1]$ , la sequenza  $\{x_i\}$  con

$$x_i = \Gamma \tan\left[\pi\left(\eta_i - \frac{1}{2}\right)\right]$$

è distribuita secondo la distribuzione di Cauchy.

ESERCIZIO: Campionare mediante il metodo d'inversione dalle distribuzioni Esponenziale e di Cauchy. Confrontare gli istogrammi ottenuti con le distribuzioni esatte.

### 2.3.4 Campionamento gaussiano

Per generare numeri casuali distribuiti in maniera gaussiana esistono vari algoritmi. Ne esaminiamo due di essi, che si basano su delle proprietà che abbiamo già studiato.

<sup>1</sup>Sui calcolatori esistono due tipi di generatori di numeri pseudocasuali: quelli che generano nell'intervallo  $0 < \eta \leq 1$  e quelli che generano nell'intervallo  $0 \leq \eta < 1$ . Nel primo caso, se fosse estratto il numero  $\bar{\eta} = 1$ , l'equazione da risolvere avrebbe una divergenza. Il problema può essere facilmente risolto tenendo conto del fatto che anche la variabile aleatoria  $1 - \eta$  è uniformemente distribuita in  $[0, 1]$ . Possiamo quindi usare in alternativa l'equazione

$$x = -\frac{1}{\lambda} \ln \eta$$

**Algoritmo basato sul Teorema del Limite Centrale**

Consideriamo  $N$  numeri casuali  $\eta_i$  generati uniformemente in  $[0, 1]$ . Sia

$$Y_N = \sum_{i=1}^N \eta_i$$

Per il Teorema del Limite Centrale la distribuzione di  $Y_N$  tende ad una gaussiana nel limite  $N \rightarrow \infty$ . Questa gaussiana ha valor medio  $N/2$  e varianza  $N/12$ . Infatti (vedi par. 1.4), la somma di variabili aleatorie ha come media la somma delle medie e come varianza, se le variabili sono incorrelate, la somma delle varianze.

Se vogliamo ottenere una distribuzione normale con media 0 e varianza 1, possiamo usare la nuova variabile

$$Z_N = \frac{Y_N - N/2}{\sqrt{N/12}}.$$

Ovviamente la distribuzione può essere considerata gaussiana solo nel limite  $N \rightarrow \infty$ . Tuttavia la distribuzione di  $Z_N$  converge abbastanza rapidamente a quella gaussiana, per cui si ottiene una distribuzione abbastanza vicina a quella gaussiana già per  $N = 12$  di modo che risulta

$$Z_{12} = Y_{12} - 6.$$

**Algoritmo di Box-Muller** Come si è già visto nell'esempio 2 del par. 1.6.2, date due variabili aleatorie  $X_1$  e  $X_2$  distribuite uniformemente in  $[0, 1]$ , le variabili aleatorie  $Z_1$  e  $Z_2$  ottenute dalla trasformazione

$$(Z_1, Z_2) = \sqrt{-2 \ln X_1} (\cos 2\pi X_2, \sin 2\pi X_2)$$

sono distribuite normalmente con media 0 e varianza 1.

**ESERCIZIO:** Generare al computer una sequenza di valori per la variabile

$$Z_N = \frac{Y_N - N/2}{\sqrt{N/12}}$$

sopra definita e mostrare come la sua distribuzione tenda rapidamente a quella Gaussiana al crescere di  $N$ .

**2.3.5 Metodo del Filtraggio**

Questo metodo consiste in una tecnica di generazione mista (come quella che abbiamo visto alla fine del par. 2.3.1, che utilizza contemporaneamente il metodo d'Inversione e quello del Rigetto.

Supponiamo di voler generare una sequenza di valori per una v. a. distribuita secondo la densità di probabilità  $f(x)$  con  $x \in [a, b]$  per la quale si verificano le seguenti condizioni

1. non conosciamo la primitiva
2. il metodo del rigetto si presenta poco efficiente

Supponiamo di poter scrivere

$$f(x) = g(x) h(x)$$

dove  $h(x)$  è una funzione monotona di  $x$ , normalizzata e di cui si conosce la primitiva  $H(x)$ . Potremo quindi scrivere

$$f(x) dx = g(x) h(x) dx = g(x) dH(x) \quad \text{con} \quad H(x) = \int_a^x h(t) dt \in [0, 1].$$

$H$  è uniformemente distribuita in  $[0, 1]$ . Ne deriva di conseguenza che la soluzione al nostro problema consiste nel ripetere il numero desiderato di volte i seguenti passi:

1. Estrarre un numero casuale  $\eta_i$  uniformemente distribuito in  $[0, 1]$
2. Trovare il valore di  $x_i$  soluzione dell'equazione

$$x_i = H^{-1}(\eta_i)$$

3. Generare un secondo numero casuale  $\bar{\eta}_i$  uniformemente distribuito in  $[0, 1]$  e accettare la generazione di  $x_i$  solo se risulta

$$g_{max} \bar{\eta}_i \leq g(x_i)$$

dove  $g_{max}$  è il massimo valore assunto da  $g(x)$  in  $[a, b]$ .

In pratica i numeri  $\{x_i\}$  vengono generati secondo la distribuzione  $h(x)$  con il metodo d'Inversione e poi vengono filtrati mediante il metodo del Rigetto tramite la funzione  $g(x)$ .

## 2.4 Analisi statistica di dati Monte Carlo

### 2.4.1 Stimatore ed Errore Statistico

Consideriamo il problema di calcolare tramite Monte Carlo un integrale. Tale integrale può essere considerato come valore di aspettazione o media  $\mu$  di una funzione <sup>2</sup>  $h(X)$  di una variabile aleatoria  $X$ , distribuita secondo una densità di probabilità  $f(x)$ .

---

<sup>2</sup> $h(x)$  viene detta *score function*, funzione punteggio.

Per definizione abbiamo

$$\mu(h) = \int_{-\infty}^{+\infty} h(x) f(x) dx. \quad (2.6)$$

Come abbiamo visto  $\mu$  può essere stimato con una simulazione Monte Carlo in cui la variabile  $X$  viene campionata dalla densità  $f(x)$  ottenendo una sequenza

$$\{x_i : i = 1, 2, \dots, N\}.$$

Lo *stimatore*

$$\bar{h}_N = \frac{1}{N} \sum_{i=1}^N h(x_i)$$

nel limite

$$N \rightarrow \infty$$

tende al valore  $\mu$  cercato. Per il Teorema del Limite Centrale, sempre nel limite  $N \rightarrow \infty$ , la distribuzione di probabilità per  $\bar{h}_N$  tende ad una gaussiana di media  $\mu$  e varianza  $\sigma^2/N$ , dove

$$\sigma^2 = \int_{-\infty}^{+\infty} [h(x) - \mu]^2 f(x) dx$$

è la varianza di  $h(x)$ .

Questo ci permette di affermare che il Monte Carlo dà una stima di  $\mu$  pari a

$$\bar{h}_N \pm \frac{\sigma}{\sqrt{N}}$$

dove l'intervallo di variabilità corrisponde all'*intervallo di confidenza di una sig-ma*. Questa affermazione significa, più precisamente, che, se  $N$  è sufficientemente grande,  $\bar{h}_N$  cadrà nell'intervallo  $[\mu - \sigma/\sqrt{N}, \mu + \sigma/\sqrt{N}]$  con una probabilità

$$P = \frac{\sqrt{N}}{\sigma\sqrt{2\pi}} \int_{\mu-\sigma/\sqrt{N}}^{\mu+\sigma/\sqrt{N}} e^{-\frac{N(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} e^{-\frac{x^2}{2}} = 0.68268.$$

In generale  $\sigma$  non è nota, ma può essere valutata direttamente dal Monte Carlo stesso tramite lo stimatore

$$S_N^2 = \frac{1}{N} \sum_{i=1}^N h^2(x_i) - \left[ \frac{1}{N} \sum_{i=1}^N h(x_i) \right]^2.$$

La quantità  $S_N/\sqrt{N}$  viene detta *errore statistico*.

Questo modo di valutare i valori di aspettazione e gli integrali, viene detto Monte Carlo Analogo, termine che indica l'uso della densità di probabilità della

variabile  $X$ . Era pertanto un Monte Carlo Analogo il primo metodo visto per valutare gli integrali generando  $X$  dalla distribuzione uniforme. Questo metodo presenta il difetto che l'errore statistico decresce con  $N$  molto lentamente, come  $1/\sqrt{N}$ . Se si vuole, ad esempio dimezzare l'errore statistico, occorre quadruplicare la dimensione del campione.

Il modo per risolvere questo problema consiste nel ricorrere a tecniche che *riducono la varianza senza alterare la media*. Queste tecniche di riduzione della varianza si basano su un metodo di campionamento che viene detto Campionamento d'Importanza (Importance Sampling) e che consiste nell'usare una più opportuna distribuzione per generare la variabile aleatoria. Abbiamo già visto all'opera tecniche simili nel metodo del Rigetto e dell'Inversione, quando erano usate per la generazione di numeri casuali secondo una distribuzione data. Questi metodi consentono anche di calcolare l'integrale della distribuzione, che, tuttavia, è già noto valere 1. Di fatto in questo caso  $h(x) = 1$  e  $S_N^2 = 0$  per ogni  $N$ ; si ottiene quindi varianza nulla, cioè il massimo di efficienza, ma, come si diceva, dal punto di vista del calcolo degli integrali questo non serve. Vedremo nel prossimo paragrafo come è possibile ottenere un miglioramento dell'efficienza anche nel caso in cui  $h(x) \neq 1$ .

### 2.4.2 Tecniche di riduzione della varianza e campionamento d'Importanza

Supporremo di voler sempre calcolare  $\mu(h)$  (2.6) tramite Monte Carlo campionando però la variabile aleatoria  $X$  secondo una nuova densità  $g(x)$ . Per non modificare la media definisco una nuova funzione punteggio

$$H(x) = \frac{h(x) f(x)}{g(x)}.$$

Risulta, infatti,

$$\mu(H) = \int_{-\infty}^{+\infty} \frac{h(x) f(x)}{g(x)} g(x) dx = \int_{-\infty}^{+\infty} h(x) f(x) dx = \mu(h).$$

Per valutare  $\mu(H)$ , una volta campionata una sequenza  $\{x_i : i = 1, 2, \dots, N\}$  dalla densità  $g(x)$ , useremo il nuovo stimatore

$$\bar{H}_N = \frac{1}{N} \sum_{i=1}^N \frac{h(x_i) f(x_i)}{g(x_i)},$$

che, nel limite  $N \rightarrow \infty$ , tende a  $\mu$ . Poichè le due medie  $\mu(H)$  e  $\mu(h)$  coincidono, per confrontare  $\sigma^2(H)$  e  $\sigma^2(h)$  basta che calcoliamo il secondo momento delle

distribuzioni di entrambi. Abbiamo

$$\begin{aligned} M_2(H) &= \int_{-\infty}^{+\infty} H^2(x) g(x) dx = \\ &= \int_{-\infty}^{+\infty} \left[ \frac{h(x) f(x)}{g(x)} \right]^2 g(x) dx = \\ &= \int_{-\infty}^{+\infty} h^2(x) \frac{f(x)}{g(x)} f(x) dx \end{aligned}$$

mentre

$$M_2(h) = \int_{-\infty}^{+\infty} h(x)^2 f(x) dx.$$

Se il rapporto  $f(x)/g(x)$  è una quantità mediamente minore di 1 per i valori di  $x$  per i quali  $h^2(x) f(x)$  è grande, la varianza di  $H(x)$  può essere ridotta in maniera significativa. Benchè si tratti di requisiti sottintesi, è bene sottolineare che:

- $g(x)$  deve essere interpretabile come densità di probabilità, deve quindi essere non negativa e normalizzata;
- deve esistere un metodo efficiente di campionamento secondo  $g(x)$ , o in quanto sia disponibile un generatore, o in quanto sia semplice usare uno dei metodi già visti. Ad esempio, per poter usare il metodo d'Inversione deve potersi calcolare  $G(x) = \int_{-\infty}^x g(x) dx$  e risolvere l'equazione  $x = G^{-1}(\eta)$ .

Dal punto di vista matematico il campionamento d'importanza corrisponde ad un cambiamento della variabile d'integrazione

$$h(x) f(x) dx = h(x) dF(x) \rightarrow h(x) \frac{f(x)}{g(x)} dG(x).$$

La variabile  $G$ , distribuita uniformemente tra 0 e 1, viene usata al posto della variabile

$$F(x) = \int_{-\infty}^x f(x) dx.$$

Questo metodo presenta delle similitudini con il metodo del Filtraggio (par. 2.3.5). In quel caso il cambiamento di variabile serviva ad aumentare l'efficienza nel metodo del rigetto, ora è invece motivato dalla finalità di ridurre l'errore statistico<sup>3</sup>. In ogni caso queste tecniche possono essere utili per rimuovere delle singolarità dalla funzione integranda.

---

<sup>3</sup>È ovvio che l'efficienza del metodo del Rigetto aumenta proprio quando riduciamo la varianza della funzione su cui pratichiamo il rigetto. Quindi anche la tecnica del Filtraggio può essere considerata una tecnica di riduzione della varianza.

### 2.4.3 Assorbimento di particelle da parte di una lastra

Per illustrare la tecnica appena vista consideriamo un semplice esempio. Supponiamo che delle particelle, p. e. neutroni o gamma, incidano normalmente su una piastra di materiale di spessore  $T$ . Supponiamo inoltre che ciascuna particella possa penetrare per una distanza  $x$  della piastra con una densità di probabilità

$$f(x) dx = \alpha e^{-\alpha x} dx.$$

$\alpha^{-1}$  rappresenta il cammino libero medio all'interno del materiale della piastra. La penetrazione può arrestarsi sia a causa di assorbimento, che a causa di una collisione che fa variare la direzione del moto. Adottando un modello semplificato, supporremo che sia presente solo il fenomeno dell'assorbimento.

Per semplificare le formule assumiamo, come unità di misura delle distanze, proprio il cammino libero medio:

$$f(x) dx = e^{-x} dx.$$

Vogliamo calcolare la frazione di particelle che riescono a superare lo spessore della piastra.

Con il Monte Carlo analogo le nostre azioni sarebbero state:

- Campionare  $x$  dalla densità esponenziale;
- segnare il punteggio 1 (0) se  $x > T$  ( $x < T$ );
- dopo  $N$  campionamenti calcolare il punteggio medio e l'errore statistico.

Questo corrisponde a calcolare un'approssimazione al valor medio della funzione punteggio

$$h(x) = \Theta(x - T)$$

sulla densità di probabilità  $f(x)$ . Per il problema in oggetto tutte le quantità possono essere calcolate analiticamente:

$$\mu(h) = \int_0^{\infty} h(x) e^{-x} dx = \int_T^{\infty} e^{-x} dx = e^{-T}$$

$$\sigma^2(h) = \int_0^{\infty} (h(x) - \mu(h))^2 e^{-x} dx = \int_0^{\infty} (h^2(x) - 2\mu h(x) + \mu^2) e^{-x} dx = \mu(1-\mu)$$

Se vogliamo, tanto per fissare un livello di precisione, che la deviazione standard per il valore medio,  $\sigma/\sqrt{N}$ , diventi pari al 10% del valore medio stesso occorre che

$$\frac{\sigma}{\sqrt{N}} = \frac{10}{100} \times \mu,$$

$T$	$\mu$	$\sigma/\mu$	$N$
3	$4.98 \cdot 10^{-2}$	4.37	$1.91 \cdot 10^3$
5	$6.74 \cdot 10^{-3}$	1.21	$1.47 \cdot 10^4$
10	$4.54 \cdot 10^{-5}$	$1.48 \cdot 10^2$	$2.20 \cdot 10^6$
20	$2.06 \cdot 10^{-9}$	$2.20 \cdot 10^4$	$4.85 \cdot 10^{10}$

Tabella 2.1: Calcolo del numero di campionamenti necessario per una precisione del 10% con il Monte Carlo analogo.

occorrerà, cioè, campionare dalla distribuzione per un numero di volte

$$N = \left( \frac{10\sigma}{\mu} \right)^2 = 100 \frac{1 - \mu}{\mu} = 100 \frac{1 - e^{-T}}{e^{-T}} \underset{T \gg 1}{\approx} 100e^T$$

Fissando alcuni valori indicativi di  $T$  possiamo costruire così la tabella 2.1 dalla quale si può dedurre che, per spessori superiori a una decina di volte il cammino libero medio, non sarà possibile ottenere la precisione richiesta su qualsiasi computer. Il Campionamento d'Importanza consiste, in questo caso, nell'uso di una densità distorta dipendente da un parametro  $b$

$$g(x, \bar{b}) = \bar{b} e^{-\bar{b}x}$$

dove  $\bar{b}$  è il particolare valore del parametro che minimizza la varianza.

La funzione punteggio diventa ora

$$H(x) = \frac{f(x)}{g(x, \bar{b})} h(x) = \frac{f(x)}{g(x, \bar{b})} \Theta(x - T),$$

per cui, estratta la sequenza  $\{x_i : i = 1, 2, \dots, N\}$  dalla densità  $g(x, \bar{b})$ , segneremo un punteggio 0 se  $x_i < T$  o punteggio

$$H(x_i) = \frac{f(x_i)}{g(x_i, \bar{b})} = \frac{1}{\bar{b}} e^{-x_i(1-\bar{b})}$$

se  $x_i > T$ . La situazione si presenta chiaramente diversa, come si può notare dalla figura 2.6 nella quale funzione peso e funzione punteggio sono mostrate nei due casi per  $T = 5$  e  $\bar{b} = 0.18$ .

Con il campionamento d'Importanza avremo una varianza

$$\begin{aligned} \sigma^2(H) &= \int_0^\infty \left[ \frac{1}{\bar{b}} e^{-x(1-\bar{b})} \Theta(x - T) \right]^2 b e^{-bx} dx - \mu^2(H) = \\ &= \int_T^\infty \frac{1}{\bar{b}} e^{-x(2-\bar{b})} dx - e^{-2T} = \frac{e^{-T(2-\bar{b})}}{b(2-\bar{b})} - e^{-2T} \end{aligned}$$

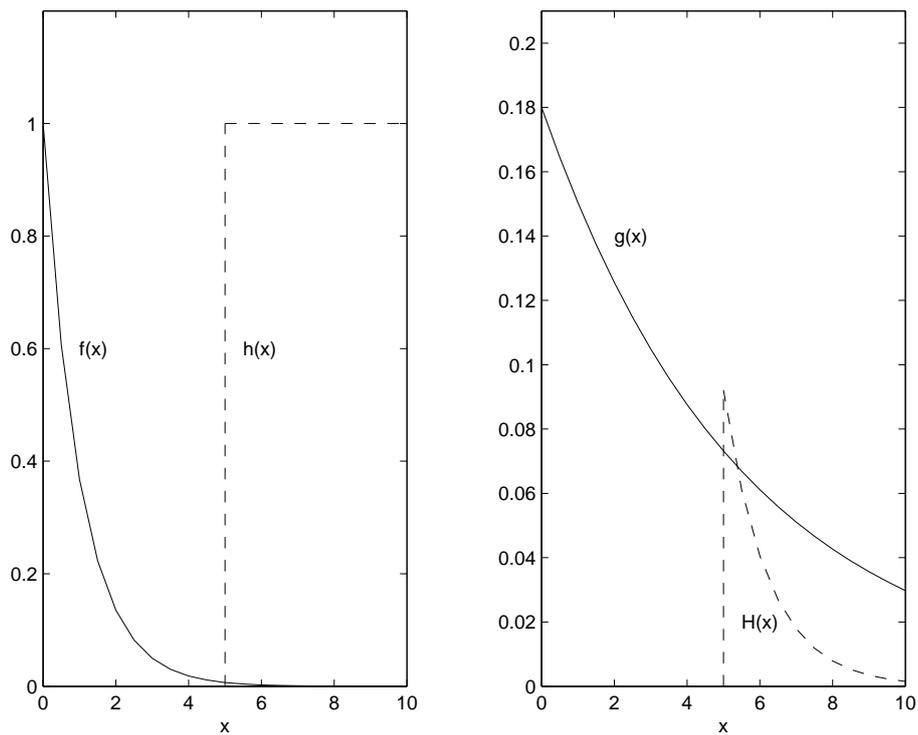


Figura 2.6: Confronto tra Monte Carlo analogo e Monte Carlo d'importanza: nel primo caso la densità di probabilità ha valori trascurabili dove la funzione punteggio è diversa da 0.

$T$	$\bar{b}$	$\sigma/\mu$	$N$
3	0.28	1.95	381
5	0.18	2.55	651
10	0.09	3.65	1329
20	0.048	5.18	2687

Tabella 2.2: Calcolo del numero di campionamenti necessario per una precisione del 10% con  $g(x, \bar{b})$ .

che assume il suo valore minimo per

$$b = \bar{b} = 1 + \frac{1}{T} - \sqrt{1 + \frac{1}{T^2}}.$$

Come nel caso precedente, riportiamo in una tabella il numero di estrazioni richieste per avere una stima della media entro l'errore del 10%.

Da questa tabella vediamo come si sia ottenuta una riduzione drastica della varianza e quindi del numero di campionamenti necessari per avere un valore affidabile per l'integrale. Nel caso  $T = 20$  il valore  $N = 2687$  va confrontato con il valore  $N = 4.85$  miliardi trovato per il Monte Carlo analogo.

ESERCIZIO: Riprodurre al computer la tabella calcolata analiticamente usando i valori di  $\bar{b}$  ivi riportati 2.2.

#### 2.4.4 Tecnica di Spanier

Si comprende facilmente che, per applicare il Campionamento d'Importanza, è necessario, oltre all'individuazione di una funzione tipo  $g(x, b)$  in grado di ridurre la varianza, anche la determinazione dei valori ottimali del parametro (o dei parametri). In generale questo non può essere fatto analiticamente, per questo Spanier ha proposto una tecnica di ottimizzazione mediante un numero di campionamenti Monte Carlo relativamente piccolo. Esemplichiamo questa tecnica nel nostro caso.

Si considera un certo numero  $M$  di valori del parametro  $\{b_i : i = 1, 2, \dots, M\}$  che coprono l'intero intervallo dei valori di variabilità di  $b$  e si studia l'espressione

$$\widetilde{M}_2(b_j) = \int \frac{h(x) f(x)}{g(x, b_j)} \frac{h(x) f(x)}{g(x, \bar{b})} g(x, \bar{b}) dx$$

che differisce da  $M_2$  per il fatto che sono presenti due valori del parametro  $b$ :  $\bar{b}$  è il valore usato per la generazione, mentre  $b_j$  è uno degli  $M$  valori fissati precedentemente.

La tecnica di Spanier procede nel modo seguente:

- Si parte con un primo valore  $\bar{b}$  e con questo valore si genera un set di  $N$  campionamenti dalla distribuzione  $g(x, \bar{b})$ .
- Si calcola l'approssimazione a  $\widetilde{M}_2(b_j)$  tramite lo stimatore

$$\widetilde{M}_2(b_j) = \frac{1}{N} \sum_{i=1}^N \frac{h(x_i) f(x_i)}{g(x_i, b_j)} \frac{h(x_i) f(x_i)}{g(x_i, \bar{b})} \quad \text{per } j = 1, \dots, M.$$

- Si trova il valore  $b_j$  corrispondente al più piccolo degli  $M_2(b_j)$ .
- Questo valore di  $b_j$  viene usato come  $\bar{b}$  per la successiva generazione.
- Si ripete la procedura finchè essa non converge, cioè finchè il valore corrispondente al minimo è proprio  $\bar{b}$ .

Ai fini del calcolo di  $\bar{b}$  non è necessario essere molto precisi, quindi il valore di  $N$  può anche essere non molto grande.

# Capitolo 3

## Campionamento in Meccanica Statistica

In questo capitolo viene esposta la tecnica di campionamento utilizzata nella Meccanica Statistica dei sistemi di spin su reticolo [3, 4]. Tecniche Monte Carlo vengono usate in Meccanica Statistica per simulare dei modelli teorici in condizioni di equilibrio e calcolarne le predizioni. Come sempre in Meccanica Statistica queste sono rappresentate da medie di grandezze dinamiche sulla distribuzione di probabilità relativa alle condizioni di equilibrio.

Si tratta, in questo caso, di un campionamento d'Importanza fatto su un sistema la cui evoluzione è simulata su calcolatore. Gli algoritmi che sono stati elaborati tengono conto di due problemi essenziali: come essere sicuri che il sistema venga portato in condizioni di equilibrio e come essere sicuri che vi permanga durante il processo di misura, in modo che le medie calcolate siano rappresentative di tali condizioni.

L'analogia esistente tra modelli in Meccanica Statistica e modelli in Teoria dei Campi consente di utilizzare le medesime tecniche anche in questo settore della Fisica Teorica.

### 3.1 Il concetto di Ensemble

Come abbiamo visto, lo studio formale della Teoria della probabilità richiede tre concetti distinti: lo spazio dei campioni  $\Omega$ , lo spazio degli eventi  $\mathbf{B}$  (Campo di Borel) e la misura di probabilità  $P$ .

I Fisici statistici utilizzano un'unica nozione, quella di *Ensemble*[5]. Un Ensemble è una collezione i cui membri sono gli elementi dello spazio dei campioni ciascuno dei quali ripetuto un numero di volte proporzionale alle sua probabilità, in modo, cioè, che la sua probabilità è proprio data dal rapporto tra tale numero e il numero totale degli elementi dell'ensemble. I campioni sono in questo caso le configurazioni microscopiche nelle quali il sistema che si sta studiando si può

trovare. Per sistemi caratterizzati da variabili dinamiche discrete e di taglia finita, gli elementi dell'Ensemble sono in numero finito. In generale, tuttavia, si ha a che fare con sistemi le cui variabili sono continue e che spesso sono studiati nel *limite Termodinamico*, cioè per un numero infinito di componenti, per cui l'Ensemble risulta infinito. Nel caso continuo è necessario introdurre il concetto di densità degli stati. Ogni stato microscopico può essere considerato un punto nell'iperspazio  $\Gamma$  delle coordinate descrittive del sistema (ad es. lo spazio delle fasi a  $6N$  dimensioni per un gas di  $N$  molecole). L'Ensemble viene così rappresentato in tale spazio da una distribuzione di punti caratterizzati da una densità. È questa densità che sostituisce il concetto di densità di probabilità.

## 3.2 Il Campionamento in Meccanica Statistica

Le simulazioni Monte Carlo in Meccanica Statistica vengono in generale utilizzate nell'ambito dell'*Ensemble Canonico*. Tale Ensemble corrisponde a sistemi composti da un numero costante di elementi in equilibrio con un bagno termico esterno a temperatura fissata  $T$  con il quale possono scambiare energia. In questo caso la distribuzione di probabilità del sistema è data da

$$p_\mu = \frac{e^{-\beta E_\mu}}{\sum_\mu e^{-\beta E_\mu}} \quad (3.1)$$

che viene detta distribuzione di Boltzmann. In questa espressione  $\mu$  etichetta lo stato microscopico,  $E_\mu$  è la sua energia,  $\beta = 1/KT$  ( $K$  costante di Boltzmann,  $T$  temperatura assoluta del bagno termico). La (3.1) riguarda sistemi che possiedono un set discreto di stati, che sono quelli che qui saranno considerati.

Come si è detto, l'obiettivo della Meccanica Statistica è quello di calcolare quantità tipiche della Termodinamica come valori d'attesa su tutti i possibili microstati di variabili fisiche del tipo di

$$\langle Q \rangle = \frac{\sum_\mu Q_\mu e^{-\beta E_\mu}}{\sum_\mu e^{-\beta E_\mu}} \quad (3.2)$$

Ad esempio  $Q_\mu$  potrebbe essere l'energia stessa e in questo caso  $\langle Q \rangle$  rappresenterebbe l'energia interna, oppure la somma dei momenti magnetici di un sistema di atomi e  $\langle Q \rangle$  rappresenterebbe la magnetizzazione.

Il calcolo di  $\langle Q \rangle$  usando la formula (3.2) è possibile solo per sistemi di taglia molto piccola. Per esemplificare la problematica consideriamo un modello molto semplice, il modello di Ising, che è il modello per un magnete. In questo modello si ha un reticolo, di cui si possono scegliere la dimensionalità, la grandezza e la geometria. A ciascun sito  $i$  del reticolo è associata una variabile dicotomica  $s_i$  che può assumere valori  $\pm 1$  e rappresenta il momento magnetico o spin di un atomo di un reticolo cristallino.

L'Hamiltoniano per questo sistema è

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} s_i s_j - H \sum_i s_i \quad (3.3)$$

dove  $J$  è una costante di accoppiamento tra gli spin,  $H$  un campo magnetico omogeneo e la somma si intende estesa a tutte le coppie di spin primi vicini sul reticolo. Nel caso  $J > 0$  gli spin tendono ad allinearsi (interazione ferromagnetica); analogamente il segno  $-$  davanti ad  $H$  significa che gli spin tendono ad allinearsi al campo magnetico esterno. Se il sistema è immerso in bagno termico a temperatura  $T$  la probabilità di una particolare configurazione  $\mu$  è data da

$$p_\mu = \frac{e^{(K \sum_{\langle i,j \rangle} s_i s_j + B \sum_i s_i)_\mu}}{\sum_\mu e^{(K \sum_{\langle i,j \rangle} s_i s_j + B \sum_i s_i)_\mu}} \quad (3.4)$$

dove  $K = \beta J$  e  $B = \beta H$ . La configurazione  $\mu$  corrisponde ad un set di valori per gli spin del reticolo

$$\mu \equiv \{s_1, s_2, \dots, s_N\}$$

La magnetizzazione definita come

$$\langle M \rangle = \frac{1}{N} \left\langle \sum_{i=1}^N s_i \right\rangle \quad (3.5)$$

è una funzione della temperatura. L'interesse del modello deriva dal fatto che esso presenta una transizione di fase paramagnetica-ferromagnetica per reticoli in dimensione  $D \geq 2$  in assenza di campo magnetico, per cui la magnetizzazione assume un valore non nullo al di sotto di una temperatura critica  $T_c$ . Benchè il modello sia molto semplice se ne conosce la soluzione esatta soltanto in  $D = 1$  e, in assenza di campo, in  $D = 2$ . La prima fu trovata nel 1925 da Ising stesso (cui il problema era stato affidato dal suo maestro Lenz nel 1920). Nel secondo caso fu trovata da Onsager nel 1944.

Tornando al problema del calcolo dei valori d'attesa, supponiamo di voler valutare per il modello di Ising la magnetizzazione (3.5), cioè

$$\langle M \rangle = \frac{1}{N} \frac{\sum_\mu (\sum_{i=1}^N s_i)_\mu e^{(K \sum_{\langle i,j \rangle} s_i s_j + B \sum_i s_i)_\mu}}{\sum_\mu e^{(K \sum_{\langle i,j \rangle} s_i s_j + B \sum_i s_i)_\mu}}.$$

Il numero totale di elementi nella somma sulle configurazioni dipende dalla taglia del reticolo. Nel caso di un reticolo quadrato con  $N$  spin per lato, poichè ogni spin può assumere due possibili valori, avremo per  $N = 3$  un numero di configurazioni pari a  $2^3 = 8$  che diventa  $2^{10} \simeq 1.024 \cdot 10^3$  nel caso di  $N = 10$ . Poichè le transizioni di fase si hanno solo nel limite termodinamico, è chiaro che con questo sistema di calcolo non ne vedremo mai la benchè minima traccia. Di qui discende l'impossibilità di sommare su tutte le configurazioni e quindi la necessità del campionamento.

### 3.3 Lo stimatore

Il modo più semplice di applicare la tecnica Monte Carlo ad un sistema termico è di scegliere un sottoinsieme di  $M$  stati  $\{\mu_1, \mu_2, \dots, \mu_M\}$  con una certa probabilità  $g_\mu$  e calcolare invece della quantità  $\langle Q \rangle$  lo stimatore

$$Q_M = \frac{\sum_i Q_{\mu_i} g_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}{\sum_i g_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}$$

dove ogni configurazione è stata ulteriormente pesata con l'inverso della probabilità di generazione per rimuovere la distorsione effettuata nel campionamento. È chiaro che  $Q_M$  è un buon stimatore:

$$\lim_{M \rightarrow \infty} Q_M = \langle Q \rangle.$$

Nel caso più semplice le  $g_\mu$  sono tutte uguali ( $g_\mu = 1/M$ ) e avremo

$$Q_M = \frac{\sum_i Q_{\mu_i} e^{-\beta E_{\mu_i}}}{\sum_i e^{-\beta E_{\mu_i}}}$$

In molti casi tuttavia è possibile campionare solo un numero limitato di stati e questo pone nuovi problemi. Consideriamo ad esempio il modello di Ising su reticolo cubico  $10 \times 10 \times 10$ . Il numero di stati è  $2^{1000} \simeq 10^{300}$ . Possiamo campionare, seppure con qualche difficoltà, un numero di stati dell'ordine di  $10^8$ . Questo vuol dire che ogni stato ha probabilità pari a  $10^{-292}$  di essere campionato. Ma, come è subito chiaro dalla espressione di  $\langle Q \rangle$ , se si va a bassa temperatura ( $\beta$  grande) solo pochissimi stati di energia più bassa avranno probabilità di Boltzmann macroscopiche, mentre per tutti gli altri la probabilità sarà trascurabile. Questo comporta che  $Q_M$  darà nella pratica una stima molto inaccurata di  $\langle Q \rangle$ .

### 3.4 Il Campionamento d'Importanza

La soluzione, anche in Meccanica Statistica, è il campionamento d'Importanza, che in questo caso consiste nel campionare direttamente secondo la distribuzione di Boltzmann, o, per lo meno, cercare di farlo al meglio. In questo modo si riduce il calcolo dello stimatore alla semplice media

$$Q_M = \frac{\sum_i Q_{\mu_i}}{M}. \quad (3.6)$$

La strategia che viene seguita consiste nel simulare su computer un processo in cui il sistema che si vuole studiare evolve fino a trovarsi in condizioni di equilibrio alla temperatura fissata. Questo vuol dire che, nel corso della successiva evoluzione, i microstati del sistema si presenteranno con una frequenza che riproduce la

probabilità di Boltzmann. Da questo punto in poi si procede al campionamento desiderato.

La simulazione del processo evolutivo avviene utilizzando un particolare metodo Monte Carlo per generare una sequenza di configurazioni del sistema

$$\mu_1 \rightarrow \mu_2 \rightarrow \mu_3 \rightarrow \dots$$

Il processo è tale che ciascuna configurazione dipende soltanto dalla configurazione precedente e non dalla storia progressa, si tratta cioè di un *processo di Markov*.

Chiamiamo

$$P(\mu \rightarrow \mu')$$

la probabilità che il sistema si trovi nella configurazione  $\mu'$  se la precedente configurazione è  $\mu$ .  $P$  è ovviamente normalizzata:

$$\sum_{\mu'} P(\mu \rightarrow \mu') = 1$$

Vengono considerati processi di Markov che soddisfano i seguenti requisiti:

**Accessibilità:** qualsiasi sia la configurazione ad un certo istante deve esservi una probabilità non nulla di passare ad una qualsiasi delle altre configurazioni. Questa condizione garantisce che, nel limite  $M \rightarrow \infty$  lo stimatore converga effettivamente al valor medio cercato.

**Microreversibilità:** detta  $p_\mu$  la distribuzione di probabilità secondo la quale si vuole campionare il generico stato  $\mu$ , deve essere soddisfatta la relazione

$$p_\mu P(\mu \rightarrow \mu') = p_{\mu'} P(\mu' \rightarrow \mu) \quad (3.7)$$

Questa condizione è detta anche del *Bilancio Dettagliato* e risulta essenziale per dimostrare la convergenza della distribuzione di probabilità dei microstati alla desiderata distribuzione  $p_\mu$ .

Per processi di Markov così caratterizzati si dimostra la proposizione seguente:

LA DEVIAZIONE DELLA DISTRIBUZIONE DI PROBABILITÀ DELLE CONFIGURAZIONI DA  $p_\mu$  DIMINUISCE PROCEDENDO LUNGO LA CATENA DI MARKOV. SE AL PASSO  $n$ -SIMO DELLA SEQUENZA LA PROBABILITÀ DELLA CONFIGURAZIONE  $\mu$  È PROPRIO  $p_\mu$ , ESSA RESTERÀ TALE PER TUTTI I PASSI SUCCESSIVI.

Per dimostrare questa proposizione occorre introdurre una misura della distanza tra le distribuzioni e fare vedere che questa distanza diminuisce nel corso della catena di Markov fino ad annullarsi. Denominiamo

$$W(\mu, n)$$

la distribuzione di probabilità vigente al passo  $n$ -simo della catena e definiamo quindi distanza tra le distribuzioni la grandezza

$$D_n = \sum_{\mu} |W(\mu, n) - p_{\mu}| \quad (3.8)$$

Al passo successivo avremo, applicando la proprietà di normalizzazione della  $P$  e la relazione di microreversibilità

$$\begin{aligned} D_{n+1} &= \sum_{\mu} |W(\mu, n+1) - p_{\mu}| = \\ &= \sum_{\mu} \left| \sum_{\mu'} W(\mu', n) P(\mu' \rightarrow \mu) - p_{\mu} \sum_{\mu'} P(\mu \rightarrow \mu') \right| = \\ &= \sum_{\mu} \left| \sum_{\mu'} W(\mu', n) P(\mu' \rightarrow \mu) - \sum_{\mu'} p_{\mu'} P(\mu' \rightarrow \mu) \right| = \\ &= \sum_{\mu} \left| \sum_{\mu'} [W(\mu', n) - p_{\mu'}] P(\mu' \rightarrow \mu) \right| \end{aligned}$$

Applicando la disuguaglianza triangolare e poichè  $P(\mu' \rightarrow \mu) \geq 0 \forall \mu', \mu$  risulta

$$\left| \sum_{\mu'} [W(\mu', n) - p_{\mu'}] P(\mu' \rightarrow \mu) \right| \leq \sum_{\mu'} |W(\mu', n) - p_{\mu'}| P(\mu' \rightarrow \mu),$$

quindi

$$D_{n+1} \leq \sum_{\mu} \sum_{\mu'} |W(\mu', n) - p_{\mu'}| P(\mu' \rightarrow \mu) = \sum_{\mu'} |W(\mu', n) - p_{\mu'}| = D_n$$

Notiamo inoltre che, perchè sia  $D_{n+1} = D_n$ , deve accadere che, per ogni  $\mu'$ , i termini  $W(\mu', n) - p_{\mu'}$  abbiano tutti lo stesso segno o che siano tutti nulli. Il primo dei due casi non può verificarsi perchè si tratta di una differenza tra quantità entrambe normalizzate a 1. Pertanto vediamo che  $D_n = 0 \Rightarrow D_{n+1} = 0$ . La proposizione risulta quindi completamente dimostrata.

### 3.5 L'algoritmo di Metropolis

Quanto si è visto al paragrafo precedente consente di costruire un processo di Markov nel quale progressivamente la distribuzione  $p_{\mu}$  di probabilità delle configurazioni di un sistema tende verso quella che è stata utilizzata nella legge di Microreversibilità utilizzata nel corso del processo stesso. Anzi, visto che non è stata necessaria alcuna ipotesi sulle probabilità di transizione  $P(\mu \rightarrow \mu')$ , la

sceita sul tipo di processo è ampia. In particolare si è interessati, nel caso della Meccanica Statistica su reticolo, a costruire un processo tale che  $p_\mu$  coincida con la distribuzione di probabilità di Boltzmann 3.1.

Il prototipo di questi processi è denominato *Algoritmo di Metropolis*, e, benchè sviluppato nell'ambito della Fisica Nucleare, è oggi molto utilizzato proprio per simulare sistemi statistici.

In maniera del tutto generale, cioè per una qualsiasi distribuzione desiderata  $p_\mu$ , l'algoritmo considera la transizione da una configurazione  $\mu$  ad una configurazione  $\mu'$  del sistema sulla base di due passaggi:

1. si sceglie una configurazione di prova  $\mu'$  con una probabilità simmetrica  $P^*(\mu \rightarrow \mu') (= P^*(\mu' \rightarrow \mu))$ ;
2. se  $p_{\mu'} \geq p_\mu$  la configurazione  $\mu'$  viene accettata.  
se  $p_{\mu'} < p_\mu$ , si genera un numero casuale  $\xi$  uniformemente distribuito in  $(0,1)$ ; dopo di che:
  - se  $\xi < p_{\mu'}/p_\mu$  la configurazione  $\mu'$  viene accettata,
  - se  $\xi > p_{\mu'}/p_\mu$  la configurazione  $\mu'$  viene rigettata

Si vede facilmente che la probabilità di transizione può essere formalizzata matematicamente nel modo seguente

$$P(\mu \rightarrow \mu') = P^*(\mu \rightarrow \mu') \times \min\left(1, \frac{p_{\mu'}}{p_\mu}\right)$$

$$P(\mu \rightarrow \mu) = P^*(\mu \rightarrow \mu) + \sum_{i=1}^N P^*(\mu \rightarrow \mu_i) \times \max\left(0, 1 - \frac{p_{\mu_i}}{p_\mu}\right)$$

Verifichiamo che la legge di Microreversibilità è soddisfatta:

**I caso:**  $p_{\mu'} \geq p_\mu \Leftrightarrow \frac{p_{\mu'}}{p_\mu} > 1 \Leftrightarrow \frac{p_\mu}{p_{\mu'}} < 1$

$$p_\mu P(\mu \rightarrow \mu') = p_\mu P^*(\mu \rightarrow \mu') = \frac{p_\mu}{p_{\mu'}} p_{\mu'} P^*(\mu' \rightarrow \mu) = p_{\mu'} P(\mu' \rightarrow \mu)$$

**II caso:**  $p_{\mu'} \leq p_\mu \Leftrightarrow \frac{p_{\mu'}}{p_\mu} < 1 \Leftrightarrow \frac{p_\mu}{p_{\mu'}} > 1$

$$p_\mu P(\mu \rightarrow \mu') = p_\mu P^*(\mu \rightarrow \mu') \frac{p_{\mu'}}{p_\mu} = p_{\mu'} P^*(\mu \rightarrow \mu') \times 1 = p_{\mu'} P(\mu' \rightarrow \mu)$$

**ESERCIZIO:** Utilizzare l'algoritmo di Metropolis per campionare dalla distribuzione esponenziale  $e^{-x}$ .

Normalmente la selezione di un nuovo valore della variabile  $x$  in tutto l'intervallo  $(0, \infty)$  può rendere molto bassa l'accettazione. Perché? È lecito restringersi ad un intervallo limitato intorno al valore attuale della  $x$ ? Confrontare la distribuzione ottenuta con quella desiderata.

### 3.5.1 Applicazione ai modelli di spin

Come dicevamo, l'algoritmo di Metropolis (e più in generale i tanti algoritmi che da esso sono derivati) ha una specifica applicazione nel caso della simulazione dei modelli di spin su reticolo in presenza di bagno termico a temperatura fissata  $T$ . In questo caso a partire da una configurazione a piacere, che può anche essere poco probabile alla temperatura  $T$ , l'algoritmo consente al sistema di portarsi in uno stato di equilibrio, in uno stato, cioè, in cui le configurazioni tendono a presentarsi, nel corso dell'evoluzione, con frequenza proporzionale alla distribuzione di probabilità di Boltzmann. A questo punto si può procedere ad effettuare le misure delle grandezze che interessano: energia interna, magnetizzazione, ....

Per quanto riguarda la configurazione di prova, essa viene scelta in maniera uniforme in una classe di configurazioni. Nel caso più semplice si inverte uno spin a caso: questo corrisponde a considerare tutte equiprobabili le configurazioni che differiscono per uno spin dalla configurazione di partenza. La condizione di accessibilità risulta soddisfatta. Infatti, come si è visto, ogni configurazione di prova ha probabilità non nulla di essere accettata, e, invertendo di volta in volta anche un solo spin, si può da una qualsiasi configurazione di partenza arrivare ad una qualsiasi altra configurazione. Notiamo che per decidere sull'accettazione della configurazione di prova va valutato il rapporto

$$\frac{p_{\mu'}}{p_{\mu}} = e^{-\beta(E_{\mu'} - E_{\mu})}.$$

Poichè questi modelli hanno interazioni a corto range, la variazione di energia coinvolgerà soltanto un piccolo blocco di spin più vicini allo spin che è stato invertito.

ESERCIZIO: Applicare il metodo di Metropolis al modello di Ising con interazione di primi vicini e in assenza di campo magnetico esterno.

## 3.6 L'autocorrelazione

È chiaro che nell'algoritmo appena descritto la configurazione al passo  $n$ -simo della catena di Markov dipende fortemente da quella al passo  $(n-1)$ -simo. I valori della quantità  $Q$  misurati in passi che sono vicini tra di loro non sono quindi statisticamente indipendenti. Occorre quindi trovare un modo per decidere quanti passi bisogna attendere perchè il sistema abbia perso memoria del precedente valore di  $Q$  ed abbia quindi senso ripetere la misura.

Misura della dipendenza statistica è l'autocorrelazione per la quantità in gioco, cioè

$$C(k) = \frac{\langle Q_{\mu_i} Q_{\mu_{i+k}} \rangle - \langle Q_{\mu_i} \rangle^2}{\langle Q_{\mu_i}^2 \rangle - \langle Q_{\mu_i} \rangle^2}$$

Come si è già visto, per una sequenza completamente scorrelata

$$C(k) = \delta_{k,0}$$

Nella pratica si misura  $C(k)$  durante i primi passi del processo e si calcola il valore di  $\bar{k}$  tale che  $C(\bar{k})$  sia piccola (ad esempio  $0.05 = 5\%$ ). Quindi si prosegue accumulando il valore di  $Q$  ogni  $\bar{k}$  passi del Monte Carlo. Il valore di  $\bar{k}$  dipenderà dalla temperatura e dal particolare processo di Markov considerato. Algoritmi che non invertono un solo spin, ma interi blocchi, sono in grado di abbattere notevolmente il valore di  $\bar{k}$ .



# Bibliografia

- [1] N. Cufaro Petroni. *Teoria della Probabilità*. Edizioni dal Sud, Modugno (Bari), 1996.
- [2] F. James. Monte carlo theory and practice. *Rep. Prog. Phys.*, 43:1145, 1980.
- [3] H. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, 1999, Oxford.
- [4] K. P. N. Murthy. *Monte Carlo: basics*. arXiv: cond-mat/0104215 v1, 2001.
- [5] K. Huang. *Meccanica Statistica*. Zanichelli, Bologna, 1997.