Metodi Monte Carlo per la generazione di numeri casuali

Albi Kerbizi

October 25, 2024

Abstract

Nota dedicata ai metodi Monte Carlo per la generazione di numeri casuali nell'ambito del corso di Laboratorio II, parte *Analisi statistica dei dati sperimentali*.

1 Introduzione

Con *metodi Monte Carlo* (MC) si intende generalmente l'insieme di metodi che sfruttano la generazione di variabili aleatorie artificiali per risolvere problemi matematici complicati. Le origini dei metodi MC risalgono già alla seconda metà del 1700, ma sono utilizzati sistematicamente dopo le prime applicazioni, condotte nell'ambito del progetto Manhattan, sui processi di diffusione e assorbimento dei neutroni nei materiali fissili[1]. Ad oggi, i metodi MC sono uno strumento indispensabile in tutti i campi della ricerca scientifica e molto diffusi grazie alla disponibilità di calcolatori e alla facilità con cui si possono generare numeri casuali.

Tra le applicazioni più comuni dei metodi MC in fisica ci sono le simulazioni degli urti ad alta energia tra particelle[2] e lo studio delle incertezze dei parametri stimati a partire dai dati sperimentali. Per questo corso, invece, i metodi MC sono uno strumento molto utile per verificare le leggi della statistica (ad esempio, il teorema del limite centrale), ottenere le distribuzioni di funzioni arbitrarie di variabili casuali e per verificare la correttezza dei metodi per le stime dei parametri.

Tutti i metodi MC si basano essenzialmente sulla generazione di numeri casuali con distribuzione uniforme nell'intervallo $[0, 1]$. Nel linguaggio `Fortran`, la subroutine che permette di generare numeri casuali con distribuzione uniforme si chiama

[1] Il nome stesso *Monte Carlo*, preso dal celebre casinò del Principato di Monaco, veniva usato da John von Neuman e Stanislav Ulam per riferirsi alle loro ricerche segrete sui materiali fissili.

[2] In gergo, i programmi relativi si chiamano Generatori di Eventi.

`random_number`. Il commando per generare un numero casuale è `call random_number(r)`, dove `r` è una variable di tipo `real` il cui valore è il numero casuale generato.

Pù in generale, è possibile generare un numero casuale r' con distribuzione uniforme nell'intervallo [a,b], dove a e b sono due numeri reali con $a < b$, tramite l'equazione

$$r' = a + r\,(b - a), \tag{1}$$

dove r ha distribuzione uniforme in [0,1].

Nelle sezioni che seguono, verranno presentati i metodi MC più comuni per generare una variabile casuale x secondo una generica funzione di distribuzione $f(x)$, e altri metodi utili.

2 Generazione di variabili casuali continue

2.1 Metodo accept-reject

Il metodo più generale per generare una variabile casuale continua x che assume valori nell'intervallo $[a, b]$ con funzione di distribuzione $f(x)$, è il metodo dell'accept-reject, o metodo del rigetto. Esistono diverse varianti del metodo, e la più semplice è mostratra in Fig. 1. Essa consiste nel racchiudere il grafico di $f(x)$ in un rettangolo avente come lati gli intervalli $[a, b]$ e $[0, f_{max}]$, dove f_{max} è il valore massimo che $f(x)$ assume in $[a, b]$.

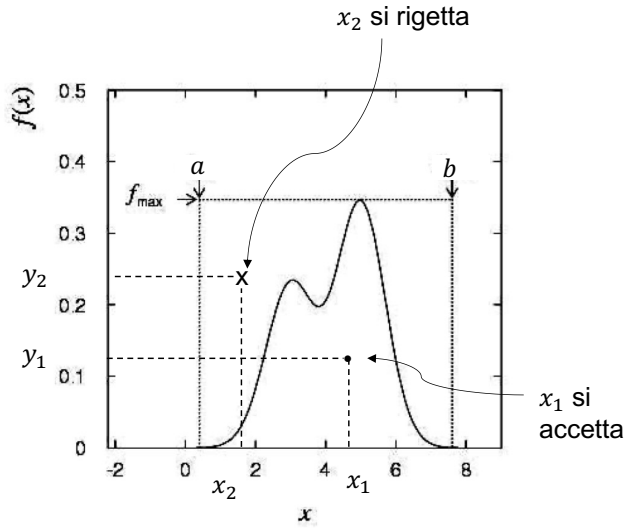

Figure 1: Illustrazione del metodo accept-reject. Il punto x_1 si accetta mentre x_2 si rigetta.

Una volta costruito il rettangolo, si applica la seguente ricetta

1. Cominciare un ciclo infinito, e all'interno del ciclo
   a. Generare x con distribuzione uniforme in $[a, b]$.
   b. Generare y con distribuzione uniforme in $[0, f_{max}]$.
   c. Calcolare il valore $f(x)$.
2. Se $y < f(x)$, accettare x [3] e uscire dal ciclo. Altrimenti rigettare x e ricominciare da 1.

Per generare N_{acc} variabili casuali $x_1, x_2, \ldots, x_n$, bisogna ripetere N_{acc} volte i punti 1-2. Per verificare che i valori generati seguono la distribuzione $f(x)$, si possono istogrammare i valori generati e confrontare l'istogramma con quanto ci si aspetta dalla funzione di distribuzione $f(x)$.

I punti 1-2 devono essere ripetuti più di una volta per generare un solo valore di x. Pertanto in generale è necessario generare N_{gen} cicli per ottenere N_{acc} valori di x. Questo introduce una ineficienza nel metodo, quantificabile come il rapporto tra il numero di coppie (x, y) generate il cui valore di x è stato accettato e il numero di coppie (x, y) generate. In formule, l'efficienza è data da

$$\eta = \frac{N_{acc}}{N_{gen}}. \tag{2}$$

Il numero N_{acc} è anche dato dalla frazione di coppie (x, y) che stanno sotto il grafico della funzione $f(x)$, ossia

$$N_{acc} = N_{gen} \frac{\int_a^b dx\, f(x)}{(b-a)\, f_{max}}. \tag{3}$$

Inserendo questa espressione in Eq. (2) si ottiene

$$\eta = \frac{1}{(b-a) \times f_{max}} \int_a^b dx\, f(x), \tag{4}$$

ovvero l'efficienza è data dal rapporto tra l'area sotto il grafico di $f(x)$ nell'intervallo $[a, b]$ e l'area del rettangolo in cui è inscritto il grafico di $f(x)$.

Esempio. Consideriamo la funzione di distribuzione

$$f(x) = \frac{1}{2\pi}(1 + a\, \cos x), \tag{5}$$

[3] Nel senso che la variabile x viene salvata, per esempio, in un array.

con $|a| \leq 1$. Il valore massimo che f assume in $[0, 2\pi]$ è $f_{max} = \frac{1}{2\pi}(1 + |a|)$. Quindi il rettangolo che inscrive $f(x)$ è dato da $[0, 2\pi] \times [0, \frac{1}{2\pi}(1 + |a|)]$.

Il grafico di $f(x)$ per $a = 0.2$ e il rettangolo che la inscrive sono mostrati in Fig. 2 (sinistra). Usando il metodo accept-reject sono state generate 585 coppie di punti (x, y). Le coppie la cui x è stata accettata sono 400 e sono mostrate con i punti verdi, mentre le coppie rigettate con i punti rossi. In Fig. 2 (destra) è invece mostrato l'istogramma di tutti i valori di x accettati, confrontato con la funzione di distribuzione aspettata.

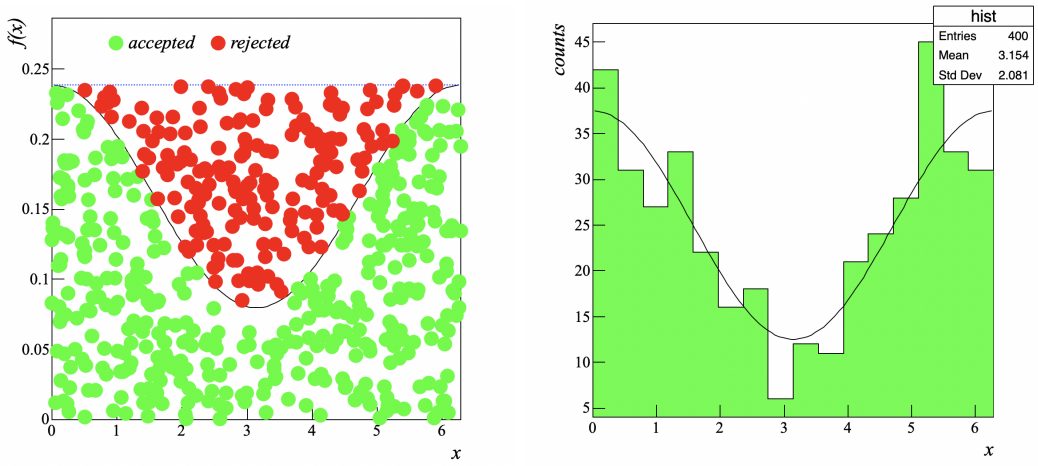

Figure 2: Grafico a sinistra: applicazione del metodo accept-reject alla funzione $f(x) = (1 + 0.5\cos(x))/(2\pi)$. I valori accettati sono i punti verdi, quelli rigettati i punti rossi. Grafico a destra: istogramma dei valori accettati e confronto con i conteggi aspettati (linea continua).

Utilizzando l'Eq. 4, l'efficienza del metodo accept-reject applicata alla funzione di distribuzione in Eq. (5) è $\eta = 1/(1 + a)$. Nel caso specifico $a = 0.2$, l'efficienza è $5/6 \simeq 0.83$. L'efficienza cresce al decrescere di a, e nel caso limite in cui $a = 0$ l'efficienza vale 1 (il grafico di $f(x)$ e il rettangolo che lo inscrive coincidono).

Se la variabile casuale x assume valori in un intervallo $[a, b]$ esteso come per esempio $[0 : \infty)$ oppure $(-\infty, +\infty)$, la funzione di distribuzione $f(x)$ è caratterizzata da una o due code. La probabilità dell'occorrenza di x nelle code è piccola e quindi il metodo accept-reject richiede svariati cicli per generare i valori di x in queste regioni. Esempi tipici sono la funzione di distribuzione dei tempi d'attesa data da $f(x) = \exp(-x/\tau)/\tau$ e la distribuzione gaussiana. In questi casi si fa riferimento ad altri metodi, come per esempio il metodo della funzione di ripartizione, descritto nella sezione successiva, e il metodo del limite centrale, presente nelle note del corso.

2.2 Metodo della funzione di ripartizione

Data una variabile casuale x che assume valori nell'intervallo $[a, b]$ con funzione di distribuzione $f(x)$, definiamo la funzione di ripartizione $F(x) = \int_a^x dx' f(x')$. Il metodo della funzione di ripartizione sfrutta il fatto che $F(x)$ è una variabile casuale con distribuzione uniforme nell'intervallo $[0, 1]$[4].

Prima si genera un valore per la funzione di ripartizione $F(x)$, quindi un numero casuale r con distribuzione uniforme in $[0, 1]$. Si pone $F(x) = r$, da cui

$$x = F^{-1}(r). \tag{6}$$

Per ogni valore di r generato si ottiene quindi un valore di x valutando la funzione inversa F^{-1} di F in r. Il metodo assume chiaramente che la funzione di ripartizione sia invertibile analiticamente [5].

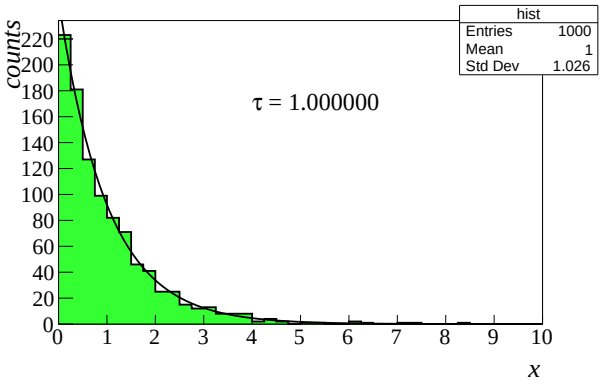

Figure 3: Esempio di generazione di 1000 numeri casuali con funzione di distribuzione esponenziale con $\tau = 1$, usando il metodo accept reject. La curva rapresenta il numero di conteggi atteso in ciascun intervallo dell'istogramma.

Il metodo della funzione di ripartizione si applica, per esempio, alla generazione di numeri casuali con distribuzione esponenziale $f(x) = \exp(-x/\tau)/\tau$. La funzione di ripartizione corrispondente è

$$F(x) = \int_0^x dx'/\tau \, \exp(-x'/\tau) = 1 - \exp(-x/\tau). \tag{7}$$

[4]Vedi note del corso.

[5]In alternativa l'Eq. (6) si può invertire numericamente, al costo di una complicazione algoritmica. Solitamente si cercano metodi alternativi più semplici.

Applicando l'Eq. (6) al caso in considerazione, si ottiene

$$x = -\tau \, \log(1 - r). \tag{8}$$

Generando N_{gen} valori di r si ottengono altrettanti valori di x, che hanno la corretta distribuzione esponenziale come si può facilmente osservare istogrammando i valori ottenuti. In Fig. 3 è mostrata un esempio di generazione di $N_{gen} = 1000$ variabili casuali con funzione di distribuzione esponenziale avente $\tau = 1$.

3 Calcolo del valore di un integrale

Il metodo Monte Carlo per il calcolo degli integrali definiti è noto come metodo *importance sampling*, e consiste in quanto segue.

Supponiamo di voler calcolare l'integrale di una funzione $g(x)$ nell'intervallo $[a, b]$. Introduciamo una funzione $f(x)$ positiva e normalizzata nell'intervallo $[a, b]$, quindi interpretabile come funzione di distribuzione della variabile x.

Possiamo allora scrivere

$$I = \int_a^b g(x)\, dx = \int_a^b \frac{g(x)}{f(x)}\, f(x)\, dx \equiv E\left[\frac{g(x)}{f(x)}\right]. \tag{9}$$

L'integrale I si può quindi interpretare come il valore di aspettazione di $g(x)/f(x)$ valutato con la funzione di distribuzione $f(x)$.

Generando N variabili casuali $x_1, \ldots, x_N$ secondo la funzione di distribuzione $f(x)$ [6] possiamo stimare l'integrale come

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \frac{g(x_i)}{f(x_i)}, \tag{10}$$

quindi mediante la media aritmentica dei valori di $g(x)/f(x)$.

La stima $\hat{I}$ dell'integrale I è essa stessa una variabile casuale, caratterizzata da una certa deviazione standard $\hat{\sigma}_{\hat{I}}$ che è l'incertezza assocciata ad $\hat{I}$. Per calcolare l'incertezza sulla stima dell'integrale bisogna prima valutare la varianza di $g(x)/f(x)$, ovvero

$$\hat{\sigma}^2_{g/f} = \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{g(x_i)}{f(x_i)} - \hat{I}\right]^2. \tag{11}$$

Da cui, utilizzando la legge di propagazione della varianza, si ottiene l'incertezza sulla stima di I

$$\hat{\sigma}^2_{\hat{I}} = \frac{\hat{\sigma}^2_{g/f}}{N}. \tag{12}$$

[6] Ad esempio utilizzando il metodo dell'accept-reject oppure della funzione di ripartizione.

Mettendo insieme Eq. (10) ed Eq. (12) si può esprimere il risultato finale come

$$\hat{I} \pm \frac{\sigma_{g/f}}{\sqrt{N}}. \tag{13}$$

Come si osserva l'incertezza associata al calcolo dell'integrale decresce come $1/\sqrt{N}$, una caratteristica comune dei metodi Monte Carlo. Incrementando il campione di variabili casuali simulate con $f(x)$, diminuisce l'incertezza associata ad $\hat{I}$.

Per diminuire $\hat{\sigma}_{\hat{I}}$ si può inoltre cercare di scegliere la funzione $f(x)$ in modo che $\sigma_{g/f}$ sia piccola. Per fare questo, è utile osservare che scegliendo $f(x) = c\,g(x)$, con c una costante, e utilizzando l'Eq. (11) si ottiene $\sigma^2_{g/f} = 0$. La varianza di $g(x)/f(x)$ si riduce in questo caso a zero. La condizione che $f(x)$ sia normalizzata impone che $c = 1/I$, pertanto la scelta $f(x) = c\,g(x)$ non è sensata. Essa tuttavia offre un criterio per ridurre $\sigma^2_{g/f}$, ovvero che la funzione $f(x)$ deve essere scelta in modo che il rapporto $g(x)/f(x)$ sia il più piatto possibile.

Vediamo ora un esempio pratico. Vogliamo stimare il valore dell'integrale

$$I = \int_{-1}^{1} \frac{\exp(-x^2/2)}{\sqrt{2\pi}}\,dx = 2\int_{0}^{1} \frac{\exp(-x^2/2)}{\sqrt{2\pi}}\,dx, \tag{14}$$

che è noto valere $I = 0.68268...$

La funzione $g(x)$ è data da $g(x) = 2\exp(-x^2/2)/(\sqrt{2\pi})$. Una scelta immediata per $f(x)$ è data dalla distribuzione uniforme $f(x) = 1$ con $x \in [0, 1]$.

Notiamo però che $\exp(-x^2/2)$ vale 1 per $x = 0$ e $\exp(-1/2)$ per $x = 1$. In questi estremi la funzione $g(x) = \exp(-x/2)$ assume gli stessi valori. Un'altra possibilità è quindi la funzione $f_{exp}(x) = \exp(-x/2)/[2\,(1 - e^{-1})]$, normalizzata nell'intervallo $[0, 1]$.

Generando $N = 100$ variabili casuali $x_1, x_2, \ldots$ distribuiti uniformemente in $[0, 1]$ si ottiene $\hat{I}_{uni.} = 0.682 \pm 0.010$. Generando invece $N = 100$ variabili casuali distribuiti secondo $f_{exp}(x)$ (utilizzando ad esempio il metodo della funzione di ripartizione in Sec. 2.2) si ottiene $\hat{I}_{exp} = 0.6855 \pm 0.0024$, caratterizzata da un'incertezza minore rispetto alla stima $\hat{I}_{uni.}$. Incrementando il numero di variabili casuali simulate a $N = 1000$ si ottengono le stime $\hat{I}_{uni.} = 0.6820 \pm 0.0010$ e $\hat{I}_{exp.} = 0.6828 \pm 0.0003$. Entrambe le stime sono vicine al valore noto $I = 0.68268..$ e compatibile con esso entro una deviazione standard, ma la stima ottenuta utilizzando $f_{exp.}(x)$ ha un'incertezza tre volte più piccola rispetto all'incertezza ottenuta assumendo $f(x)$ una distribuzione uniforme.

4 Calcolo di π

Consideriamo il quadrato $\Omega_Q = \{(x, y) \in \mathbb{R} \,|\, -1 \leq x \leq 1, -1 \leq y \leq 1\}$ nel piano (x, y), come mostrato in Fig. 4. All'interno del quadrato, consideriamo il cerchio

$\Omega_R = \{(x, y) \in \mathbb{R} | x^2 + y^2 \leq R^2\}$ di raggio $R = 1$ centrato nell'origine del sistema di assi.

Generiamo N coppie di variabili casuali indipendenti $(x_1, y_1), \ldots, (x_N, y_N)$ all'interno del quadrato. Una coppia (x_i, y_i) si può generare utilizzando l'Eq. (1) con $a = -1$ e $b = 1$, ovvero $x_i = -1 + 2\, r_i$ e $y_i - 1 + 2\, u_i$, con r_i e u_i variabili casuali con distribuzione uniforme. I punti $(x_1, y_1), \ldots, (x_N, y_N)$ sono pertanto distribuiti uniformemente nel quadrato, secondo la funzione di distribuzione $f(x, y) = 1/4$.

Indichiamo con M il numero di punti che cadono all'interno del cerchio. Ci aspettiamo che $M = N \times p$, dove p è la probabilità che un punto cada all'interno del cerchio. Quest'ultima è data da

$$p = \int_{\Omega_R} f(x, y)\, dx\, dy = \frac{1}{4} \int_0^R \rho\, d\rho \int_0^{2\pi} d\phi = \frac{1}{4}\, \frac{R^2}{2} \times 2\pi = \pi/4, \tag{15}$$

dove Ω_R indica la regione interna al cerchio, ovvero l'insieme dei punti $\Omega_R = \{(x, y) \in \mathbb{R} | x^2 + y^2 \leq R^2\}$. Nell'ultima uguaglianza è stato usato il valore del raggio $R = 1$.

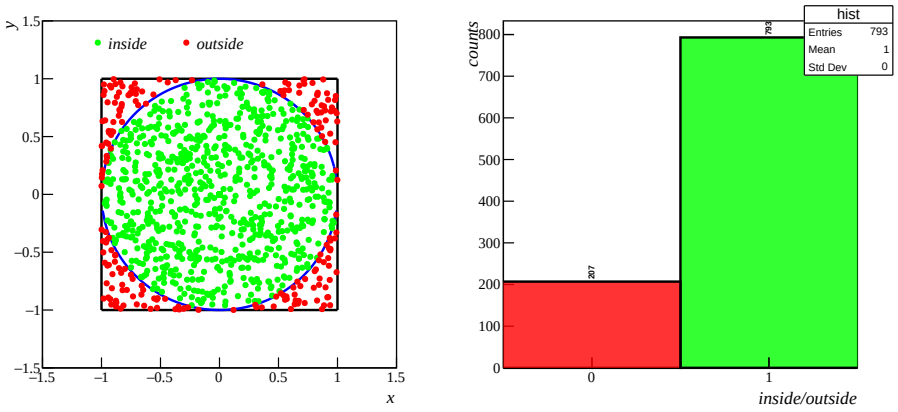

Figure 4: Sinistra: Esempio di calcolo di π tramite la generazione di $N = 1000$ variabili casuali uniformemente distribuiti all'interno di un quadrato. Destra: Conteggi dei punti che sono caduti all'interno del cerchio (verde).

Usando l'espressione per p e il valore atteso $M = N \times p$, si può calcolare π come

$$\hat{\pi} = 4 \times \frac{M}{N}. \tag{16}$$

Infine, considerando come "successo" un punto all'interno del cerchio la distribuzione del valore atteso M all'interno del cerchio è binomiale. L'incertezza associata al

valore atteso è pertanto $\sigma_M = \sqrt{N\,\hat{p}\,(1-\hat{p})}$, dove per la probabilità p può essere stimata come $\hat{p} = M/N$. L'incertezza associata alla stima $\hat{\pi}$ risulta per tanto

$$\sigma_{\hat{\pi}} = 4 \times \frac{\sigma_{\hat{M}}}{N}. \tag{17}$$

Nell'esempio in Fig. 4, sono stati generati $N = 1000$ punti nel quadrato. Di questi $M = 793$ punti (verde) sono caduti all'interno del cerchio. Utilizzando l'Eq. (16) e l'Eq. (17), si ottiene $\hat{\pi} = 3.172 \pm 0.051$, compatibile con $3.14\ldots$ entro una deviazione standard.

5 Simulazione della funzione di distribuzione di una combinazione di variabili casuali

I metodi Monte Carlo possono essere utilizzati per calcolare la distribuzione di combinazioni di variabili casuali. Un esempio è la distribuzione della variabile

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}, \tag{18}$$

con $x_1, \ldots, x_N$ un campione della variabile casuale x distribuita secondo la funzione di distribuzione $f(x)$. Quest'ultima è caratterizzata da un valore di aspettazione μ e da una deviazione standard σ. Il valore medio di x è dato dalla media aritmetica $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. Il teorema del limite centrale afferma che, se σ è finita, allora per $N \to \infty$ la variabile z segue una distribuzione normale standard. La deviazione standard associata a z sarà quindi $\sigma_z = 1$ e il significato probabilitisto di σ_z è chiaro, ad esempio $P(|z - 0| < \sigma_z) = 0.68268.\ldots.$

In pratica, N assume un valore finito e z non è detto segua una distribuzione normale standard. Di conseguenza il significato probabilistico di σ_z non è chiaro. Quello che si conosce è il valore di aspettazione $\mu_z = E[z] = 0$ e la deviazione standard $\sigma_z = 1$, come si può dedurre dall'Eq. (18).

Scelta la funzione di distribuzione $f(x)$ e fissato un certo N, la distribuzione di z si può calcolare tramite una simulazione Monte Carlo come segue:

1. Generare N variabili casuali $x_1, \ldots, x_N$ con una funzione di distribuzione $f(x)$,
2. Calcolare z come in Eq. (18),
3. Ripetere M volte i punti 1 e 2, e riempire un istogramma con i valori di $z_1, \ldots, z_M$.

Inoltre, avendo a disposizione il campione $z_1, \ldots, z_M$ si può calcolare la probabilità $P(|z - \mu_z| < \sigma_z) = N_{|z-\mu_z|<\sigma_z}/M$, dove $N_{|z-\mu_z|<\sigma_z}$ è il numero di variabili z che cadono nell'intervallo $|z - \mu_z| < \sigma_z$.

N	$f(x) = 1$	$f(x) = e^{-x}$
2	0.651	0.737
5	0.672	0.700
10	0.678	0.692
100	0.683	0.682
1000	0.682	0.685

Table 1: Valori di $P(|z - \mu_z| < \sigma_z)$ ottenuti mediante simulazioni Monte Carlo per diversi N.

In Fig. 5 sono mostrati i risultati di una simulazione Monte Carlo assumendo che la variabile x sia distribuita secondo una funzione di distribuzione uniforme $f(x) = 1$ in $[0, 1]$ (istogramma a sinistra) e secondo la funzione di distribuzione $f(x) = e^{-x}$ (istogramma a destra). I valori di N sono, dall'alto in basso, 2, 5, e 100. Per ogni istogramma sono stati generati $M = 50000$ eventi. Le distribuzioni sono confrontate con quanto atteso dalla distribuzione normale standard (punti). Inoltre l'area colorata in ogni istogramma mostra l'intervallo $\mu_z - \sigma_z < z < \mu_z + \sigma_z$.

Come si può vedere, la distribuzione della variabile z ottenuta come somma di di variabili casuali con funzione di distribuzione esponenziale è molto diversa dalla distribuzione normale standard per $N \lesssim 100$. Converge quindi lentamente alla distribuzione normale standard, contrariamente al caso $f(x) = 1$. I risultati delle probabilità $P(|z - \mu_z| < \sigma_z)$ ottenute per i diversi N considerati, sono riassunti in Tab. 1. I valori si avvicinano a 0.68268.. atteso nel caso di distribuzione normale standard per $N \gtrsim 100$ nel caso $f(x) = e^{-x}$ e per $N \gtrsim 10$ nel caso $f(x) = 1$.

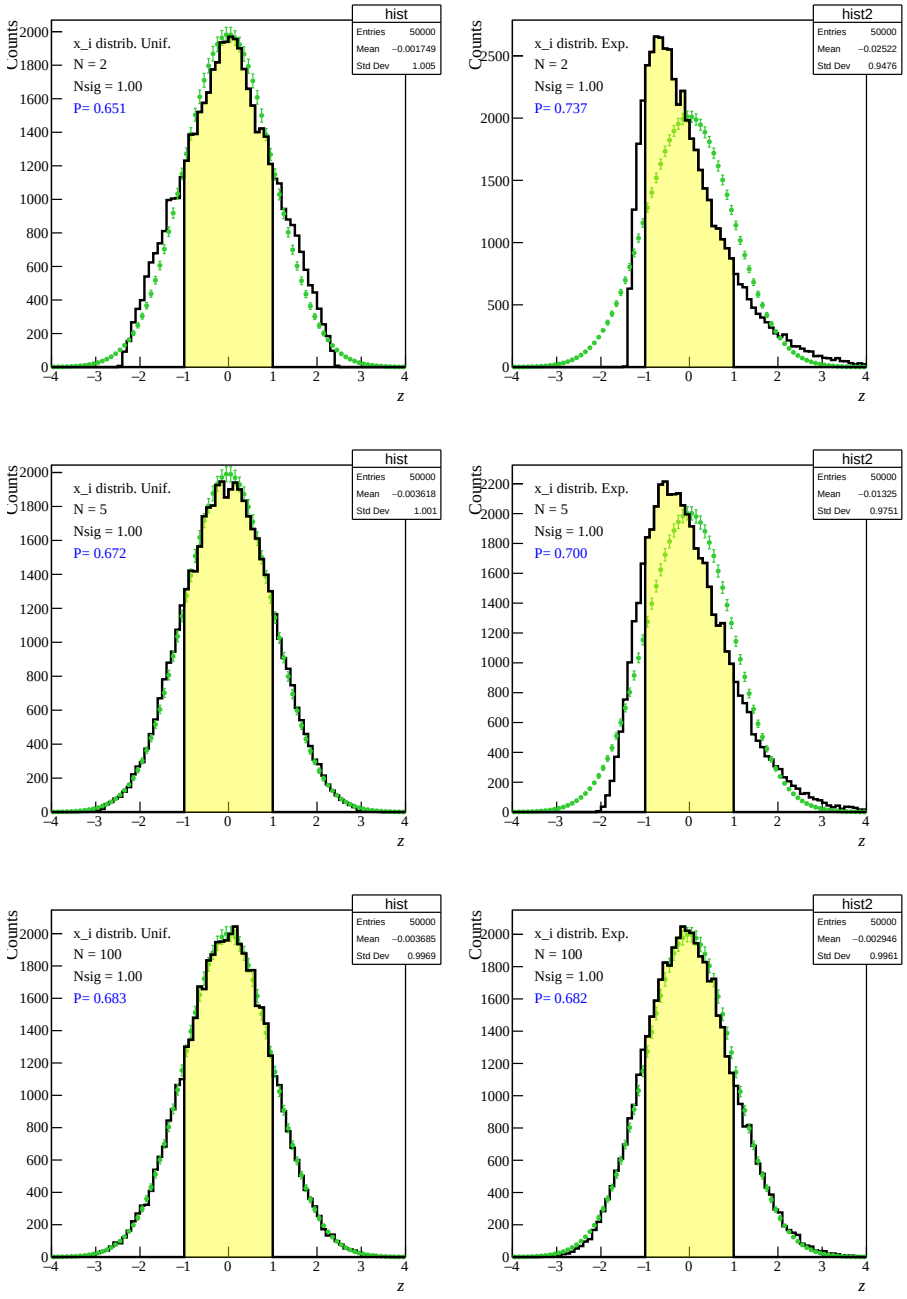

Figure 5: Distribuzioni della variabile z in Eq. (18) per diversi valori di N ottenute tramite simulazioni Monte Carlo, assumendo $f(x) = 1$ (a sinistra) e $f(x) = e^{-x}$ (a destra). L'area colorate indica l'intervallo $\mu_z - \sigma_z < z < \mu_z + \sigma_z$. I punti mostrano i conteggi attesi secondo la distribuzione normale standard.