Metodi Monte Carlo per la generazione di numeri casuali

Albi Kerbizi

October 27, 2023

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 `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 $f(x)$.

[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.

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 $[a : b]$.

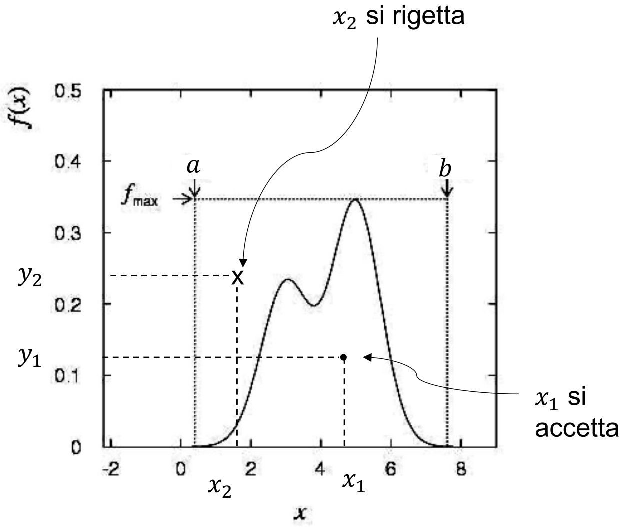

Figure 1: Illustrazione del metodo accept-reject. Il punto x_1 si accetta mentre x_2 si rigetta.

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[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

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

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

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

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

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 $f(x)$ e' 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) e' invece mostrato l'istogramma di tutti i valori di x accettati, confrontato con la funzione di distribuzione aspettata.

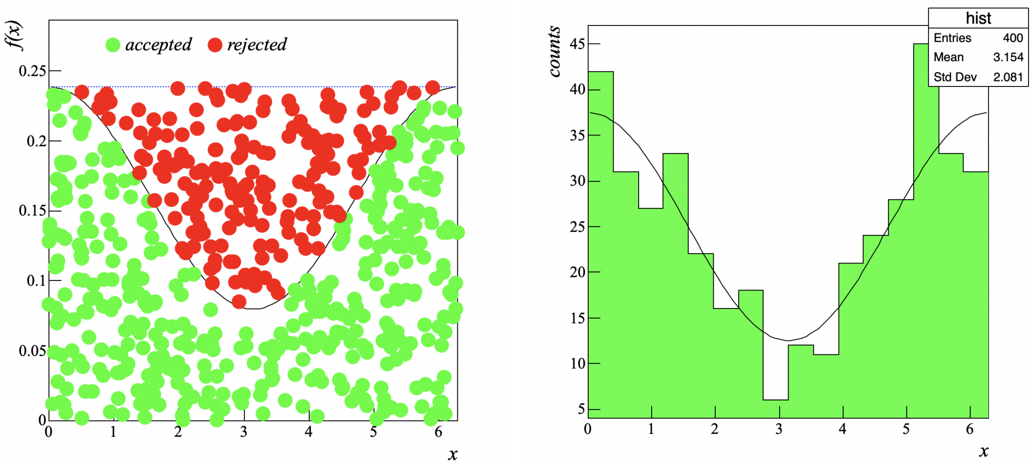

Figure 2: Grafico a sinistra: applicazione del metodo accept-reject alla funzione $f(x) = (1 + 0.5\, \cos(x))/(2\pi)$. I valori accettati sono i punti verdi, quelli rigettati i punti rossi. Grafico a destra: istogramma dei valori accettati e confronto con i conteggi aspettati (linea continua).

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].

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

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

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 e' mostrata un esempio di generazione di $N_{gen} = 1000$ variabili casuali con funzione di distribuzione esponenziale avente $\tau = 1$.

3 Generazione di variabili casuali discrete

3.1 Numeri casuali con distribuzione binomiale

La distribuzione binomiale descrive la propabilità che dati N campionamenti di una variabile casuale che assume valori binari (ad esempio la faccia di una moneta), k di questi siano favorevoli se la probabilità di un evento favorevole è p. Considerando il significato della distribuzione binomiale e fissato il valore di p, è possibile generare un valore della variabile k tramite la ricetta

1. Generare M variabili casuali $r_i, \ldots, r_M$ con distribuzione uniforme in $[0 : 1]$.

[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

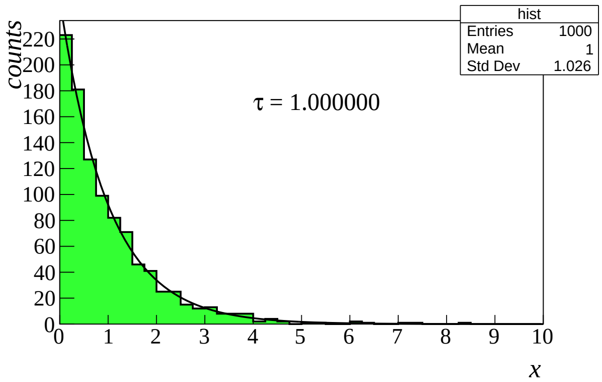

Figure 3: Esempio di generazione di 1000 numeri casuali con funzione di distribuzione esponenziale con $\tau = 1$, usando il metodo accept reject. La curva rapresenta il numero di conteggi atteso in ciascun intervallo dell'istogramma.

2. Contare quante volte la condizione $r_i \leq p$ è soddisfatta. Il numero che si ottiene è il valore di k.

Per generare N variabili $k_1, \ldots, k_N$ con distribuzione binomiale, bisogna ripetere N volte i punti 1-2. Complessivamente, quindi, bisogna generare $N \times M$ variabili casuali con distribuzione uniforme per ottenere N variabili con distribuzione binomiale.

Un esempio con $N = 6$ e $p = 0.4$ e' mostrato nel grafico a sinistra in Fig. 4. Le linee nera e rossa rappresentano il confronto con una distribuzione binomiale e di Poisson, rispettivamente.

3.2 Numeri casuali con distribuzione di Poisson

La distribuzione di Poisson si ottiene dalla distribuzione binomiale al limite $p \to 0$ e $N \to \infty$ con $N\,p = cost \equiv \nu$, dove ν è il valore di aspettazione della distribuzione di Poisson.

Fissato il valore di ν, la generazione di una variabile k con funzione di distribuzione di Poisson si ottiene applicando i punti 1 e 2 in Sez. 3.1, scegliendo p sufficientemente piccolo ed N in modo tale che $\nu = N\,p$.

Un esempio con $\nu = 4$ ottenuto scegliendo $N = 40$ e $p = 0.1$ e' mostrato nel grafico a destra in Fig. 4. Le linee nera e rossa rappresentano il confronto con una distribuzione binomiale e di Poisson, rispettivamente.

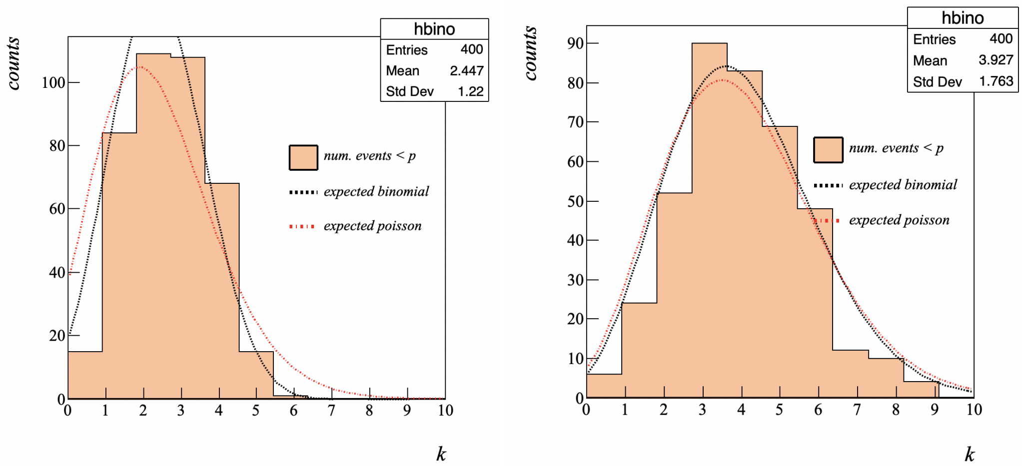

Figure 4: Grafico a sinistra: generazione di variabili casuali con funzione di distribuzione binomiale avente $p = 0.4$ e $N = 6$. Grafico a destra: generazione di variabili casuali con distribuzione di Poisson con $\nu = 4$ ($p = 0.1$ e $N = 40$). In entrambi grafici, i valori generati sono rappresentati dall'istogramma. Le linee tratteggiate nera e rossa rappresentano il confronto con la distribuzione aspettata, assumendo rispettivamente una distribuzione binomiale e di Poisson.