Simulazione dei dati

- Scopo della simulazione
- Fasi della simulazione
- Generazione di numeri casuali
- Esempi

Scopo della simulazione

Le distribuzioni di riferimento usate per determinare grandezze fisiche mediante “fit” non sempre sono esprimibili mediante funzioni analitiche; spesso esse vengono costruite mediante procedure di simulazione dei dati. Nei fenomeni studiati non è usualmente possibile determinare esattamente, sulla base di modelli teorici da verificare nell'esperimento, le grandezze che si andranno a misurare. Ciò avviene per molteplici motivi:

- le condizioni iniziali non sono note con precisione infinita
- nelle misure intervengono errori casuali
- i fenomeni stessi hanno natura non deterministica (meccanica quantistica)
- ...

Non è, in conseguenza di ciò, né possibile né conveniente cercare di costruire una simulazione deterministica dei dati raccolti nello studio di un fenomeno; l'indeterminazione presente nei fenomeni e nell'osservazione sperimentale viene simulata utilizzando la generazione di numeri casuali.

Tecniche “Montecarlo”

L’uso della generazione di numeri casuali nella simulazione di dati ha dato il nome “Montecarlo” alla tecnica che la prevede.

Nelle simulazioni Montecarlo le varie grandezze vengono rappresentate da numeri generati casualmente con densità di probabilità uguale a quella nota (o prevista).

Le grandezze simulate mediante generazione casuale possone avere la natura piú diversa:

- angolo di emissione in un decadimento
- probabilità di decadimento in un canale tra molti possibili
- energia persa da una particella nell’attraversamento di un materiale
- errore casuale commesso in un processo di misura

È facile notare che, come già accennato, parte dell’indeterminazione si trova nei fenomeni stessi e parte nel loro studio sperimentale: risulta quindi opportuno dividere il processo di simulazione in due fasi, nelle quali si generano gli eventi prima e si simula la risposta dell’apparato sperimentale poi.

Fasi della simulazione

Generazione – Simulazione ↑

Dati grezzi (raw data)

Dati fisici (punti, energie)

Segmenti, tracce, particelle

Istogrammi

Eventi

Risultati

Analisi ↓

Generazione di numeri casuali

Nella simulazione Montecarlo riveste una importanza primaria, come si è visto, la possibilità di generare numeri casuali con una densità di probabilità predeterminata, che può avere la forma più diversa.

Nella rappresentazione di due possibili canali di decadimento di una particella, con probabilità P_1 e P_2 , per esempio, è necessario generare un numero casuale distribuito uniformemente tra 0 e 1 e scegliere il primo o il secondo canale se tale numero è, rispettivamente, minore o maggiore di $P1$.

Nella rappresentazione degli errori casuali è usualmente necessario generare un numero distribuito in modo gaussiano, con media e varianza determinate.

In altri casi la densità di probabilità è descritta da altre funzioni.

Molte librerie, facilmente accessibili, forniscono generatori di numeri casuali con distribuzione uniforme o gaussiana; è quindi necessario utilizzare procedure che permettano di ottenere, partendo da essi, numeri casuali distribuiti con la probabilità voluta.

Metodo analitico

Un metodo che permette di costruire una variabile casuale con densità di probabilità descritta dalla generica funzione $f(x)$ con $x_{min} < x < x_{max}$ a partire da una variabile distribuita uniformemente tra 0 e 1 si basa sulla normalizzazione unitaria della probabilità. Integrando la funzione da x_{min} al generico x si ottiene infatti una funzione $F(x)$ compresa tra 0 e 1 :

$$0 < F(x) = \int_{x_{min}}^{x} f(t)dt < 1$$

Si verifica facilmente che la variabile $y = F(x)$ è distribuita uniformemente:

$$n(x, x + dx) = Nf(x)dx = N\frac{dF}{dx}dx = N(F(x + dx) - F(x)) = Ndy$$

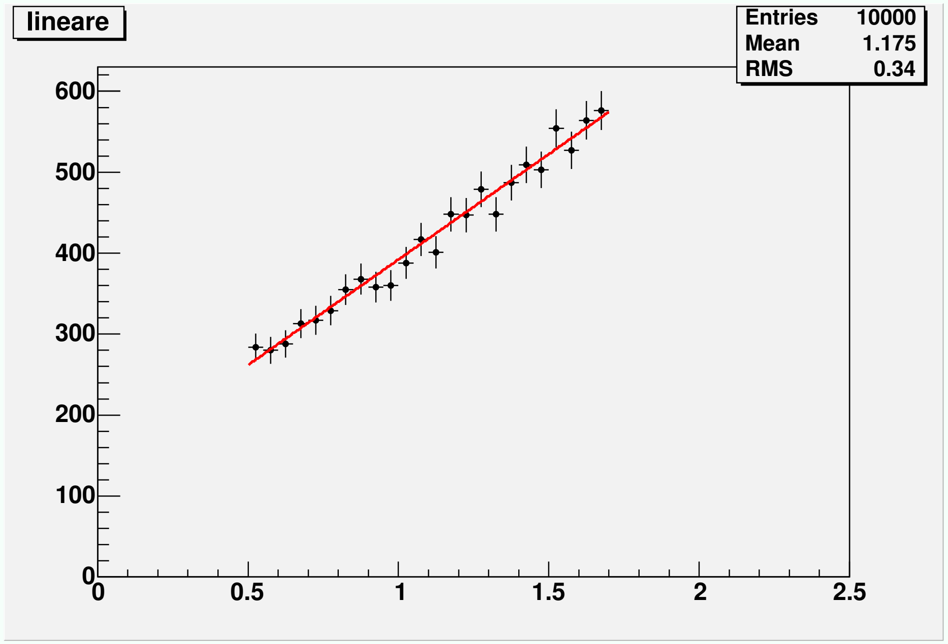
È quindi sufficiente generare un numero casuale y con distribuzione uniforme tra 0 e 1 ed applicare la funzione inversa per trovare un numero distribuito nel modo richiesto:

$$x = F^{-1}(y)$$

Esempio - funzione lineare

$$f(x) = ax+b \quad \Rightarrow \quad F(x) = \frac{ax^2}{2}+bx \quad \Rightarrow \quad F^{-1}(y) = \frac{-b+\sqrt{b^2+2ay}}{a}$$

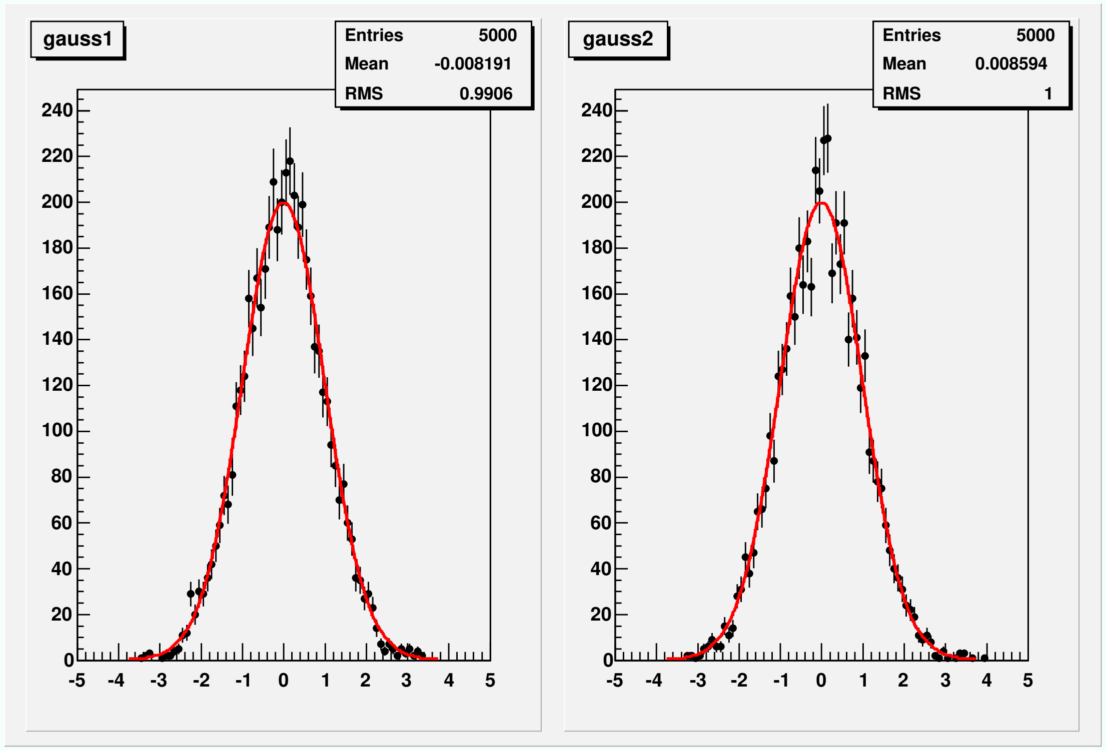
$$x = F^{-1}(F(x_{min}) + (F(x_{max}) - F(x_{min})) * y)$$



Distribuzione gaussiana

La densità di probabilità gaussiana non è integrabile, diventa integrabile però combinando 2 variabili gaussiane ed eseguendo una trasformazione in coordinate polari. Eseguendo tutti i calcoli si trova, per media nulla e varianza unitaria:

$$x_1 = \sqrt{-2\ln(1-y_1)}\cos(2\pi y_2) \quad ; \quad x_2 = \sqrt{-2\ln(1-y_1)}\,\mathrm{sen}(2\pi y_2)$$



Metodo della scelta

Un metodo alternativo, che non richiede l'integrazione analitica della densità di probabilità, si può applicare alle funzioni, definite in un intervallo finito $[x_{min}, x_{max}]$, con un massimo:

$$f(x) < f_{max}$$

Il metodo si basa sulla scelta tra i numeri generati:

1 si genera un numero casuale y distribuito uniformemente tra 0 e 1

2 si costruisce una variabile $x = x_{min} + y(x_{max} - x_{min})$

3 si genera un nuovo numero casuale y distribuito uniformemente tra 0 e 1

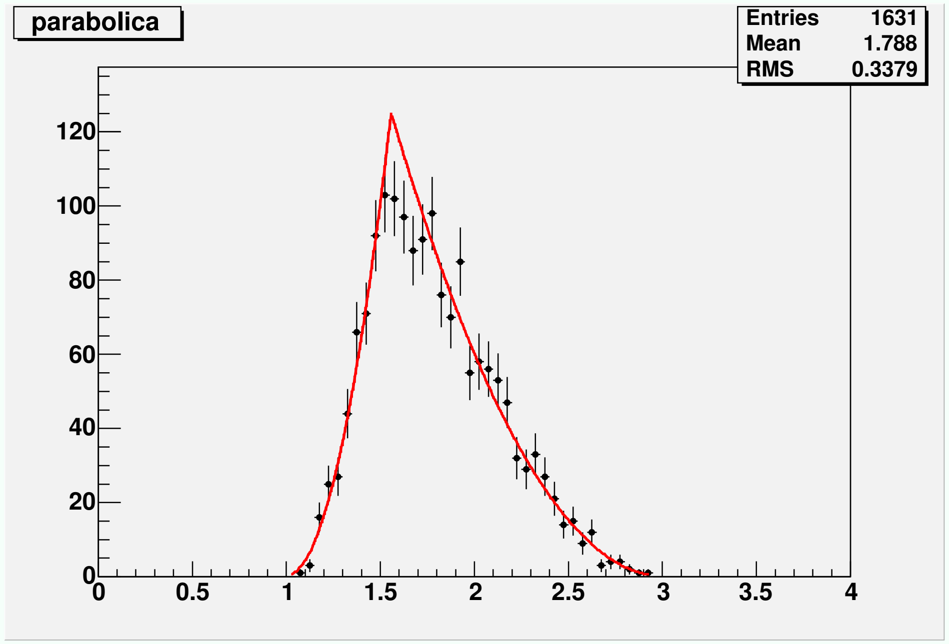
4 se $y < f(x)/f_{max}$ si accetta x, altrimenti si rifiuta

Il metodo diventa inefficiente quando $f_{max}(x_{max} - x_{min}) \gg 1$; in tal caso infatti si spende gran parte del tempo nel generare numeri casuali che verranno poi rifiutati.

Esempio - funzione parabolica

$$f(x) = min(a(x - x_{min})^2, b(x_{max} - x)^2) \qquad \Rightarrow$$

$$f_{max} = \frac{ab}{\left(\sqrt{a} + \sqrt{b}\right)^2} (x_{max} - x_{min})^2$$



generati 5000 numeri: efficienza 32.6% (prevista 33.3%)

Altri metodi

I metodi visti certamente non esauriscono le possibilità.

- Per rendere meno inefficiente il metodo della scelta quando la densità di probabilità presenta forti oscillazioni si può ricorrere ad un cambio di variabile (per es. $y = \ln x$)
- Una distribuzione $f(x)$ data dalla somma di più componenti $f_i(x)$ può essere simulata generando un intero i con probabilità relative pari agli integrali di f_i e generando poi una variabile x con la distribuzione corrispondente
- Piú metodi si possono combinare per simulare una variabile x con densità $g(x) > f(x)$ ed una variabile y distribuita uniformemente ed accettare x se $y < f(x)/g(x)$
- ...

Generazione di variabili correlate

Talvolta è necessario generare piú variabili correlate tra loro; ciò può essere facilmente realizzato mediante una opportuna combinazione di variabili scorrelate.

Date due variabili u e v con varianza unitaria, scorrelate tra loro, ed un parametro r compreso tra -1 e 1 le combinazioni

$$x = u\sqrt{1-r^2} + vr \quad , \quad y = v$$

hanno ancora varianza unitaria ma il loro coefficiente di correlazione è pari a r .

Per ottenere variabili con varianze definite σ_x^2 e σ_y^2 è a questo punto sufficiente moltiplicare rispettivamente per σ_x e σ_y :

$$x = \sigma_x \left(u\sqrt{1-r^2} + vr \right) \quad , \quad y = \sigma_y v$$