

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

---

SCUOLA DI SCIENZE  
Corso di Laurea in Matematica

# Il Metodo Monte Carlo

Tesi di Laurea in Probabilità e Statistica

Relatore:  
Chiar.mo Prof.  
Andrea Pascucci

Presentata da:  
Lorenzo Drumond

---

---

Sessione Unica  
Anno Accademico 2018/2019



# Indice

<b>Introduzione</b>	<b>i</b>
<b>1 Introduzione al Metodo</b>	<b>1</b>
1.1 Notazioni . . . . .	1
1.2 Legge dei grandi numeri . . . . .	3
1.3 Metodo Monte Carlo . . . . .	4
<b>2 Approssimazione dell'errore in MCM</b>	<b>7</b>
2.1 Teorema Centrale del Limite . . . . .	7
2.2 Teorema di Berry-Esseen . . . . .	9
<b>3 Generatori di Numeri Pseudocasuali</b>	<b>19</b>
3.1 Linear Congruential Generator . . . . .	20
3.2 Combinazioni di Generatori . . . . .	21
3.3 Metodi di Campionamento . . . . .	22
3.3.1 Metodo della Trasformazione Inversa . . . . .	23
3.3.2 Metodo di Acceptance-Rejection . . . . .	25
<b>4 Esempi e Applicazioni</b>	<b>29</b>
4.1 Implementazione . . . . .	29
4.2 Approssimazione di $\pi$ . . . . .	30
4.3 Un esempio di Business Planning . . . . .	31
<b>Bibliografia</b>	<b>35</b>



# Introduzione

Il Metodo Monte Carlo è uno strumento indispensabile in moltissimi ambiti, dalla finanza all'ingegneria, dalla ricerca scientifica alla matematica, ma il suo sviluppo e crescente utilizzo sono cosa recente, e conseguente alla creazione dei primi computer. Al giorno d'oggi, "Monte Carlo" si riferisce ad un qualunque metodo che utilizzi sequenze di numeri casuali per condurre simulazioni statistiche. Questo metodo nacque durante la Seconda Guerra Mondiale grazie all'avvenimento concomitante di due eventi: la creazione del primo computer, ENIAC; e la costruzione della bomba atomica.

Furono questi eventi a dare la spinta allo sviluppo del metodo: la difficoltà della risoluzione dei problemi di carattere termonucleare, come la descrizione del moto dei neutroni all'interno di una sfera di materiale fissile, unita alla possibilità di eseguire decine di migliaia di calcoli in breve tempo, fecero resuscitare quelli che erano una volta noti come "metodi di campionamento statistico", e che erano caduti nel dimenticatoio proprio a causa della lunghezza e difficoltà dei calcoli. Ma con la nascita del computer, i matematici Stan Ulan e John von Neumann videro le potenzialità di questo metodo.

Il metodo conobbe una rapidissima crescita e la sua applicazione si estese oltre al campo della fisica, arrivando a toccare il calcolo di integrali in  $n$ -dimensioni e la risoluzione di equazioni differenziali alle derivate parziali.

L'obiettivo di questa tesi è quello di introdurre le tecniche di approssimazione numerica note come Metodo Monte Carlo. Dopo una breve presentazione dell'apparato teorico che giustifica il funzionamento di questo metodo nel Capitolo 1, nel secondo Capitolo si prosegue con uno studio sulla stima dell'errore, focalizzando l'interesse sul Teorema Centrale del Limite e sulla dimostrazione del Teorema di Berry-Esseen. Il Capitolo 3 affronta il tema della generazione di numeri pseudocasuali e i metodi di campionamento di variabili aleatorie come il Metodo della Trasformazione Inversa e quello di Acceptance-Rejection, fondamentali nel Metodo Monte Carlo. La tesi si conclude col quarto Capitolo, presentando tre esempi di generazione di numeri pseudocasuali e applicazione del metodo Monte Carlo.



# Capitolo 1

## Introduzione al Metodo

### 1.1 Notazioni

**Definizione 1.** Uno spazio di probabilità  $(\Omega, \mathcal{F}, P)$  è costituito da un insieme non vuoto  $\Omega$ , una  $\sigma$ -algebra  $\mathcal{F}$  e da una misura  $P$ , detta "di probabilità", su  $\mathcal{F}$  tale che  $P(\Omega) = 1$ .

Dato  $B \in \mathcal{F}, P(B) > 0$ , la probabilità di  $A \in \mathcal{F}$  condizionata a  $B$  è data da

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

**Definizione 2.** Consideriamo uno spazio metrico  $(\mathbb{M}, \rho)$ , e sia  $\mathcal{B}_\rho$  la  $\sigma$ -algebra di Borel su  $\mathbb{M}$ . Una *distribuzione* è una misura di probabilità su  $(\mathbb{M}, \mathcal{B}_\rho)$ .

Consideriamo il caso  $\mathbb{M} = \mathbb{R}^d$ . Indicheremo  $\mathcal{B}_\rho$  con  $\mathcal{B}_d$ .

**Definizione 3.** Fissato  $x_0 \in \mathbb{R}^d$ , la distribuzione *Delta di Dirac*  $\delta_{x_0}$  centrata in  $x_0$  è data da

$$\delta_{x_0}(H) = \begin{cases} 1 & x_0 \in H \\ 0 & x_0 \notin H \end{cases} \quad H \in \mathcal{B}_d.$$

Una *distribuzione discreta* è una distribuzione della forma

$$\mu(H) := \sum_{n=1}^{\infty} p_n \delta_{x_n}(H), \quad H \in \mathcal{B}_d,$$

con  $(x_n)_{n \in \mathbb{N}}$  successione in  $\mathbb{R}^d$ , e  $p_n \geq 0, \sum_{n=1}^{\infty} p_n = 1$ .

**Esempio 1.1.** La distribuzione uniforme discreta su  $H$  è data da

$$\text{Unif}_H = \frac{1}{n} \sum_{k=1}^n \delta_{x_k},$$

con  $H = \{x_1, \dots, x_n\}$  sottoinsieme finito di  $\mathbb{R}^d$ .

**Definizione 4.** Una distribuzione  $\mu$  su  $\mathcal{B}_d$  è *assolutamente continua* (e si scrive  $\mu \in AC$ ) se esiste una funzione  $\gamma: \mathbb{R}^d \rightarrow [0, \infty[$  tale che

$$\int_{\mathbb{R}^d} \gamma(x) dx = 1 \quad \text{e} \quad \mu(H) = \int_H \gamma(x) dx, \quad H \in \mathcal{B}_d.$$

La funzione  $\gamma$  è detta *densità* di  $\mu$ , o PDF.

**Esempio 1.2.** Dato un sottoinsieme Lebesgue-misurabile di misura  $> 0$   $K \in \mathbb{R}^d$ , la distribuzione uniforme  $\text{Unif}_K$  su  $K$  ha densità

$$\gamma = \frac{1}{\text{Leb}_d(K)} \chi_K,$$

dove  $\text{Leb}_d$  indica la misura di Lebesgue in  $\mathbb{R}^d$  e  $\chi_K$  la funzione caratteristica<sup>1</sup> di  $K$ .

**Esempio 1.3.** La distribuzione normale reale  $\mathcal{N}_{\mu, \sigma^2}$  di parametri  $\mu \in \mathbb{R}, \sigma > 0$  è la distribuzione su  $\mathcal{B}_1$  con densità

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}.$$

$\mathcal{N}_{0,1}$  è detta distribuzione normale standard.

**Definizione 5.** La *funzione di ripartizione* (abbreviata CDF) di una distribuzione  $\mu$  su  $(\mathbb{R}, \mathcal{B}_1)$  è definita da

$$F_\mu(x) := \mu([-\infty, x]), \quad x \in \mathbb{R}.$$

Si vede facilmente che una CDF è sempre monotona debolmente crescente, continua da destra e tale che

$$\lim_{x \rightarrow -\infty} F_\mu(x) = 0, \quad \lim_{x \rightarrow \infty} F_\mu(x) = 1.$$

**Definizione 6.** Una *variabile aleatoria* su  $(\Omega, \mathcal{F}, P)$  a valori in  $\mathbb{R}^d$  è una funzione

$$X: \Omega \rightarrow \mathbb{R}^d$$

tale che

$$(X \in H) := X^{-1}(H) \in \mathcal{F}, \quad H \in \mathcal{B}_d.$$

cioè  $X \in m\mathcal{F}$ .

Indichiamo con  $\sigma(X) := \{X^{-1}(H) | H \in \mathcal{B}_d\}$ .

**Definizione 7.** Data una variabile aleatoria  $X: \Omega \rightarrow \mathbb{R}^d$ , la distribuzione  $\mu_X$  su  $\mathbb{R}^d$  associata a  $X$  è data da

$$\mu_X(H) = P(X \in H),$$

e si scrive  $X \sim \mu_X$ .

**Definizione 8.** In uno spazio di probabilità  $(\Omega, \mathcal{F}, P)$ , il *valore atteso* di una variabile aleatoria integrabile è definito da

$$E[X] := \int_{\Omega} X dP.$$

Siano  $X \in L^1(\Omega, \mathcal{F}, P)$  a valori in  $\mathbb{R}^d$  e  $\mathcal{G}$  una sotto- $\sigma$ -algebra di  $\mathcal{F}$ . Esiste una v.a.  $Z \in L^1(\Omega, P)$  a valori in  $\mathbb{R}^d$  tale che

<sup>1</sup>In generale,  $\chi_H$  indica la funzione caratteristica dell'insieme  $H$ .



1.  $Z \in m\mathcal{G}$ ;
2. per ogni v.a.  $W \in m\mathcal{G}$  limitata, vale

$$E[ZW] = E[XW].$$

Scriviamo  $Z = E[X|\mathcal{G}]$  e diciamo che  $Z$  è una *versione dell'attesa condizionata* di  $X$  a  $\mathcal{G}$ .

Se  $\mathcal{G} = \sigma(Y)$ ,  $Y$  v.a. su  $(\Omega, \mathcal{F}, P)$ , scriviamo  $Z = E[X|Y]$ . Se  $X = \chi_F$ ,  $F \in \mathcal{F}$ , scriviamo  $Z = P(F|\mathcal{G})$ .

**Definizione 9.** Sia  $X: \Omega \rightarrow \mathbb{R}^d$  una variabile aleatoria su  $(\Omega, \mathcal{F}, P)$ . La funzione

$$\varphi_X: \mathbb{R}^d \rightarrow \mathbb{C}$$

definita da

$$\varphi_X(\eta) = E[e^{i\langle \eta, X \rangle}] = E[\cos\langle \eta, X \rangle] + iE[\sin\langle \eta, X \rangle], \quad \eta \in \mathbb{R}^d,$$

è detta *funzione caratteristica*, o CHF, di  $X$ .

**Definizione 10.** Diciamo che le v.a.  $X_1, \dots, X_n$  definite nello spazio  $(\Omega, \mathcal{F}, P)$  sono *indipendenti* in  $P$  se vale

$$P\left(\bigcap_{i=1}^n (X_i \in H_i)\right) = \prod_{i=1}^n P(X_i \in H_i), \quad H_i \in \mathcal{B}_i, i = 1, \dots, n.$$

Da qui in avanti mi riferirò al metodo Monte Carlo con **MCM**, e indicherò le variabili aleatorie con v.a.

## 1.2 Legge dei grandi numeri

**MCM** si basa sulla Legge debole dei grandi numeri.

**Lemma 1.2.1 (Disuguaglianza di Markov).** Per ogni  $X$  v.a. a valori in  $\mathbb{R}^d$ ,  $\epsilon > 0$  e  $p \in [0, +\infty[$  vale

$$P(|X| \geq \epsilon) \leq \frac{E[|X|^p]}{\epsilon^p} \quad (1.1)$$

In particolare, se  $Y \in L^2(\Omega, P)$  è una v.a. reale allora si ha

$$P(|Y - E[Y]| \geq \epsilon) \leq \frac{\text{var}(Y)}{\epsilon^2} \quad (1.2)$$

*Dimostrazione.* Per la prima disuguaglianza, se  $E[|X|^p] = +\infty$  non c'è nulla da provare, altrimenti vediamo che, per monotonia

$$E[|X|^p] \geq E[|X|^p \chi_{(|X| \geq \epsilon)}] \geq \epsilon^p E[\chi_{(|X| \geq \epsilon)}] = \epsilon^p P(|X| \geq \epsilon)$$

dove  $\chi_{(|X| \geq \epsilon)}$  è la funzione caratteristica dell'insieme  $(|X| \geq \epsilon) = \{\omega \in \Omega \mid |X(\omega)| \geq \epsilon\}$ . La seconda disuguaglianza è un'ovvia conseguenza, ricordando che  $\text{var}(Y) = E[(Y - E[Y])^2]$ . □

**Teorema 1.2.2 (Legge debole dei grandi numeri).** Sia  $(X_n)_{n \in \mathbb{N}}$  una successione di v.a. reali scorrelate<sup>2</sup>

$$E[X_n] = \mu, \quad \text{var}(X_n) \leq C, \quad n \in \mathbb{N},$$

con  $\mu, C \in \mathbb{R}$ . Posto

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \quad n \in \mathbb{N},$$

si ha

$$\text{var}(M_n) = E[(M_n - \mu)^2] \leq \frac{C}{n}, \quad (1.3)$$

e, per ogni  $\epsilon > 0$ ,

$$P(|M_n - \mu| \geq \epsilon) \leq \frac{C}{n\epsilon^2}. \quad (1.4)$$

Quindi  $M_n$  converge in norma  $L^2$  e in probabilità al valore medio  $\mu$

*Dimostrazione.* Per linearità

$$E[M_n] = \frac{1}{n} \sum_{k=1}^n E[X_k] = \mu.$$

Per l'ipotesi di scorrelazione

$$E[(M_n - \mu)^2] = \text{var}(M_n) = \frac{\text{var}(X_1 + \cdots + X_n)}{n^2} = \frac{\text{var}(X_1) + \cdots + \text{var}(X_n)}{n^2} \leq \frac{C}{n},$$

che prova (1.3). Combinando (1.3) con (1.2) otteniamo (1.4). □

Data una v.a.  $X_1 \in L^2(\Omega, P)$ , è sempre possibile costruire una successione di v.a. indipendenti e identicamente distribuite (i.i.d.). Allora il Teorema 1.2.2 mostra come, per  $n \gg 1$ , la media aritmetica  $M_n = \frac{X_1 + \cdots + X_n}{n}$  assume molto probabilmente un valore "vicino a  $E[X_1]$ ".

### 1.3 Metodo Monte Carlo

Siamo pronti a delineare il principio di funzionamento del **MCM**: il suo scopo è quello di calcolare il valore atteso di una variabile aleatoria.

Consideriamo una v.a.  $X$  e una funzione  $f \in m\mathcal{B}_1$  tale che  $f(X) \in L^2(\Omega, P)$ . Allora, per (1.4), vale

$$E[f(X)] \approx \frac{1}{n} \sum_{k=1}^n f(X^{(k)}), \quad \text{per } n \gg 1, \quad (1.5)$$

dove  $X^{(1)}, \dots, X^{(n)}$  indicano dei valori casuali della v.a.  $X$  generati in modo indipendente.

---

<sup>2</sup>Date due v.a. reali  $X, Y \in L^2(\Omega, P)$ ,  $\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])]$ . Sono scorrelate se  $\text{cov}(X, Y) = 0$ .

Dunque **MCM** permette di approssimare numericamente il valore atteso di una funzione di una v.a. di cui si sia in grado di generare dei valori casuali in modo indipendente.

Un primo esempio di come **MCM** può essere applicato è il seguente: consideriamo una v.a.  $X \sim \text{Unif}_{[0,1]}$  e  $f \in L^1([0,1])$ . Il Teorema del Calcolo della Media afferma che se  $Y$  è una v.a. su  $\mathbb{R}^d$  tale che  $Y \sim \mu_Y$  e  $f: \mathbb{R}^d \rightarrow \mathbb{R}^N$ ,  $f \in m\mathcal{B}_d$ ,  $f(Y) \in L^1(\Omega, P)$  allora

$$E[f(Y)] = \int_{\mathbb{R}^d} f(y) \mu_Y(dy).$$

Pertanto

$$E[f(X)] = \int_{\Omega} f(X) dP = \int_{\mathbb{R}} f(x) \chi_{[0,1]} dx = \int_0^1 f(x) dx$$

e quindi possiamo usare **MCM** per approssimare il valore di un integrale, anche di dimensione maggiore di uno. I principali vantaggi rispetto ad altri metodi sono:

- nessuna ipotesi di regolarità sulla funzione integranda  $f$ ;
- l'ordine di convergenza del metodo è indipendente dalla dimensione;

Al momento **MCM** è l'unico metodo numerico conosciuto per risolvere problemi di grandi dimensioni che tipicamente sorgono nelle applicazioni reali. Data la natura di approssimazione probabilistica, le problematiche che si affrontano nell'utilizzo di questo metodo sono dunque quelle di stima dell'errore e di generazione di valori casuali e indipendenti.



# Capitolo 2

## Approssimazione dell'errore in MCM

Due strumenti fondamentali nella stima dell'errore di MCM sono il Teorema Centrale del Limite e il Teorema di Berry-Esseen.

### 2.1 Teorema Centrale del Limite

Il seguente lemma ci permette di dimostrare il Teorema centrale del limite:

**Lemma 2.1.1 (Teorema di Lévy).**<sup>1</sup> Sia  $(\mu_n)_{n \in \mathbb{N}}$  successione di distribuzioni reali, e sia  $(\varphi_n)_{n \in \mathbb{N}}$  la successione delle corrispondenti funzioni caratteristiche CHF. Allora:

1. se  $\mu_n \xrightarrow{d} \mu$  (convergenza debole di distribuzioni) allora  $\varphi_n$  converge puntualmente alla CHF di  $\mu$ , ossia  $\varphi_n(\eta) \xrightarrow[n \rightarrow \infty]{} \varphi(\eta)$  per ogni  $\eta \in \mathbb{R}$
2. viceversa, se  $\varphi_n$  converge puntualmente a una funzione  $\varphi$  continua in 0, allora  $\varphi$  è la CHF di una distribuzione  $\mu$  e vale  $\mu_n \xrightarrow{d} \mu$

**Teorema 2.1.2 (Teorema centrale del limite).** Sia  $(X_n)_{n \in \mathbb{N}}$  successione di v.a. reali i.i.d.<sup>2</sup> in  $L^2(\Omega, P)$ . Posto

$$\begin{aligned} \mu &:= E[X_1], & \sigma^2 &:= \text{var}(X_1), \\ S_n &:= X_1 + \dots + X_n, & M_n &:= \frac{S_n}{n}, \end{aligned}$$

allora  $E[M_n] = \mu$  e  $\text{var}(M_n) = \frac{\sigma^2}{n}$ . Definiamo l' $n$ -esima media aritmetica normalizzata come

$$\widetilde{M}_n := \frac{M_n - E[M_n]}{\sqrt{\text{var}(M_n)}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

Allora la successione delle medie aritmetiche normalizzate  $\widetilde{M}_n$  converge debolmente a una v.a. con distribuzione normale standard:

$$\widetilde{M}_n \xrightarrow{d} Z \sim \mathcal{N}_{0,1}$$

---

<sup>1</sup>Per la dimostrazione, si rimanda a [7].

<sup>2</sup>indipendenti e identicamente distribuite.

*Dimostrazione.* Utilizziamo il Lemma 2.1.1: poichè

$$\widetilde{M}_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{X_k - \mu}{\sigma},$$

allora abbiamo

$$\varphi_{\widetilde{M}_n}(\eta) = E \left[ e^{i \frac{\eta}{\sqrt{n}} \sum_{k=1}^n \frac{X_k - \mu}{\sigma}} \right] =$$

per indipendenza, dato che se  $X, Y$  indipendenti allora  $E[XY] = E[X]E[Y]$

$$= \left( E \left[ e^{i \frac{\eta}{\sqrt{n}} \frac{X_1 - \mu}{\sigma}} \right] \right)^n =$$

e poichè, se  $X \in L^p(\Omega, P)$  allora  $\varphi_X(\eta) = \sum_{k=0}^p \frac{E[(iX)^k]}{k!} \eta^k + o(\eta^p)$  (si veda [7]), allora

$$= \left( 1 + \frac{(i\eta)^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \xrightarrow{n \rightarrow \infty} e^{-\frac{\eta^2}{2}}$$

che è la CHF della distribuzione normale standard. Per il Lemma 2.1.1, questo conclude la prova. □

Il Teorema centrale del limite ci permette di dare una stima dell'errore numerico di **MCM**. Poniamo

$$p_\lambda := P \left( |M_n - \mu| \leq \lambda \frac{\sigma}{\sqrt{n}} \right) = P \left( |\widetilde{M}_n| \leq \lambda \right), \quad \lambda > 0;$$

allora per il Teorema centrale del limite si ha la stima, per  $n \gg 1$ :

$$p_\lambda \approx P(|Z| \leq \lambda), \quad Z \sim \mathcal{N}_{0,1}.$$

Del resto

$$\begin{aligned} P(|Z| \leq \lambda) &= P(-\lambda \leq Z \leq \lambda) = \\ &= P(Z \leq \lambda) - P(Z < -\lambda) = \\ &= F_Z(\lambda) - (1 - P(Z \geq -\lambda)) = \\ &= F_Z(\lambda) + P(-Z \leq \lambda) - 1 = \end{aligned}$$

poichè la distribuzione normale standard ha densità pari

$$= 2F_Z(\lambda) - 1$$

con  $F_Z$  la funzione di ripartizione di  $Z$ .

Per la stima dell'errore numerico del metodo Monte Carlo, si parte dai valori di probabilità  $p$  più usati comunemente, ossia  $p = 95\%$  e  $p = 99\%$ ; posto  $\lambda = F^{-1}\left(\frac{p+1}{2}\right)$ , si ottiene

$$P\left(|M_n - \mu| \leq 1,96 \frac{\sigma}{\sqrt{n}}\right) \approx 95\% \quad \text{e} \quad P\left(|M_n - \mu| \leq 2,57 \frac{\sigma}{\sqrt{n}}\right) \approx 99\%.$$

Per questo motivo

$$r_{95} := 1,96 \frac{\sigma}{\sqrt{n}} \quad \text{e} \quad r_{99} := 2,57 \frac{\sigma}{\sqrt{n}}$$

sono detti *raggi degli intervalli di confidenza al 95% e al 99% per  $\mu$* : se  $M_n$  è il risultato aleatorio dell'approssimazione di **MCM** per  $\mu$ , allora

$$[M_n - r_{95}, M_n + r_{95}] \quad \text{e} \quad [M_n - r_{99}, M_n + r_{99}]$$

sono gli intervalli a cui  $\mu$  appartiene con probabilità rispettivamente pari al 95% e al 99%.

## 2.2 Teorema di Berry-Esseen

Sotto ipotesi più forti, una stima esplicita della velocità di convergenza per il Teorema centrale del limite è data dal Teorema di Berry-Esseen.

**Teorema 2.2.1 (Teorema di Berry-Esseen).** *Sia  $(X_n)_{n \in \mathbb{N}}$  successione di v.a. reali i.i.d in  $L^3(\Omega, P)$  tali che  $E[X_i] = 0$ ,  $E[X_i^2] = \sigma^2$  e  $E[X_i^3] = \rho < \infty$ . Se  $F_n$  indica la CDF di  $\widetilde{M}_n = \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}$  e  $N$  la CDF della distribuzione normale standard  $\mathcal{N}_{0,1}$ , allora si ha*

$$|F_n(x) - N(x)| \leq \frac{3\rho}{\sigma^3\sqrt{n}}, \quad x \in \mathbb{R}, n \in \mathbb{N} \quad (2.1)$$

**Osservazione 2.1.** La costante 3 non è il valore migliore che si può ottenere. Al momento non è noto il valore ottimale per tale costante: si sa solo che è compreso tra 0.409 e 0.56

**Osservazione 2.2.** Per la disuguaglianza di Jensen

$$\sigma^3 = (\sigma^2)^{\frac{3}{2}} = \left|E[X_i^2]\right|^{\frac{3}{2}} \leq E\left[|X_i|^2\right]^{\frac{3}{2}} = \rho$$

Pertanto  $\rho \geq \sigma^3$ . Inoltre, se  $n \leq 9$  allora  $\frac{3}{\sqrt{n}} \geq 1$  e quindi

$$\frac{3\rho}{\sigma^3\sqrt{n}} \geq 1$$

il che rende la (2.1) triviale. Assumeremo quindi che  $n \geq 10$  da qui in avanti.

Per dimostrare il Teorema 2.2.1 ci dovremo servire di diversi lemmi e della seguente funzione di densità<sup>3</sup>

$$h_L(x) = \frac{1 - \cos(Lx)}{\pi Lx^2} \quad (2.2)$$

con CHF associata

$$\omega_L(\theta) = \left(1 - \left|\frac{\theta}{L}\right|\right)^+ \quad \text{per } |\theta| \leq L \quad (2.3)$$

Indicheremo la CDF corrispondente con  $H_L$ . Inoltre, seguendo la notazione di [1], date due CDF  $F$  e  $G$  definiamo l'operazione di convoluzione

$$(F * G)(z) := \int_{\mathbb{R}} F(z - y) dG(y) \quad (2.4)$$

dove intendiamo l'integrale calcolato rispetto alla distribuzione associata a  $G$  (questo è il significato della scrittura  $dG(y)$ ).

**Lemma 2.2.2 (Lemma di Riemann-Lebesgue).** *Se  $f \in L^1(\mathbb{R})$ , allora la trasformata di Fourier di  $f$  soddisfa*

$$\hat{f}(z) := \int_{\mathbb{R}} f(x) e^{-izx} dx \longrightarrow 0 \quad \text{per } |z| \longrightarrow \infty \quad (2.5)$$

*Dimostrazione.* Se  $f = \chi_{(a,b)}(x)$ ,  $a, b \in \mathbb{R}$  allora

$$\int_{\mathbb{R}} f(x) e^{-izx} dx = \int_a^b e^{-izx} dx = \frac{e^{-iza} - e^{-izb}}{iz} \xrightarrow{|z| \rightarrow \infty} 0.$$

Se  $f = \sum_{k=1}^N x_k \chi_{(a_k, b_k)}$ ,  $x_k, a_k, b_k \in \mathbb{R}$  allora, per linearità dell'integrale di Lebesgue

$$\int_{\mathbb{R}} f(x) e^{-izx} dx = \int_{\mathbb{R}} \sum_{k=1}^N x_k \chi_{(a_k, b_k)} e^{-izx} dx = \sum_{k=1}^N x_k \int_{\mathbb{R}} e^{-izx} \chi_{(a_k, b_k)} dx \xrightarrow{|z| \rightarrow \infty} 0.$$

Infine, sia  $f \in L^1(\mathbb{R})$  arbitraria. Poichè le funzioni semplici sono dense in  $L^1$ , per ogni  $\epsilon > 0$  esiste una funzione semplice  $g$  tale che

$$\int_{\mathbb{R}} |f(x) - g(x)| dx < \epsilon$$

Allora, per quanto visto prima, esiste  $N \in \mathbb{R}$  tale che

$$\left| \int_{\mathbb{R}} g(x) e^{-izx} dx \right| < \epsilon \quad \text{per ogni } |z| > N$$

Per additività degli integrali e per disuguaglianza triangolare si ha, per  $|z| > N$

$$\left| \int_{\mathbb{R}} f(x) e^{-izx} dx \right| \leq \int_{\mathbb{R}} |f(x) - g(x)| dx + \left| \int_{\mathbb{R}} g(x) e^{-izx} dx \right| \leq 2\epsilon$$

cioè

$$\lim_{|z| \rightarrow \infty} \hat{f}(z) = 0$$

□

---

<sup>3</sup>si veda [1, p. 135]



**Lemma 2.2.3.** *Vale<sup>4</sup>*

$$\left| e^{ix} - \sum_{m=0}^n \frac{(ix)^m}{m!} \right| \leq \frac{|x|^{n+1}}{(n+1)!} \quad \text{per ogni } n \in \mathbb{N}, x \in \mathbb{R} \quad (2.6)$$

*Dimostrazione.* Integriamo per parti il seguente integrale:

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1} \int_0^x (x-s)^{n+1} e^{is} ds$$

Dunque, per  $n = 0$  vale

$$\int_0^x e^{is} ds = x + i \int_0^x (x-s) e^{is} ds$$

da cui

$$e^{ix} = 1 + ix + i^2 \int_0^x (x-s) e^{is} ds$$

Se  $n = 1$  allora

$$\int_0^x (x-s) e^{is} ds = \frac{x^2}{2} + \frac{i}{2} \int_0^x (x-s)^2 e^{is} ds$$

e quindi ricaviamo

$$e^{ix} = 1 + ix + \frac{(ix)^2}{2} + \frac{i^3}{2} \int_0^x (x-s)^2 e^{is} ds$$

Iterando questo procedimento otteniamo

$$e^{ix} - \sum_{m=0}^n \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds$$

Poichè  $|e^{is}| \leq 1$  per ogni  $s \in \mathbb{R}$ , allora

$$\left| \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds \right| \leq \frac{|x|^{n+1}}{(n+1)!}$$

□

Applicando il valore atteso ad entrambi i membri della disuguaglianza, dove poniamo  $x = tX$ ,  $X$  v.a., e usando la disuguaglianza di Jensen (essendo  $|\cdot|$  funzione convessa) si ha

$$\left| E[e^{itX}] - \sum_{m=0}^n E\left[\frac{(itX)^m}{m!}\right] \right| \leq E\left[\frac{|tX|^{n+1}}{(n+1)!}\right] \quad (2.7)$$

**Lemma 2.2.4.** *Siano  $X, Y$  v.a. indipendenti su  $(\Omega, \mathcal{F}, P)$  a valori in  $\mathbb{R}^d$ . Se  $X \in AC$  con densità  $\gamma_X$ , allora anche  $X + Y \in AC$  con densità*

$$\gamma_{X+Y}(z) = \int_{\mathbb{R}^d} \gamma_X(z-y) \mu_Y(dy), \quad z \in \mathbb{R}^d \quad (2.8)$$

<sup>4</sup>in realtà si può ottenere una approssimazione dell'errore migliore, ma per i nostri scopi questo è sufficiente.

*Dimostrazione.* Sia  $H \in m\mathcal{B}_d$ . Allora

$$\begin{aligned} P(X + Y \in H) &= E[\chi_H(X + Y)] = \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \chi_H(x + y) \mu_{(X,Y)}(dx dy) = \end{aligned}$$

per indipendenza

$$= \int_{\mathbb{R}^d \times \mathbb{R}^d} \chi_H(x + y) (\mu_X \otimes \mu_Y) =$$

per Fubini e l'assoluta continuità di  $X$

$$= \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \chi_H(x + y) \gamma_X(x) dx \right) \mu_Y(dy) =$$

applicando il cambiamento di variabili  $z = x + y$  nell'integrale interno

$$= \int_{\mathbb{R}^d} \left( \int_H \gamma_X(z - y) dz \right) \mu_Y(dy) =$$

per Fubini

$$= \int_H \left( \int_{\mathbb{R}^d} \gamma_X(z - y) \mu_Y(dy) \right) dz$$

□

**Lemma 2.2.5.** *Siano  $F$  e  $G$  delle CDF con  $G'(x) \leq \lambda < \infty$ . Sia  $\Delta(x) = F(x) - G(x)$ ,  $\eta = \sup|\Delta|$ ,  $\Delta_L = \Delta * H_L$  e  $\eta_L = \sup|\Delta_L(x)|$ . Allora*

$$\eta_L \geq \frac{\eta}{2} - \frac{12\lambda}{\pi L} \quad \text{o equivalentemente} \quad \eta \leq 2\eta_L + \frac{24\lambda}{\pi L} \quad (2.9)$$

*Dimostrazione.* Per le proprietà delle CDF<sup>5</sup> si ha che  $\Delta \xrightarrow{x \rightarrow \pm\infty} 0$ ,  $F$  continua da destra e  $G$  continua (perchè derivabile). Allora esiste  $x_0 \in \mathbb{R}$  tale che  $\Delta(x_0) = \pm\eta$  oppure  $\Delta(x_0-) = \pm\eta$ . Infatti, se  $\eta = 0$  allora  $|\Delta(x)|$  è costante. Se  $\eta > 0$ , allora per ogni  $\epsilon > 0$  esiste  $x$  tale che  $\eta - |\Delta(x)| < \epsilon$ . Sia  $(x_n)_{n \in \mathbb{N}}$  una successione tale che  $\eta - |\Delta(x_n)| < \frac{1}{n}$ . La classe limite di una successione è sempre chiusa e non vuota, e nel caso di  $(x_n)_{n \in \mathbb{N}}$  non contiene  $\pm\infty$ , pertanto esiste  $x_0 \in \mathbb{R}$  tale che una sottosuccessione  $(x_{n_k})_{k \in \mathbb{N}}$  converge a  $x_0$ . Se  $x_0$  è un punto di continuità di  $|\Delta(x)|$ , allora

$$0 \leq \eta - |\Delta(x_{n_k})| < \frac{1}{n_k}$$

e quindi

$$\lim_{k \rightarrow \infty} (\eta - |\Delta(x_{n_k})|) = \eta - \lim_{k \rightarrow \infty} |\Delta(x_{n_k})| = \eta - |\Delta(x_0)| = 0$$

---

<sup>5</sup>si veda [7]

Se non lo è, allora è un punto di discontinuità di salto (essendo  $\Delta$  continua da destra). Se  $|\Delta(x_0)| \neq \eta$ , allora  $(x_{n_k})_{k \in \mathbb{N}}$  converge da sinistra a  $x_0$  e quindi

$$\lim_{k \rightarrow \infty} (\eta - |\Delta(x_{n_k})|) = \eta - \lim_{k \rightarrow \infty} |\Delta(x_{n_k})| = \eta - |\Delta(x_0-)| = 0$$

Assumiamo dunque, senza perdita di generalità, che  $\Delta(x_0) = \eta$ . Sia  $s > 0$ . Per la monotonia di  $F$  vale che  $F(x_0 + s) \geq F(x_0)$ . Inoltre, poichè  $G'(x) \leq \lambda$  allora per qualche  $c \in \mathbb{R}$  vale  $\frac{G(x_0+s)-G(x_0)}{s} = G'(c) \leq \lambda$ . Pertanto  $\Delta(x_0 + s) \geq F(x_0) - G(x_0) - \lambda s = \eta - \lambda s$ .

Ora poniamo  $\delta = \frac{\eta}{2\lambda}$  e  $t = x_0 + \delta$ ; allora se  $|x| \leq \delta$ , vale che  $\delta - x \geq 0$  e quindi

$$\Delta(t - x) = \Delta(x_0 + (\delta - x)) \geq \eta - \lambda(\delta - x) = \frac{\eta}{2} + \lambda x$$

Dunque

$$\Delta(t - x) \geq \begin{cases} \frac{\eta}{2} + \lambda x & \text{se } |x| \leq \delta \\ -\eta & \text{altrimenti} \end{cases}$$

Per stimare la convoluzione  $\Delta_L$  osserviamo che

$$2 \int_{\delta}^{\infty} h_L(x) dx \leq 2 \int_{\delta}^{\infty} \frac{2}{\pi L x^2} dx = \frac{4}{\pi L \delta}$$

Del resto,  $h_L$  è pari e quindi  $\int_{-R}^R x h_L(x) dx = 0$ ; pertanto

$$\eta_L \geq \Delta_L(t) = (\Delta * H_L)(t) = \int_{\mathbb{R}} \Delta(t - x) dH_L(x) \geq$$

per la proprietà delle distribuzioni assolutamente continue

$$\begin{aligned} &\geq \int_{-\delta}^{\delta} \left( \frac{\eta}{2} + \lambda x \right) h_L(x) dx + 2(-\eta) \int_{\delta}^{\infty} h_L(x) dx = \\ &= \frac{\eta}{2} \left( 1 - 2 \int_{\delta}^{\infty} h_L(x) dx \right) - 2\eta \int_{\delta}^{\infty} h_L(x) dx \geq \\ &\geq \frac{\eta}{2} - \frac{6\eta}{\pi L \delta} = \frac{\eta}{2} - \frac{12\lambda}{\pi L} \end{aligned}$$

□

**Lemma 2.2.6.** *Siano  $K_1$  e  $K_2$  CDF di v.a. con valore atteso nullo, le cui CHF  $\kappa_i$  sono sommabili. Allora*

$$K_1(x) - K_2(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-e^{-itx}) \frac{\kappa_1(t) - \kappa_2(t)}{it} dt \quad (2.10)$$

*Dimostrazione.* Poichè  $\kappa_i \in L^1(\mathbb{R})$ , allora per la formula di inversione la densità delle rispettive v.a. è

$$k_i(y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ity} \kappa_i(t) dt.$$

Per definizione vale

$$K_i(x) = \int_{-\infty}^x k_i(y) dy$$

Posto  $\Delta K := K_1 - K_2$ , allora si ha

$$\begin{aligned} \Delta K(x) - \Delta K(a) &= \int_a^x (k_1(y) - k_2(y)) dy = \\ &= \frac{1}{2\pi} \int_a^x \left[ \int_{\mathbb{R}} e^{-ity} (\kappa_1(t) - \kappa_2(t)) dt \right] dy = \end{aligned}$$

per Fubini

$$= \frac{1}{2\pi} \int_{\mathbb{R}} (\kappa_1(t) - \kappa_2(t)) \left( \frac{e^{-ita} - e^{-itx}}{it} \right) dt$$

Si verifica facilmente che la funzione

$$\frac{\kappa_1(t) - \kappa_2(t)}{it}$$

è sommabile. Per le proprietà delle CDF,  $\Delta K(a) \rightarrow 0$  per  $a \rightarrow \infty$ . Inoltre, per il Lemma 2.2.2

$$\int_{\mathbb{R}} \frac{\kappa_1(t) - \kappa_2(t)}{it} e^{-ita} dt \rightarrow 0 \quad \text{per } a \rightarrow \infty.$$

Dunque troviamo che

$$K_1(x) - K_2(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-e^{-itx}) \frac{\kappa_1(t) - \kappa_2(t)}{it} dt$$

□

Adesso possiamo passare alla dimostrazione del Teorema di Berry-Esseen

*Dimostrazione del teorema 2.2.1.* Assumiamo inizialmente che  $\sigma^2 = 1$ .

Siano ora  $F, G$  delle CDF che soddisfino le ipotesi dei lemmi 2.2.5 e 2.2.6, con CHF  $\varphi_F$  e  $\varphi_G$ . Poniamo  $F_L := F * H_L$  e  $G_L := G * H_L$ . Vediamo allora che

$$\begin{aligned} F_L(z) &= \int_{\mathbb{R}} H_L(z - y) dF(y) = \\ &= \int_{\mathbb{R}} \left( \int_{-\infty}^{z-y} h_L(x) dx \right) dF(y) = \end{aligned}$$

applicando la sostituzione  $x = \zeta - y$  nell'integrale interno

$$= \int_{\mathbb{R}} \left( \int_{-\infty}^z h_L(\zeta - y) d\zeta \right) dF(y) =$$

per Fubini

$$= \int_{-\infty}^z \left( \int_{\mathbb{R}} h_L(\zeta - y) dF(y) \right) d\zeta$$

cioè, per il lemma 2.2.4,  $F_L$  è la CDF della somma di v.a. indipendenti  $X$  (con CDF  $F$ ) e  $Y$  (con CDF  $H_L$ ). Pertanto, la CHF di  $F_L$  è data da

$$\varphi_{X+Y}(t) = E\left[e^{it(X+Y)}\right] = E\left[e^{itX}\right]E\left[e^{itY}\right] = \varphi_F(t)\omega_L(t)$$

Analogamente, la CHF di  $G_L$  è  $\varphi_G\omega_L$ . Applichiamo allora a  $F$  e  $G$  il lemma 2.2.6: abbiamo

$$\begin{aligned} |F_L(x) - G_L(x)| &= \frac{1}{2\pi} \left| \int_{\mathbb{R}} (-e^{-itx}) \frac{\varphi_F(t)\omega_L(t) - \varphi_G(t)\omega_L(t)}{it} dt \right| \leq \\ &\leq \frac{1}{2\pi} \int_{\mathbb{R}} \frac{|\varphi_F(t)\omega_L(t) - \varphi_G(t)\omega_L(t)|}{|t|} dt \leq \end{aligned}$$

poichè  $|\omega_L(t)| \leq 1$  per  $|t| \leq L$  allora

$$\leq \frac{1}{2\pi} \int_{-L}^L \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|} dt$$

Applicando ora il lemma 2.2.5 si ottiene

$$|F(x) - G(x)| \leq \frac{1}{\pi} \int_{-L}^L \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|} dt + \frac{24\lambda}{\pi L}$$

dove  $\lambda = \sup_x G'(x)$ . Allora sostituendo  $F$  con  $F_n$  e  $G$  con  $N$  otteniamo

$$|F_n(x) - N(x)| \leq \frac{1}{\pi} \int_{-L}^L \frac{1}{|\theta|} \left| \varphi^n\left(\frac{\theta}{\sqrt{n}}\right) - \psi(\theta) \right| d\theta + \frac{24\lambda}{\pi L} \quad (2.11)$$

dove  $\varphi := \varphi_{X_1}$  è la CHF di  $X_1$ ,  $\varphi_{\widetilde{M}_n}(\theta) = \varphi^n\left(\frac{\theta}{\sqrt{n}}\right)$  è la CHF di  $\widetilde{M}_n$  e  $\psi(\theta) := e^{-\frac{\theta^2}{2}}$  è la CHF di  $\mathcal{N}_{0,1}$ .

Adesso dobbiamo cercare una stima del secondo membro della disuguaglianza; per quanto riguarda il secondo termine, osserviamo che

$$N(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

e quindi

$$N' = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

e si vede quindi facilmente che il punto di massimo di  $N'$  è  $x = 0$  e il suo massimo è  $\lambda = \frac{1}{\sqrt{2\pi}}$ . Per il primo termine, osserviamo che dati  $\alpha, \beta, \gamma \in \mathbb{R}, n \in \mathbb{N}$  tali che  $|\alpha|, |\beta| \leq \gamma$  allora

$$|\alpha^n - \beta^n| \leq n|\alpha - \beta|\gamma^{n-1} \quad (2.12)$$

Infatti

$$\begin{aligned} |\alpha^n - \beta^n| &= \left| (\alpha - \beta) \sum_{m=0}^{n-1} \alpha^{n-m-1} \beta^m \right| \leq \\ &\leq \sum_{m=0}^{n-1} |\alpha^{n-m} \beta^m - \alpha^{n-m-1} \beta^{m+1}| = \\ &= \sum_{m=0}^{n-1} |\alpha|^{n-m-1} |\beta|^m |\alpha - \beta| \leq n|\alpha - \beta| \gamma^{n-1} \end{aligned}$$

Applicando la (2.7) (ricordiamo che in generale, la CHF di una v.a  $X$  reale è definita come  $E[e^{itX}]$ ) abbiamo

$$\left| \varphi(t) - \left(1 - \frac{t^2}{2}\right) \right| \leq \frac{\rho|t|^3}{6} \quad (2.13)$$

quindi se  $t^2 \leq 2$  allora

$$|\varphi(t)| - \left(1 - \frac{t^2}{2}\right) \leq \left| \varphi(t) - \left(1 - \frac{t^2}{2}\right) \right|$$

e quindi

$$|\varphi(t)| \leq 1 - \frac{t^2}{2} + \frac{\rho|t|^3}{6} \quad (2.14)$$

Poniamo  $L = \frac{4\sqrt{n}}{3\rho}$ . Se  $|\theta| \leq L$  allora  $\frac{\rho|\theta|}{\sqrt{n}} \leq \frac{\rho L}{\sqrt{n}} = \frac{4}{3}$  e quindi

$$\begin{aligned} \left| \varphi\left(\frac{\theta}{\sqrt{n}}\right) \right| &\leq 1 - \frac{\theta^2}{2n} + \frac{\rho|\theta|^3}{6n^{3/2}} \leq 1 - \frac{\theta^2}{2n} + \frac{4}{3} \frac{\theta^2}{6n} = \\ &= 1 - \frac{5\theta^2}{18n} \leq e^{-\frac{5\theta^2}{18n}} \end{aligned}$$

dove nell'ultimo passaggio abbiamo usato la disuguaglianza elementare  $1 - x \leq e^{-x}$ .

Poniamo  $\alpha = \varphi\left(\frac{\theta}{\sqrt{n}}\right)$ ,  $\beta = e^{-\frac{\theta^2}{2n}}$  e  $\gamma = e^{-\frac{5\theta^2}{18n}}$ . Allora, usando (2.12) abbiamo

$$\left| \varphi^n\left(\frac{\theta}{\sqrt{n}}\right) - e^{-\frac{\theta^2}{2}} \right| \leq n \left| \varphi\left(\frac{\theta}{\sqrt{n}}\right) - e^{-\frac{\theta^2}{2n}} \right| \gamma^{n-1} \quad (2.15)$$

Osserviamo che  $\gamma^{n-1} \leq e^{-\frac{\theta^2}{4}}$ . Inoltre

$$n \left| \varphi\left(\frac{\theta}{\sqrt{n}}\right) - e^{-\frac{\theta^2}{2n}} \right| \leq n \left| \varphi\left(\frac{\theta}{\sqrt{n}}\right) - \left(1 - \frac{\theta^2}{2n}\right) \right| + n \left| \left(1 - \frac{\theta^2}{2n}\right) - e^{-\frac{\theta^2}{2n}} \right|$$

Per la (2.13) vale che

$$n \left| \varphi\left(\frac{\theta}{\sqrt{n}}\right) - \left(1 - \frac{\theta^2}{2n}\right) \right| \leq \frac{\rho|\theta|^3}{6n^{1/2}} \quad (2.16)$$

Per il secondo termine del membro a destra, consideriamo la seguente relazione:

$$\left| e^{-x} - (1 - x) \right| = \left| \frac{x^2}{2!} - \frac{x^3}{3!} + \dots \right| = \left| \sum_{k=2}^{\infty} \frac{(-x)^k}{k!} \right|$$

Se  $0 \leq x < 1$  allora  $x^k \geq x^{k+1}$ ; quindi  $\frac{x^k}{k!} \geq \frac{x^{k+1}}{(k+1)!}$ . Poichè la serie converge assolutamente, allora i suoi termini possono essere riarrangiati a piacimento senza modificare la somma della serie. Dunque

$$\sum_{k=2}^{\infty} \frac{(-x)^k}{k!} = \frac{x^2}{2!} + \sum_{k=2}^{\infty} \left( \frac{x^{2k}}{(2k)!} - \frac{x^{2k-1}}{(2k-1)!} \right) \leq \frac{x^2}{2!}$$

e quindi

$$|e^{-x} - (1-x)| \leq \frac{x^2}{2!}.$$

Del resto, avendo assunto  $|\theta| \leq L$ , ed essendo  $L = \frac{4\sqrt{n}}{3\rho} \leq \frac{4\sqrt{n}}{3} < \sqrt{2n}$ , allora  $0 \leq \frac{|\theta|}{\sqrt{2n}} < 1$ . Prendendo dunque  $x = \frac{\theta^2}{2n}$  soddisfiamo l'ipotesi della disuguaglianza ed abbiamo

$$n \left| \left(1 - \frac{\theta^2}{2n}\right) - e^{-\frac{\theta^2}{2n}} \right| \leq n \frac{x^2}{2} = \frac{\theta^4}{8n} \quad (2.17)$$

Pertanto, combinando la (2.16) con la (2.17) nella (2.15) otteniamo

$$\frac{1}{|\theta|} \left| \varphi^n \left( \frac{\theta}{\sqrt{n}} \right) - e^{-\frac{\theta^2}{2}} \right| \leq e^{-\frac{\theta^2}{4}} \left( \frac{\rho\theta^2}{6n^{\frac{1}{2}}} + \frac{|\theta|^3}{8n} \right) = \frac{1}{L} e^{-\frac{\theta^2}{4}} \left( \frac{2\theta^2}{9} + \frac{|\theta|^3}{6\sqrt{n}\rho} \right)$$

Avendo assunto  $n \geq 10$  allora  $\sqrt{n} \geq 3$ . Inoltre  $\rho \geq 1$  e quindi  $\frac{1}{\rho\sqrt{n}} \leq \frac{1}{3}$ , da cui deriviamo

$$\frac{1}{|\theta|} \left| \varphi^n \left( \frac{\theta}{\sqrt{n}} \right) - e^{-\frac{\theta^2}{2}} \right| \leq \frac{1}{L} e^{-\frac{\theta^2}{4}} \left( \frac{2\theta^2}{9} + \frac{|\theta|^3}{18} \right) \quad (2.18)$$

Sostituendo la (2.18) nella (2.11), ricordandoci che  $\lambda = \frac{1}{\sqrt{2\pi}} < \frac{2}{5}$  e osservando che entrambi i membri di (2.18) sono positivi, troviamo che

$$|F_n(x) - N(x)| \leq \frac{1}{\pi L} \left( \int_{\mathbb{R}} e^{-\frac{\theta^2}{4}} \left( \frac{2\theta^2}{9} + \frac{|\theta|^3}{18} \right) d\theta + 9,6 \right)$$

che si può riscrivere come

$$|F_n(x) - N(x)| \leq C \frac{\rho}{\sqrt{n}} \quad (2.19)$$

con

$$C = \frac{3}{4\pi} \left( \int_{\mathbb{R}} e^{-\frac{\theta^2}{4}} \left( \frac{2\theta^2}{9} + \frac{|\theta|^3}{18} \right) d\theta + 9,6 \right). \quad (2.20)$$

Calcoliamo la costante  $C$ . Sfruttando l'integrale Gaussiano, si trova che

$$\int_{\mathbb{R}} \frac{1}{\sqrt{(2\pi a)}} x^2 e^{-\frac{x^2}{2a}} dx = a \quad \text{per ogni } a \in \mathbb{R}.$$

Allora

$$\int_{\mathbb{R}} e^{-\frac{x^2}{4}} \frac{2x^2}{9} dx = \frac{2}{9} \sqrt{2\pi} \int_{\mathbb{R}} \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{4}} dx = \frac{8\sqrt{\pi}}{9}$$

mentre scrivendo  $x^3 = 2x^2 \cdot \frac{x}{2}$  e integrando per parti si ha

$$\int_{\mathbb{R}} e^{-\frac{x^2}{4}} \frac{|x|^3}{18} dx = \frac{1}{9} \left( \left[ -2x^2 e^{-\frac{x^2}{4}} \right]_0^{\infty} + \int_0^{\infty} 4x e^{-\frac{x^2}{4}} dx \right) = \frac{8}{9}.$$

Pertanto, sostituendo quanto trovato in (2.20) otteniamo

$$C = \frac{3}{4\pi} \left( \frac{8\sqrt{\pi}}{9} + \frac{8}{9} + 9,6 \right) < 3.$$

Quindi, sostituendo in (2.19) si ha la tesi per  $\sigma = 1$ .

Infine, consideriamo una successione di v.a.  $(X_n)_{n \in \mathbb{N}}$  i.i.d., con  $E[X_n] = 0$ ,  $E[X_n^2] = \sigma^2 \neq 1$  e  $E[|X_n|^3] = \rho$ . Consideriamo la successione  $\tilde{X}_n := \frac{X_n}{\sigma}$ . Allora  $E[\tilde{X}_n] = 0$ ,  $E[\tilde{X}_n^2] = 1$  e  $E[|\tilde{X}_n|^3] = \frac{\rho}{\sigma^3} =: \tilde{\rho}$ . Inoltre osserviamo che

$$\tilde{M}_n := \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} = \frac{\tilde{X}_1 + \dots + \tilde{X}_n}{\sqrt{n}}$$

e quindi, per quanto visto, se  $F_n$  è la CDF di  $\tilde{M}_n$  allora

$$|F_n(x) - N(x)| \leq \frac{3\tilde{\rho}}{\sqrt{n}} = \frac{3\rho}{\sigma^3\sqrt{n}}.$$

□



## Capitolo 3

# Generatori di Numeri Pseudocasuali

Come abbiamo visto, al cuore di **MCM** c'è una sequenza di numeri apparentemente casuali, usati per condurre la simulazione. Nello studio di **MCM** abbiamo considerato questi numeri come puramente casuali, ma all'atto pratico quei numeri sono in realtà generati da algoritmi deterministici, e quindi sono casuali solo in apparenza.

Per produrre tali sequenze ci si serve di un *generatore di numeri pseudocasuali*. Per capire cos'è, dobbiamo prima specificare cosa si intende per generatore di numeri casuali: esso è un algoritmo per la generazione di una sequenza di v.a.  $U_1, U_2, \dots$  tali che

(i) ogni  $U_i$  è uniformemente distribuita tra 0 e 1;

(ii) gli  $U_i$  sono indipendenti tra loro;

La proprietà (i) è una arbitraria ma conveniente normalizzazione; come vedremo più avanti, v.a. con distribuzione uniforme sull'intervallo unitario possono essere trasformate in v.a. con qualsiasi altra distribuzione. La proprietà (ii) è quella chiave: in particolare, implica che tutte le coppie di valori devono essere scorrelate e che, in generale, il valore di  $U_i$  non deve poter essere determinabile dai valori di  $U_1, \dots, U_{i-1}$

Un generatore di numeri pseudocasuali produce una sequenza finita di numeri  $u_1, \dots, u_K$  dell'intervallo unitario. Ciascuna di queste successioni costituisce un insieme dei valori possibili campionati da v.a.  $U_1, \dots, U_K$  indipendenti e uniformemente distribuite su  $[0, 1]$ . Un buon generatore soddisfa la condizione (effettivamente piuttosto vaga) che piccole (relativamente a  $K$ ) porzioni della sequenza  $u_1, \dots, u_K$  siano difficili da distinguere da un vero campionamento di v.a. uniformi e indipendenti.

Quindi un generatore di numeri pseudocasuali produce valori che appaiono soddisfare le condizioni (i) e (ii), cioè, per  $K$  grande, la frazione di valori che cadono in un qualsiasi sottointervallo di  $[0, 1]$  vale approssimativamente quanto la lunghezza del sottointervallo stesso (uniformità), e non ci sono pattern riconoscibili tra i valori (indipendenza).

Un generatore di numeri pseudocasuali ha la seguente forma generale:

$$x_{i+1} = f(x_i), \quad u_{i+1} = g(x_{i+1}) \quad (3.1)$$

per qualche funzione deterministica  $f$  e  $g$ , dato un valore iniziale  $x_0$  detto *seme*.

Poichè un computer può rappresentare solo un numero finito di valori <sup>1</sup>, ogni formula ricorsiva della forma (3.1) finirà per ripetere un valore  $x_i$  ad un certo punto, e ripeterà quindi ogni altro valore successivo a  $x_i$ . Un buon generatore dovrebbe però essere in grado di generare milioni di valori distinti prima di ripetersi.

Dobbiamo dunque tenere a mente le seguenti considerazioni generali nella costruzione di un generatore di numeri pseudocasuali:

- *Lunghezza del periodo.* Come abbiamo visto, ogni generatore della forma (3.1) si ripeterà ad un certo punto. In generale, si preferiscono generatori con periodi molto lunghi.
- *Riproducibilità.* È possibile costruire dispositivi che generino sequenze di numeri veramente casuali. Tuttavia lo svantaggio di queste sequenze è che non sono facilmente riproducibili. Spesso infatti è necessario poter condurre una nuova simulazione usando esattamente gli stessi input della precedente, cosa facilmente ottenibile con un generatore di numeri pseudocasuali.
- *Velocità.* In una simulazione un generatore viene fatto girare migliaia o perfino milioni di volte, pertanto una caratteristica fondamentale è la velocità.
- *Portabilità.* Un algoritmo per la generazione di numeri pseudocasuali deve produrre la stessa successione di valori su tutte le piattaforme. A volte però questa condizione viene trascurata per garantire una maggiore velocità e un periodo più lungo sfruttando caratteristiche specifiche di una particolare macchina.
- *Casualità.* Questa è la condizione più importante eppure la più difficile da definire. Vi sono delle proprietà teoriche che un buon generatore deve soddisfare per garantire apparente casualità nei valori che genera; generatori con buone proprietà teoriche sono poi sottoposti a test statistici per verificare l'effettiva casualità del risultato. Al giorno d'oggi il settore è sufficientemente sviluppato da fornire numerosi generatori con buone proprietà teoriche che hanno passato rigorosi test statistici<sup>2</sup>.

### 3.1 Linear Congruential Generator

Un **LCG** (*linear congruential generator*) è un generatore dato dalla formula ricorsiva

$$x_{i+1} = ax_i \pmod{m} \quad (3.2)$$

$$u_{i+1} = \frac{x_{i+1}}{m} \quad (3.3)$$

dove  $a, m \in \mathbb{N}$  e il valore iniziale  $x_0$  è un intero tra 1 e  $m - 1$  fornito dall'utente. L'operazione  $y \pmod{m}$  restituisce il resto della divisione intera di  $y$  per  $m$ , cioè

$$y \pmod{m} = y - [y/m]m \quad (3.4)$$

<sup>1</sup>si veda ad esempio [9]

<sup>2</sup>per qualche esempio di test statistico, come il *Next-Bit Test*, si veda [8]

Tabella 3.1: parametri ottimali per **LCG**

Modulo $m$	Moltiplicatore $a$
$2^{31} - 1$	16807
	39373
	742938285
	950706376
	1226874159
2147483399	40692
2147483563	40014

dove  $[x]$  è il più grande intero minore di  $x$ . Pertanto  $x_i$  è compreso tra 0 e  $m - 1$  per ogni  $i$  e quindi gli  $u_i$  giacciono nell'intervallo unitario. La semplicità e efficienza di **LCG** lo rendono uno dei generatori di numeri pseudocasuali più usati.

Un **LCG** che produce tutti gli  $m - 1$  valori distinti possibili prima di ripetersi ha un *periodo completo*. Purtroppo, perchè ciò avvenga non è sufficiente scegliere un  $m$  molto grande. Esistono però delle semplici condizioni che garantiscono che il generatore abbia un periodo completo: se  $m$  è primo, allora si ha un periodo completo per ogni  $x_0 \neq 0$  se

$$a^{m-1} - 1 \text{ è un multiplo di } m$$

$$a^j - 1 \text{ non è un multiplo di } m \text{ per } j = 1, \dots, m - 2$$

Un numero  $a$  che soddisfa tali proprietà è detto una *radice primitiva* di  $m$ . Osserviamo che la sequenza di numeri generati è

$$x_0, ax_0, a^2x_0, a^3x_0, \dots \pmod{m}.$$

Dunque la successione ritorna al valore  $x_0$  non appena arriva al più piccolo  $k$  per cui  $a^k x_0 \pmod{m} = x_0$ , cioè al più piccolo  $k$  per cui  $a^k \pmod{m} = 1$ , cioè il più piccolo  $k$  per cui  $a^k - 1$  è un multiplo di  $m$ , e quindi si ha un periodo completo esattamente quando  $a$  è una radice primitiva di  $m$ .

La tabella 3.1 mostra moduli e moltiplicatori per sette **LCG** che sono stati testati e raccomandati in letteratura. Il modulo è sempre un intero non più grande di  $2^{31} - 1$ , che è il più grande intero rappresentabile in precisione singola (32 bit), ed è anche un primo — un primo di Mersenne. Ciascun moltiplicatore è una radice primitiva del corrispondente modulo, quindi ogni generatore ha periodo completo.

## 3.2 Combinazioni di Generatori

Metodi che combinano vari **LCG** si presentano particolarmente vantaggiosi perchè preservano le caratteristiche computazionali dei **LCG** come la velocità data dalla loro particolare semplicità, e al contempo permettono di estenderne il periodo.

Un modo elementare di combinare più **LCG** è attraverso l'addizione. Consideriamo  $J$  generatori, il  $j$ -esimo avente parametri  $a_j, m_j$ :

$$x_{j,i+1} = a_j x_{j,i} \pmod{m_j}, \quad u_{j,i+1} = \frac{x_{j,i+1}}{m_j}, \quad j = 1, \dots, J$$

La combinazione di Wichmann-Hill pone

$$u_{i+1} = u_{1,i+1} + \cdots + u_{J,i+1} - [u_{1,i+1} + \cdots + u_{J,i+1}] \quad (3.5)$$

ossia  $u_{i+1}$  è la parte frazionaria di  $u_{1,i+1} + \cdots + u_{J,i+1}$ . La combinazione di L'Ecuyer pone invece

$$x_{i+1} = \sum_{j=1}^J (-1)^{j-1} x_{j,i+1} \pmod{m_1 - 1} \quad (3.6)$$

$$u_{i+1} = \begin{cases} \frac{x_{i+1}}{m_1} & x_{i+1} > 0 \\ \frac{m_1-1}{m_1} & x_{i+1} = 0 \end{cases} \quad (3.7)$$

dove si assume che  $m_1$  è il più grande tra gli  $m_j$ .

Una combinazione di generatori può avere un periodo molto più lungo del periodo di ciascun suo componente. Un periodo molto lungo si potrebbe ottenere con un normale **LCG** utilizzando un modulo molto grande, ma questo potrebbe portare a problemi di *overflow* nel calcolatore.

Un altro modo di combinare più generatori lineari è quello di combinare i moltiplicatori in una formula ricorsiva del tipo

$$x_i = (a_1 x_{i-1} + a_2 x_{i-2} + \cdots + a_k x_{i-k}) \pmod{m} \quad (3.8)$$

seguita da  $u_i = \frac{x_i}{m}$ , con  $x_i$  interi compresi tra 0 e  $m-1$ . Questo tipo di generatore è detto *multiple recursive generator*, o **MRG**. Un seme per questo generatore consiste in un vettore con  $k$  componenti  $(x_{k-1}, \dots, x_0)$ . Poichè nella (3.8) ogni valore  $x_i$  è generato a partire dal vettore  $(x_{i-1}, \dots, x_{i-k})$ , e poichè questo vettore può assumere  $m^k$  valori distinti, la sequenza degli  $x_i$  si ripeterà una volta che tale vettore ritorna ad un valore già visitato. Quindi, il periodo più lungo possibile per un **MRG** è di  $m^k - 1$  valori.

### 3.3 Metodi di Campionamento

Assumiamo d'ora in avanti di avere a disposizione una successione di v.a.  $U_1, U_2, \dots$  indipendenti e uniformemente distribuite su  $[0, 1]$ , cioè con CDF

$$P(U_i \leq u) = \begin{cases} 0 & u < 0 \\ u & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases} \quad (3.9)$$

Quello che vogliamo fare ora è ottenere delle v.a. con determinate CDF a partire dalle  $U_1, U_2, \dots$ . Il seguente lemma è molto utile in questo senso.

**Lemma 3.3.1.** *Siano*

$$X: \Omega \rightarrow I \text{ e } f: I \rightarrow J$$

*una v.a. sullo spazio di probabilità  $(\Omega, \mathcal{F}, P)$  a valori nell'intervallo reale  $I$  e una funzione continua e monotona strettamente crescente a valori nell'intervallo reale  $J$ . Allora la CDF della v.a.  $Y := f(X)$  è*

$$F_Y = F_X \circ f^{-1} \quad (3.10)$$

dove  $F_X$  indica la CDF di  $X$ .

*Dimostrazione.* La (3.10) segue da

$$P(Y \leq y) = P(f(X) \leq y) = P(X \leq f^{-1}(y)) = F_X(f^{-1}(y)), \quad y \in J,$$

dove nella seconda uguaglianza abbiamo usato il fatto che  $f$  è monotona strettamente crescente.  $\square$

### 3.3.1 Metodo della Trasformazione Inversa

Supponiamo ora di voler campionare da una v.a. con CDF  $F$ , cioè generare una v.a.  $X$  con la proprietà che  $P(X \leq x) = F(x)$  per ogni  $x$  (più generalmente con campionare si intende generare una immagine di tale variabile aleatoria). Nel *metodo della trasformazione inversa* si pone

$$X = F^{-1}(U), \quad U \sim \text{Unif}_{[0,1]}, \quad (3.11)$$

dove  $F^{-1}$  è l'inversa di  $F$ . Una delle proprietà delle CDF è quella di essere monotona debolmente crescente. Se  $F$  è strettamente crescente allora  $F^{-1}$  è ben definita, e per il lemma 3.3.1  $X$  ha CDF  $F$ . Nel caso invece in cui non sia strettamente crescente, abbiamo bisogno di definire  $F^{-1}$  laddove la controimmagine non è un singolo. Potremmo ad esempio porre

$$F^{-1}(u) = \inf_x \{F(x) = u\} \quad (3.12)$$

Verifichiamo ora che il metodo della trasformazione inversa genera effettivamente una v.a. con CDF  $F$ . Vediamo che

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(U) \leq x) = \\ &= P(U \leq F(x)) = \\ &= F(x). \end{aligned}$$

dove la seconda uguaglianza segue dal fatto che, per come abbiamo definito  $F^{-1}$ , gli eventi  $\{F^{-1}(u) \leq x\}$  e  $\{u \leq F(x)\}$  coincidono per ogni  $u$  e  $x$ , mentre l'ultima uguaglianza deriva da (3.9).

**Esempio 3.1. Distribuzione esponenziale.** Consideriamo la distribuzione esponenziale con media  $\frac{1}{\lambda}$ :

$$\text{Exp}_\lambda(H) = \lambda \int_H e^{-\lambda x} \chi_{[0,\infty)}(x) dx$$

Vediamo che, per  $x \geq 0$

$$\begin{aligned} \text{Exp}_\lambda([-\infty, x]) &= \lambda \int_0^x e^{-\lambda t} dt = \\ &= 1 - e^{-\lambda x} \end{aligned}$$

e quindi ha per CDF

$$F(x) = (1 - e^{-\lambda x}) \chi_{[0,\infty)}(x).$$

Data dunque una v.a. con distribuzione uniforme su  $[0, 1]$ , il metodo della trasformazione inversa dà

$$X = -\frac{1}{\lambda} \log(1 - U) \quad (3.13)$$

È possibile applicare il metodo della trasformazione inversa anche quando l'inversa della CDF non è nota, approssimando  $F^{-1}$  tramite metodi numerici. Calcolare  $F^{-1}(u)$  è equivalente a trovare la radice  $x$  dell'equazione  $F(x) - u = 0$ . Per una CDF  $F$  assolutamente continua, la cui distribuzione  $\mu$  ha densità  $f$ , allora  $F' = f$  e quindi il metodo di Newton per la ricerca delle radici<sup>3</sup> ha la forma

$$x_{n+1} = x_n - \frac{F(x_n) - u}{f(x_n)}, \quad (3.14)$$

dato un valore iniziale  $x_0$ .

**Esempio 3.2. Distribuzioni discrete.** Consideriamo una v.a. discreta con valori  $c_1 < \dots < c_n$ . Sia  $p_i$  la probabilità corrispondente al punto  $c_i, i = 1, \dots, n$ , e poniamo

$$q_0 = 0, \quad q_i = \sum_{j=1}^i p_j \quad i = 1, \dots, n.$$

Questi sono i valori che assume la CDF  $F$  della v.a. :

$$F(x) = \begin{cases} q_0 & \text{per } x < c_1 \\ q_j & \text{per } c_j \leq x < c_{j+1} \end{cases}$$

Allora con il metodo della trasformazione inversa, per campionare da questa distribuzione possiamo usare il seguente algoritmo:

1. campionare un valore  $U$  uniformemente distribuito su  $[0, 1]$ ;
2. trovare  $K \in \{1, \dots, n\}$  tale che  $q_{K-1} < U \leq q_K$ ;
3. porre  $X = c_K$ .

Il secondo passaggio può essere implementato con una ricerca binaria.

**Esempio 3.3. Distribuzione condizionata.** Supponiamo che  $X$  abbia distribuzione  $F$ , e consideriamo il problema di campionare  $X$  con la condizione che  $a < X \leq b$ , con  $F(a) < F(b)$ . Se  $U \sim \text{Unif}_{[0,1]}$ , allora la v.a.  $V$  definita da

$$V = F(a) + [F(b) - F(a)]U$$

è uniformemente distribuita tra  $F(a)$  e  $F(b)$ , e  $F^{-1}(V)$  ha la distribuzione desiderata. Infatti

$$\begin{aligned} P(F^{-1}(V) \leq x) &= P(F(a) + [F(b) - F(a)]U \leq F(x)) = \\ &= P(U \leq [F(x) - F(a)]/[F(b) - F(a)]) = \\ &= \frac{F(x) - F(a)}{F(b) - F(a)} \end{aligned}$$

---

<sup>3</sup>si veda [9] per maggiori dettagli

e del resto

$$\begin{aligned} P(X \leq x | X \in ]a, b]) &= \frac{P((X \leq x) \cap (X \in ]a, b]))}{P(X \in ]a, b])} = \\ &= \frac{P(X \in ]a, x])}{P(X \in ]a, b])} = \\ &= \frac{F(x) - F(a)}{F(b) - F(a)}. \end{aligned}$$

### 3.3.2 Metodo di Acceptance-Rejection

Affrontiamo ora un altro metodo di campionamento, introdotto per la prima volta da Von Neumann: il **Metodo di Acceptance-Rejection**. Supponiamo di voler generare campioni da una v.a. con densità  $f$  definita su un insieme  $H$ . Sia  $g$  una densità su  $H$  da cui sappiamo già generare campioni, e con la proprietà che

$$f(x) \leq cg(x), \quad x \in H$$

per qualche costante  $c$ . Nel metodo di acceptance-rejection, generiamo un campione  $X$  da  $g$  e generiamo  $U$  uniformemente distribuito su  $[0, 1]$ . Se  $U \leq \frac{f(X)}{cg(X)}$  allora *accettiamo*  $X$  come campione di  $f$ . Se  $X$  è rifiutato, campioniamo un nuovo candidato da  $g$  e il test viene ripetuto finché non viene accettato un valore.

Verifichiamo la validità di questo metodo. Sia  $Y$  il campione restituito dall'algoritmo; allora  $Y$  ha la distribuzione di  $X$  condizionata a  $U \leq \frac{f(X)}{cg(X)}$ . Pertanto, per ogni  $A \subset H$ ,

$$\begin{aligned} P(Y \in A) &= P(X \in A | U \leq f(X)/cg(X)) = \\ &= \frac{P((X \in A) \cap (U \leq f(X)/cg(X)))}{P(U \leq f(X)/cg(X))}. \end{aligned} \quad (3.15)$$

Il denominatore di (3.15) è

$$P\left(U \leq \frac{f(X)}{cg(X)}\right) = E\left[\chi_{(U \leq f(X)/cg(X))}\right] =$$

per la formula della probabilità totale per l'attesa condizionata

$$\begin{aligned} &= E\left[E\left[\chi_{(U \leq f(X)/cg(X))} | X\right]\right] = \\ &= E\left[P\left(U \leq \frac{f(X)}{cg(X)} | X\right)\right] = \end{aligned}$$

per l'uniformità di  $U$

$$\begin{aligned} &= E\left[\frac{f(X)}{cg(X)}\right] = \\ &= \int_H \frac{f(x)}{cg(x)} g(x) dx = \frac{1}{c} \end{aligned}$$

Pertanto

$$P(Y \in A) = cP((X \in A) \cap (U \leq f(X)/cg(X))) =$$

analogamente a sopra

$$= cE \left[ E \left[ \chi_{(U \leq f(X)/cg(X))} \chi_{(X \in A)} \mid X \right] \right] =$$

essendo  $\chi_{(X \in A)} \in m\sigma(X)$

$$\begin{aligned} &= cE \left[ \chi_{(X \in A)} P \left( U \leq \frac{f(X)}{cg(X)} \mid X \right) \right] = \\ &= c \int_A \frac{f(x)}{cg(x)} g(x) dx = \int_A f(x) dx \end{aligned}$$

e quindi  $Y$  ha densità  $f$ .

Osserviamo che la condizione di accettazione ha probabilità  $\frac{1}{c}$ . Poichè ogni tentativo è mutualmente indipendente, il numero di candidati generati finchè un campione è accettato è una v.a. con distribuzione geometrica e valore atteso  $c$ . È pertanto preferibile trovare un valore  $c$  che sia il più possibile vicino ad 1, così da velocizzare l'algoritmo. Ovviamente non potrà mai essere  $< 1$  in quanto in tal caso andrebbe a mancare la condizione  $f(x) \leq cg(x)$ . La velocità del metodo dipenderà dunque sia da  $c$  che dalla velocità nel campionare da  $g$ .

**Esempio 3.4. Distribuzione condizionata.** Consideriamo il problema di generare una v.a.  $X$  con la condizione  $X \in A$  per qualche insieme  $A$ . Nel caso scalare è facilmente risolvibile con il metodo della trasformazione inversa se  $A$  è un intervallo. Nel caso generale diventa però difficile campionare direttamente dalla distribuzione condizionata. Tuttavia, se è possibile generare campioni non condizionati, si può sempre ricorrere alla seguente procedura:

1. genera  $X$  finchè  $X \in A$
2. restituisci  $X$

Potremmo vedere questo algoritmo come una forma di acceptance-rejection. Se  $f$  è la densità condizionata e  $g$  quella non condizionata, allora si ha

$$\begin{aligned} P(X \in H \mid X \in A) &= \frac{P((X \in H) \cap (X \in A))}{P(X \in A)} = \\ &= \int_H \frac{1}{P(X \in A)} \chi_A(x) g(x) dx = \\ &= \int_H f(x) dx, \end{aligned}$$

cioè

$$\frac{f(x)}{g(x)} = \begin{cases} \frac{1}{P(X \in A)} & x \in A \\ 0 & x \notin A \end{cases} \quad (3.16)$$

Pertanto  $c = \frac{1}{P(X \in A)}$  è il massimo del rapporto. Inoltre, poichè il rapporto  $\frac{f(x)}{cg(x)}$  è 0 o 1, la condizione  $X \in A$  è sufficiente come condizione di accettazione.



Sebbene abbiamo ristretto l'attenzione al campionamento da densità, il metodo di acceptance-rejection si può applicare anche quando  $f$  e  $g$  sono delle funzioni di distribuzione.

Il modo migliore per campionare da una determinata distribuzione dipenderà sempre da caratteristiche specifiche a quella distribuzione. Per questo il metodo di acceptance-rejection è spesso combinato con altri metodi di campionamento per sfruttare caratteristiche particolari delle distribuzioni analizzate.



# Capitolo 4

## Esempi e Applicazioni

### 4.1 Implementazione

Evitare l'overflow è uno dei punti cruciali nell'implementazione di un **LCG**, in quanto il prodotto  $ax_i$  può diventare molto grande a seconda del modulo e del moltiplicatore scelti; poichè a moduli maggiori corrispondono periodi più lunghi, il modulo  $m$  sarà quasi certamente grande. Se il prodotto  $ax_i$  può essere rappresentato esattamente per ogni  $x_i$  nella successione, non si ha overflow. Se però il moltiplicatore  $a$  è grande, nemmeno la precisione doppia potrebbe essere sufficiente per una rappresentazione esatta di ogni prodotto  $ax_i$ . In tal caso, il generatore può essere implementato utilizzando

$$ax_i \bmod m = (a_1(2^\alpha x_i \bmod m) + (a_2 x_i \bmod m)) \bmod m,$$

dove abbiamo usato la scomposizione  $a = 2^\alpha a_1 + a_2$  con  $a_1, a_2 < 2^\alpha$ .

Dunque una semplice implementazione in Octave di un **LCG** potrebbe essere:

```
function U = LCG(m, A, x0)
V(1) = x0;
U(1) = x0 / m;
for n = 2:m
    V(n) = mod(A(1)*mod(2^A(3) * V(n-1), m) + mod(A(2)*V(n-1), m), m);
    U(n) = V(n) / m;
endfor
endfunction
```

che restituisce un vettore orizzontale  $U$  contenente gli  $m$  valori generati, e che richiede come input il modulo  $m$ , un vettore riga  $A = [a_1, a_2, \text{alfa}]$  tale che  $a = a_1 \cdot 2^{\text{alfa}} + a_2$ , e il seme  $x_0$ . Ricordiamo che per avere un periodo completo,  $m$  dev'essere primo e  $a$  una sua radice primitiva.

A questo punto, un'implementazione del Metodo Monte Carlo in Octave per il calcolo del valore atteso di una v.a. reale  $E[f(X)]$  è:

```
function E = Montecarlo(fun, U)
E = (1/length(U))*sum(feval(fun, U));
endfunction
```

dove  $\mathbf{E}$  è il valore atteso restituito,  $\text{fun}$  è la funzione  $f$  e  $\mathbf{U}$  è il vettore orizzontale contenente i valori indipendenti e identicamente distribuiti.

Proviamo a calcolare

$$I = \int_{-10}^{10} e^{-x^2} dx$$

con questo script e a compararlo con il metodo di Cavalieri-Simpson<sup>1</sup>, con un numero di nodi pari a 1000000.

Utilizziamo come modulo  $m = 1115839$ , come moltiplicatore  $a = 7776 = 60 \cdot 2^7 + 96$  e come seme  $x_0 = 1$ . L'approssimazione può migliorare o peggiorare a seconda del moltiplicatore e seme scelti, e questo dipende in parte dalla prestazione del generatore con questi valori. Dopo aver generato il vettore  $\mathbf{U}$  di valori indipendenti e uniformemente distribuiti su  $[0, 1]$ , dobbiamo trasformarli in valori con distribuzione  $\text{Unif}_{[-10,10]}$ . Applicando il metodo della trasformazione inversa, si trova che i valori cercati sono

$$\tilde{U} = 20U - 10.$$

Vediamo che, posto  $f(x) = e^{-x^2}$ ,

$$E[f(\tilde{U})] = \int_{-10}^{10} \frac{1}{20} e^{-x^2} dx$$

e quindi

$$I = 20E[f(\tilde{U})],$$

quindi una volta trovato il valore di  $E[f(\tilde{U})]$  con **MCM**, dovremo moltiplicarlo per 20 per ottenere l'approssimazione di  $I$ .

Facendo correre il programma, si ottiene

$$I_M = 1,772453850904938, \quad I_{sc} = 1,772453850905053$$

dove  $I_M$  è l'approssimazione di  $I$  trovata con **MCM** e  $I_{sc}$  è l'approssimazione di  $I$  trovata con Cavalieri-Simpson.

## 4.2 Approssimazione di $\pi$

Consideriamo la funzione

$$H(x, y) = \begin{cases} 1 & \text{per } x^2 + y^2 \leq 1 \\ 0 & \text{altrimenti} \end{cases} \quad (4.1)$$

e poniamo  $\Omega = [-1, 1] \times [-1, 1]$ . Poichè

$$I = \int_{\Omega} H(x, y) dx dy = \pi$$

possiamo usare il metodo Monte Carlo per approssimare quell'integrale e dare quindi una stima di  $\pi$ . Condurre un esperimento Monte Carlo è l'equivalente

<sup>1</sup>si veda [9, p.305-306] per una sua implementazione in Matlab

matematico del metodo rudimentale del calcolo dell'area che consisteva nel versare una quantità di fagioli nel quadrato  $\Omega$ , contare quanti erano caduti all'interno del cerchio di raggio 1 inscritto, e calcolarne il rapporto rispetto al totale.

Date due v.a. indipendenti  $U_1, U_2 \sim \text{Unif}_{[-1,1]}$ , vale  $U = (U_1, U_2) \sim \text{Unif}_\Omega$ . Pertanto

$$E[H(U)] = \int_{\Omega} \frac{1}{4} H(x, y) dx dy$$

e quindi

$$\pi = 4E[H(U)].$$

Modifichiamo lo script della funzione `Montecarlo` visto sopra per accomodare funzioni a due variabili:

```
function E = Montecarlo2(fun, U)
[height, length] = size(U);
E = 0;
for n = 1:length
    E = E + feval(fun, U(1, n), U(2, n));
endfor
E = (1/length)*value;
endfunction
```

dove stavolta  $U$  è una matrice  $2 \times n$  dove ciascuna riga è generata da un **LCG** con medesimo modulo. In questo modo, se prima e seconda riga sono costituite da campionamenti indipendenti della distribuzione  $\text{Unif}_{[-1,1]}$ , allora ogni colonna rappresenta un campionamento della distribuzione  $\text{Unif}_\Omega$ , ogni colonna indipendente dalle altre. Generando la prima riga di  $U$  con  $m = 1115839, a = 7776 = 60 \cdot 2^7 + 96, x_0 = 345$  e la seconda riga con  $m = 1115839, a = 149381 = 291 \cdot 2^9 + 389, x_0 = 876$  e trasformandole nella distribuzione opportuna con

$$\tilde{U} = 2U - 1$$

otteniamo

$$I_M = 3,141592653589793$$

che, rispetto al valore  $\pi$  di Octave, ha errore assoluto e relativo

$$\text{Err}_{\text{ass}} = |I_M - \pi| = 2,011 \cdot 10^{-4},$$

$$\text{Err}_{\text{rel}} = |I_M - \pi|/\pi = 6,401 \cdot 10^{-5}.$$

## 4.3 Un esempio di Business Planning

Un'azienda vuole lanciare un nuovo prodotto sul mercato, ma ha prima bisogno di una stima del profitto netto delle vendite nel primo anno del lancio. Possiamo supporre che il profitto netto  $NP$  dipenda da

1. volume di vendite  $SV$ ;
2. prezzo di vendita per unità  $SP$ ;

3. costo di produzione per unità UC;
4. costi fissi FC.

Il profitto netto può essere calcolato come

$$NP = SV(SP - UC) - FC \quad (4.2)$$

Supponiamo che, dopo aver condotto una ricerca di mercato, si trova che ci sono tre scenari di vendita equiprobabili:

- Vengono vendute 50.000 unità al prezzo di 11,00€ l'una;
- Vengono vendute 75.000 unità al prezzo di 10,00€ l'una;
- Vengono vendute 100.000 unità al prezzo di 8,00€ l'una.

Si trova anche che il costo per unità può trovarsi tra i 5,50€ e i 7,50€, ma che con grande probabilità si aggira sui 6,50€.

Vediamo allora che abbiamo quattro quantità aleatorie, che possono essere modellizzate da v.a.: NP, SV, SP e UC. SV e SP hanno distribuzione uniforme discreta e sono correlate. Possiamo invece supporre che UC abbia distribuzione triangolare. Una distribuzione triangolare di estremi  $a, b, a < b$  e centro  $c, a < c < b$  ha funzione di densità

$$\gamma(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & c < x \leq b \end{cases}$$

e CDF

$$F(x) = \begin{cases} 0 & x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & c < x \leq b \\ 1 & b \leq x \end{cases}$$

Questa distribuzione ha la caratteristica di addensare la probabilità intorno a  $c$ , e quindi è adatta a rappresentare UC, posto  $a = 5,50€$ ,  $b = 7,50€$  e  $c = 6,50€$ .

La procedura per calcolare la stima di NP è la seguente: generare tre vettori di uguale lunghezza contenenti campionamenti di SV, SP e UC rispettivamente, applicare la formula (4.2) componente per componente, e infine applicare il Metodo Monte Carlo con  $f = id$  al vettore ottenuto per determinare una stima di NP. Di seguito uno script in Octave che fa quanto appena descritto.

```
prime = 1115839; #il modulo per il generatore
primroot1 = 7776; A1 = [60, 96, 7];
primroot2 = 6; A2 = [1, 2, 2];

#generiamo due vettori con semi e radici primitive
#diverse per migliorare l'indipendenza.
V1 = LCG(prime, A1, 1);
V2 = LCG(prime, A2, 3);
```

```

#Generiamo, a partire da V1, i campionamenti di SV e SP.
for n = 1:prime
  if V1(n) <= 1/3
    SV(n) = 50000;
    SP(n) = 11;
  elseif 1/3 < V1(n) && V1(n) <= 2/3
    SV(n) = 75000;
    SP(n) = 10;
  else
    SV(n) = 100000;
    SP(n) = 8;
  endif
endfor

#Ora generiamo, a partire da V2, i campionamenti di UC.
a = 5.5; b = 7.5; c = 6.5;
A = sqrt((b-a)*(c-a)); B = sqrt((b-a)*(b-c)); C = (c-a)/(b-a);
for n = 1:prime
  if V2(n) <= C
    UC(n) = a + A*sqrt(V2(n));
  else
    UC(n) = b - B*sqrt(1-V2(n));
  endif
endfor

#Otteniamo i campionamenti di NP.
NP = SV .* (SP - UC) - 120000;

#Applichiamo Monte Carlo.
E = Montecarlo('id', NP)

```

dove `id` è la funzione identità definibile a parte in Octave.

Il valore di `E` che questo script restituisce è

$$E = 92506,31450076785.$$





# Bibliografia

- [1] Rick Durrett. *Probability: Theory and Examples*. 2019.
- [2] FrontlineSolvers, cur. *A Business Planning Example using Monte Carlo Simulation*. 2020. URL: <https://www.solver.com/monte-carlo-simulation-example>.
- [3] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. New York: Springer, 2003.
- [4] Dirk P. Kroese et al. *Why the Monte Carlo method is so important today*. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/wics.1314>.
- [5] Michael Mascagni. *Monte Carlo Methods: Early History and the Basics*. URL: [www.cs.fsu.edu/~mascagni/MC\\_Basics.pdf](http://www.cs.fsu.edu/~mascagni/MC_Basics.pdf).
- [6] N. Metropolis. «The Beginning of the Monte Carlo Method». In: *Los Alamos Science*. Special Issue 15 (1987). URL: <https://library.lanl.gov/cgi-bin/getfile?00326866.pdf>.
- [7] Andrea Pascucci. *Teoria della Probabilità*. 2018.
- [8] Rafael Pass e Abhi Shelat. *A Course in Cryptography*. 2010.
- [9] Alfio Quarteroni et al. *Matematica Numerica*. Vol. 77. UNITEXT-La Matematica per il 3+2. Springer, 2014.
- [10] Anthony D. Rollett e Priya Manoliar. *The Monte Carlo Method*. URL: [www.semanticscholar.org](http://www.semanticscholar.org).
- [11] Wikipedia, cur. *Monte Carlo integration*. 2020. URL: [https://en.wikipedia.org/wiki/Monte\\_Carlo\\_integration](https://en.wikipedia.org/wiki/Monte_Carlo_integration).