



Generazione di distribuzione di probabilità arbitrarie

Distribuzioni
probabilità
arbitrarie

Reiezione

Inversione

Gaussiana
2D

Abbiamo visto come generare eventi con distribuzione di probabilità uniforme, ed abbiamo anche visto in quale contesto tali eventi sono utili.

Tuttavia la maggior parte dei problemi di statistica applicati alla fisica richiedono eventi con **distribuzioni di probabilità non uniformi**.

È quindi importante imparare a **generare eventi distribuiti secondo una generica densità di probabilità** a partire da eventi distribuiti uniformemente.

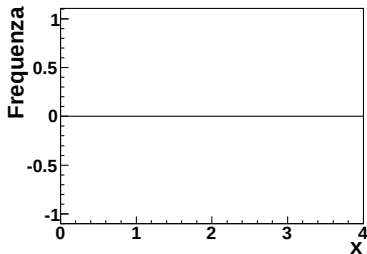
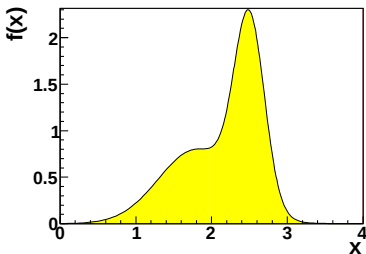


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x,y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.



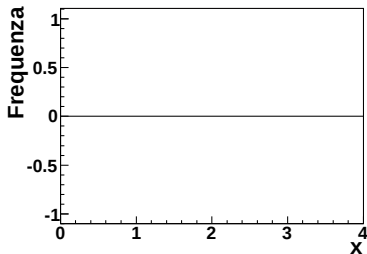
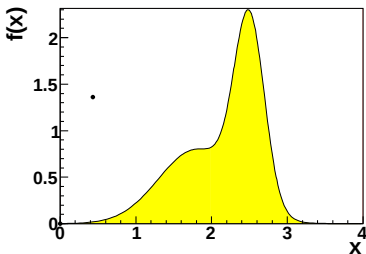


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x, y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.



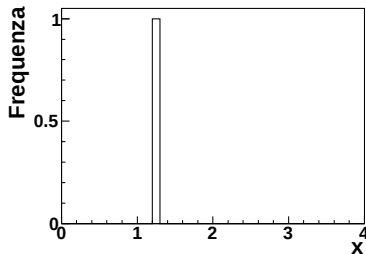
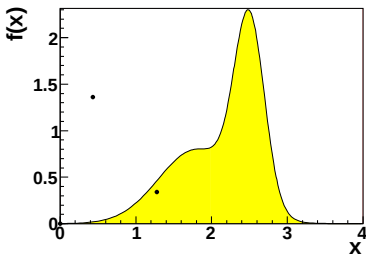


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x,y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.



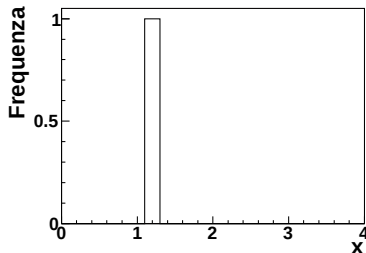
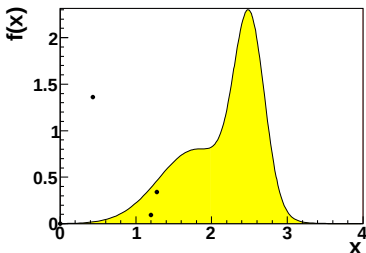


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x,y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.



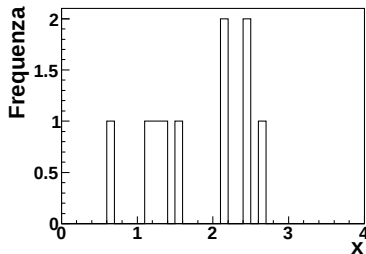
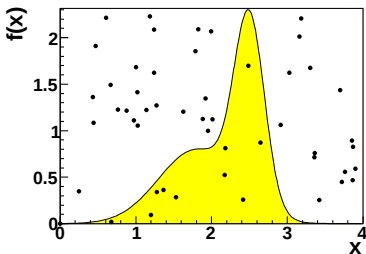


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x,y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.



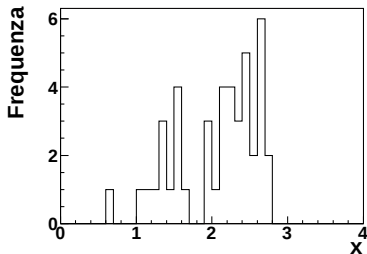
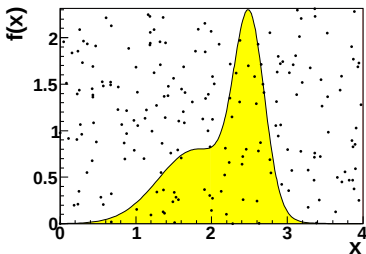


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x,y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.



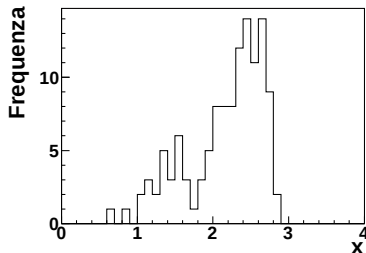
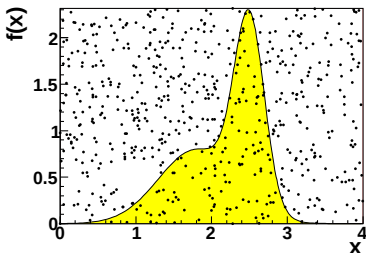


Metodo di reiezione

Supponiamo che la distribuzione da generare $f(x)$

- sia definita nell' intervallo $[x_{min}, x_{max}]$
- sia, in tale intervallo, compresa tra 0 e f_{max}

Se estraiamo un coppia di valori di (x,y) uniformemente distribuiti in $[x_{min}, x_{max}] \times [0, f_{max}]$ accetteremo x se y è minore di $f(x)$ altrimenti lo rigettiamo.





Metodo di reiezione (II)

Distribuzioni
probabilità
arbitrarie

Reiezione

Inversione

Gaussiana
2D

Questo metodo genera, per costruzione, la distribuzione di probabilità desiderata (pesa le x sulla base di $f(x)$).

Il metodo è quello usato per il calcolo dell'integrale con la sola differenza che in quel caso si conta il numero di punti all'interno della funzione mentre in questo caso si prende come numero della sequenza di numeri "random" l'ascissa di ognuno di questi punti.

L'efficienza di un metodo, definita come segue, vale per la reiezione

$$\epsilon = \frac{\text{Numeri accettati}}{\text{Numeri estratti}} = \frac{\int_{x_{\min}}^{x_{\max}} f(x) dx}{(x_{\max} - x_{\min}) y_{\max}}$$



Metodo di reiezione (II)

Distribuzioni
probabilità
arbitrarie

Reiezione

Inversione

Gaussiana
2D

Questo metodo genera, per costruzione, la distribuzione di probabilità desiderata (pesa le x sulla base di $f(x)$).

Il metodo è quello usato per il calcolo dell'integrale con la sola differenza che in quel caso si conta il numero di punti all'interno della funzione mentre in questo caso si prende come numero della sequenza di numeri "random" l'ascissa di ognuno di questi punti.

L'efficienza di un metodo, definita come segue, vale per la reiezione

$$\epsilon = \frac{\text{Numeri accettati}}{\text{Numeri estratti}} = \frac{\int_{x_{\min}}^{x_{\max}} f(x) dx}{(x_{\max} - x_{\min}) y_{\max}}$$



Metodo di reiezione (IV)

Implementazione del metodo di reiezione

Funzione reiezione

Input: funzione ($f(x)$), $f_{\max}$, $x_{\min}$, $x_{\max}$;

Output: numero casuale distribuito secondo $f(x)$

Estraggo coppie di numeri casuali (x, y)

(x in $[x_{\min}, x_{\max}]$ e y in $[0, y_{\max}]$)

finche' y non e' minore di $f(x)$

Quando cio' succede ritorno x

Fine funzione

Prototipi possibili

```
double reiezione(double f(double*,double*), double *par, double t,
                 double xmin, double xmax)
```

```
double reiezione(TF1 f, double fmax, double xmin, double xmax)
```



Limiti del metodo di reiezione

Limiti del metodo:

- si può applicare solo a funzioni limitate e con dominio limitato;
- risulta molto inefficiente nel caso si manipolino distribuzioni che si addensano in piccole regioni dell'intervallo (con picchi stretti).

Il fatto che modulando una distribuzione uniforme con il valore della densità di probabilità si ottenga l'integrale di tale distribuzione ci suggerisce che possa esistere un legame funzionale tra l'integrale di una densità di probabilità ed una variabile aleatoria distribuita uniformemente.

Questa idea è alla base del metodo di inversione.



Metodo di inversione

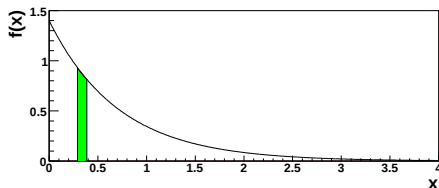
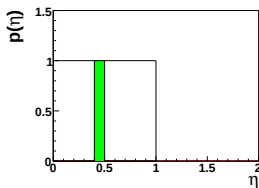
Data $f(x)$, definita in $[x_{min}, x_{max}]$, cerchiamo una funzione g tale che, se η è una variabile aleatoria distribuita uniformemente in $[0, 1]$, allora $g(\eta)$ è distribuita secondo f in $[x_{min}, x_{max}]$.

Poniamo:

$$g(0) = x_{min} , \quad g(1) = x_{max}$$

$$p(\eta)d\eta = f(x)dx$$

(la seconda condizione assicura la conservazione della probabilità)



In questo caso $x_{min} = 0$, $x_{max} = \infty$.



Metodo di inversione (II)

Integriamo la relazione precedente

$$\int_0^\eta d\eta' = \int_{g(0)=x_{min}}^{g(\eta)=x} f(y)dy \quad \Rightarrow \quad \eta = \int_{x_{min}}^x f(y)dy$$

se $f(x)$ non è una densità di probabilità occorre normalizzarla

$$f(x) \longrightarrow \frac{f(x)}{\int_{x_{min}}^{x_{max}} f(x)dx}$$

Risolvendo l'equazione ottenuta con la soluzione dell'integrale si ottiene

$$x = g(\eta)$$



Metodo di inversione (III): esempio

Problema generare numeri casuali secondo $e^{-\lambda x}$ in $[0, \infty)$:

- ottengo la densità di probabilità $f(x) = \lambda e^{-\lambda x}$
- integro:

$$\eta = \int_0^x f(y) dy$$

$$\eta = -e^{-\lambda y} \Big|_0^x = 1 - e^{-\lambda x}$$

- ... e risolvo rispetto a x :

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

Si noti che il metodo di reiezione non richiede che il dominio della funzione sia limitato ma solo che se ne conosca l'integrale analitico.



Generazione della gaussiana

Distribuzioni
probabilità
arbitrarie

Reiezione

Inversione

Gaussiana
2D

La gaussiana non può essere generata con il metodo della reiezione (non ha dominio limitato) né con il metodo di inversione (non ha un semplice integrale analitico).

Metodo della gaussiana 2D

Considero due gaussiane indipendenti (con $\bar{x} = 0$ e $\sigma = 1$) lungo x e y

$$f(x, y) dx dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}} dx dy$$

passo in coordinate polari:

$$r^2 = x^2 + y^2 \quad \theta = \arctan\left(\frac{y}{x}\right)$$

$$dx dy = r dr d\theta$$

$$f(r, \theta) dr d\theta = e^{-\frac{r^2}{2}} dr d\theta$$



Generazione della gaussiana (II)

Effettuo il cambio di variabile $u = r^2/2$, segue $du = r dr$ e quindi:

$$f(x, y) dx dy = f(r, \theta) dr d\theta = e^{-u} du \frac{d\theta}{2\pi}$$

La variabile u potrà essere generata secondo e^{-u} tra $[0, \infty)$ e θ come uniformemente distribuita tra $[0, 2\pi]$; per ricavare x e y basterà effettuare i cambi di variabile a ritroso

$$u = -\ln(1 - \eta_1) \quad r = \sqrt{2u} \quad \theta = 2\pi\eta_2$$

$$x = r \cos \theta$$

$$y = r \sin \theta$$

Abbiamo quindi definito un metodo per generare due numeri casuali indipendenti distribuiti secondo una gaussiana con media 0 e deviazione standard 1 partendo da due numeri uniformemente distribuiti da $[0, 1]$.



Generazione della gaussiana (II)

Come possiamo implementare una funzione che ritorni un solo numero casuale gaussiano senza per questo “buttare via” il secondo numero che viene calcolato dal metodo della gaussiana 2D ?

```
double gaussiana2D(){
    double x;
    static double y;
    static bool compute = true;
    if (compute){
        // calcolo x e y
        compute = false;
        return x;
    } else {
        compute = true;
        return y;
    }
}
```

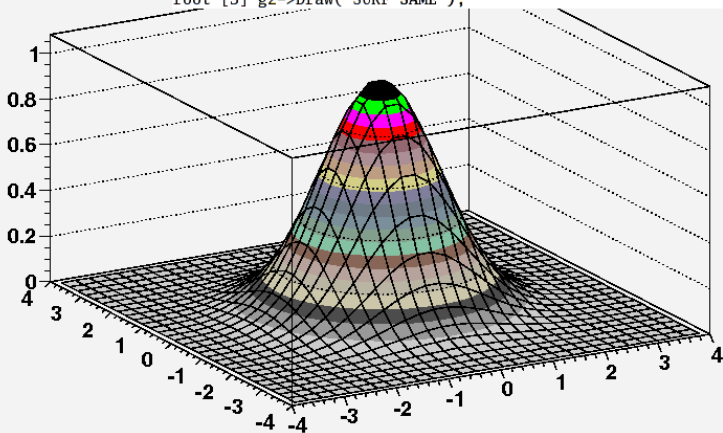
Per estrarre una variabile X distribuita secondo una gaussiana con media $\bar{x}$ e dev. std. σ data x (variabile gaussiana normalizzata) basterà effettuare la trasformazione

$$X = \bar{x} + x\sigma$$



$$\exp(-(x^2+y^2)/2)$$

```
root [1] TF2 *g2 = new TF2("g2","exp(-(x^2+y^2)/2)",-4,4,-4,4);  
root [2] g2->Draw("SURF2");  
root [3] g2->Draw("SURF SAME");
```



Distribuzioni
probabilità
arbitrarie

Reiezione

Inversione

Gaussiana
2D



Traccia dell'esperienza

Distribuzioni
probabilità
arbitrarie

Reiezione

Inversione

Gaussiana
2D

- Generare una distribuzione arbitraria usando il metodo di reiezione.
- Generare una distribuzione arbitraria (con integrale analitico noto) usando il metodo di inversione.
- Generare una gaussiana usando la distribuzione bi-dimensionale.
- Verificare con un fit la correttezza delle distribuzioni ottenute nei tre casi (i parametri ricavati dal fit devono essere compatibili con quelli impostati in generazione).
- La generazione di una distribuzione gaussiana in due dimensioni è una buona occasione per familiarizzare con gli istogrammi bi-dimensionali.