

Lezione 5

Numeri Casuali

<http://www.mi.infn.it/~palombo/didattica/Lab-TNDS/CorsoLab/LezioniFrontali/>

Fernando Palombo

Variabili Casuali

Variabile casuale (vc) è una variabile che assume un valore distinto per ogni elemento dello spazio campione. La variabile assume valori con una probabilità che segue un tipo di distribuzione ben definito.

Può assumere valori discreti o continui.

Le variabili casuali possono essere qualitative e quantitative. Quelle quantitative si riferiscono generalmente a misure sperimentali e noi ci riferiamo a quelle in queste lezioni.

Noi possiamo rifare un esperimento molte volte e vedere la frequenza con cui si hanno i valori osservati all'interno di determinati classi (bin).

Si realizza cioè un istogramma. Vediamo come caratterizzare la distribuzione di probabilità nella misura di qualche processo fisico.

Funzione Densità di Probabilità

$F(x_0)$ = probabilità che la variabile casuale X assuma un valore minore o uguale ad x_0 . Quindi $F(-\infty) = 0$ e $F(+\infty) = 1$.

$$P(x \in [x, x + dx]) = F(x + dx) - F(x)$$

La funzione $f(x) = dF/dx$ è detta funzione densità di probabilità (p.d.f.):

$$P(x \in [x, x + dx]) = F(x + dx) - F(x) = f(x)dx$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad \text{Condizione di normalizzazione}$$

La variabile casuale è qui supposta continua. Se fosse discreta l'integrale verrebbe sostituito da una sommatoria.

Misura con Due Variabili Casuali

$$P(x \in [x, x + dx], y \in [y, y + dy]) = f(x, y) dx dy$$

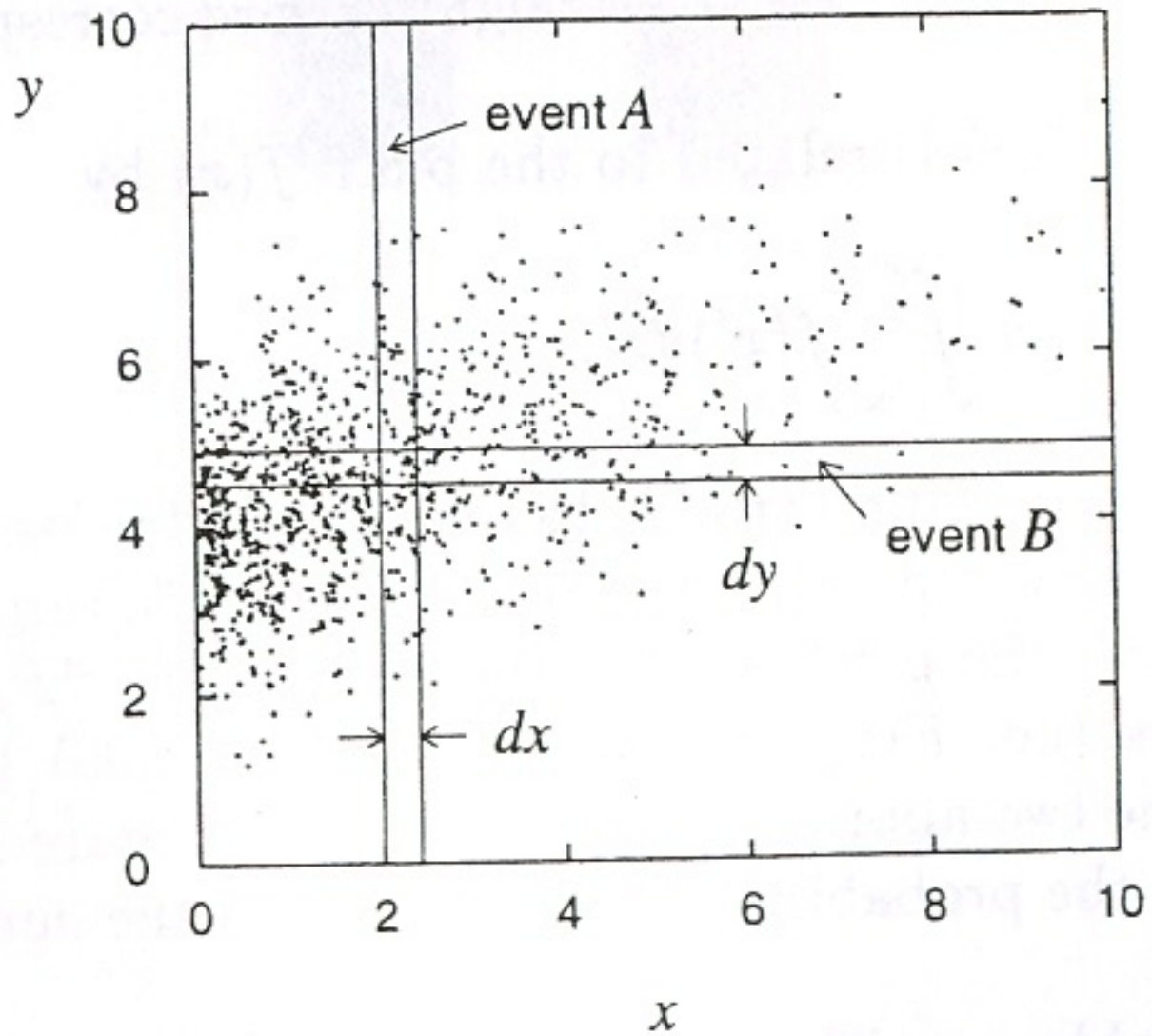
$f(x, y)$ è detta funzione densità di probabilità congiunta.

Se siamo interessati alla p.d.f. della variabile x indipendentemente dai valori assunti dalla variabile y allora

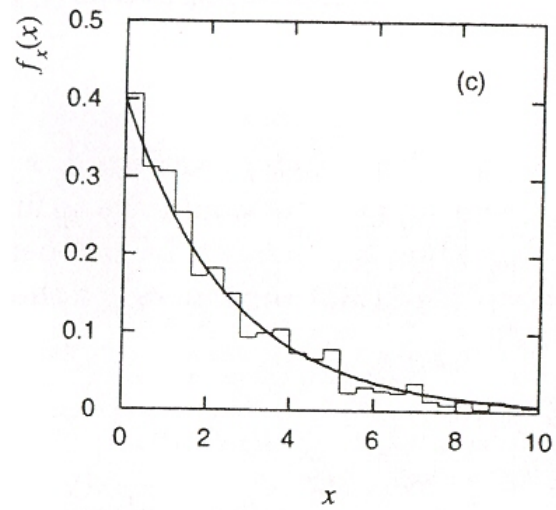
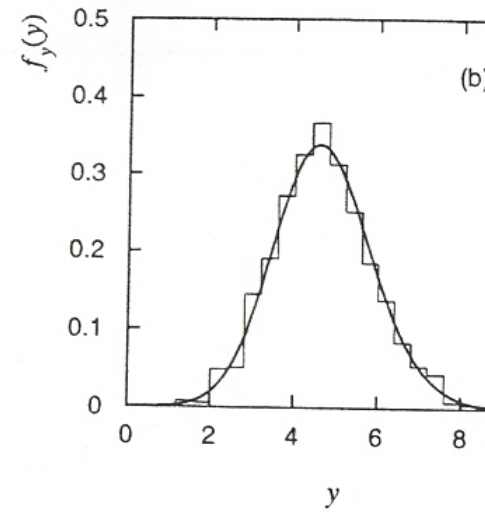
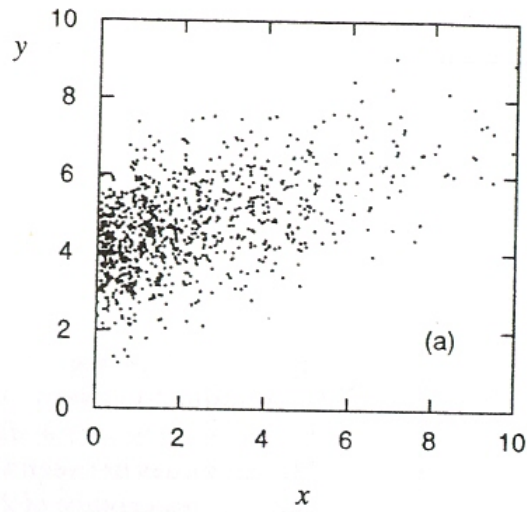
$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

$f_x(x)$ si chiama p.d.f. marginale per x . Analogamente si definisce la marginale per y . $f_y(y)$. In una distribuzione bidimensionale le p.d.f. marginali sono le proiezioni sugli assi (scatter plot)

Se le due variabili sono statisticamente indipendenti (il valore di una variabile non dipende dal valore dell'altra variabile) $\rightarrow f_x(x) = f(x)$; $f_y(y) = f(y)$
 $f(x, y) = f(x) f(y)$



Scatter Plot



Funzione distribuzione Cumulativa

La distribuzione di probabilità di una variabile casuale può essere caratterizzata dalla funzione di distribuzione cumulativa (cdf) $F(x)$:

$$F(x) = \int_{-\infty}^x f(u) du \quad f(x) \text{ è la pdf}$$

$F(x)$ rappresenta la probabilità che la variabile casuale X assuma un valore minore o uguale ad x . La c.d.f. è una funzione crescente da 0 ad 1

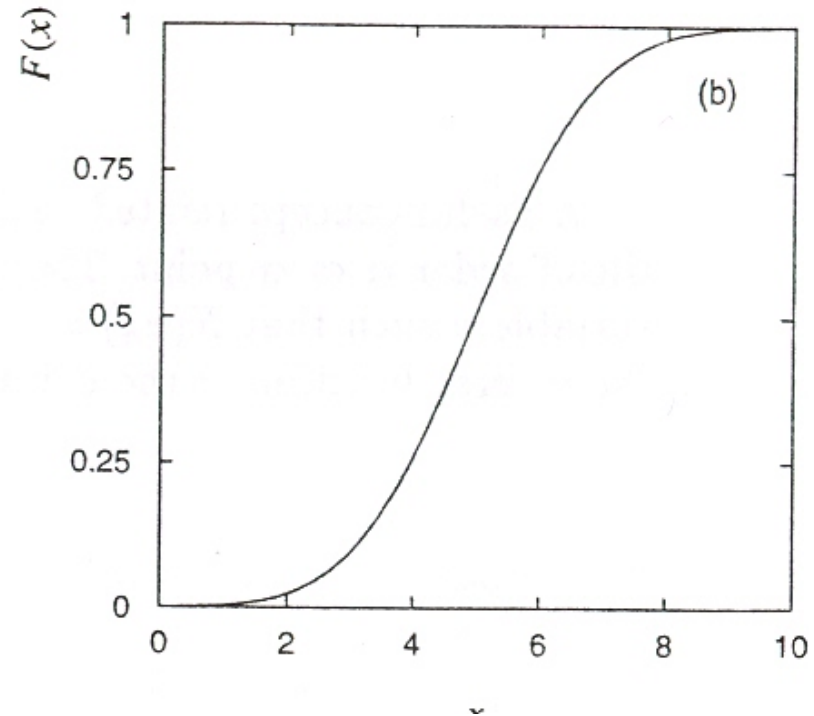
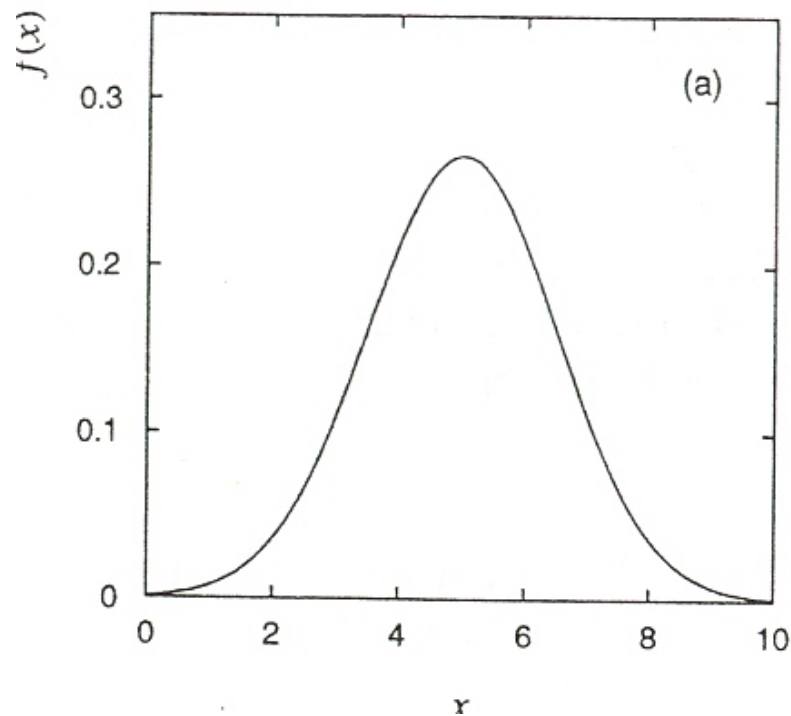
Se $F(x)$ è strettamente crescente, allora tra x e $F(x)$ c'è corrispondenza biunivoca e la $F(x)$ può essere invertita. Chiamiamo punto α o quantile di ordine α il valore x_α della variabile x per il quale **$F(x_\alpha) = \alpha$**

Il quantile non è altro che l'inverso della c.d.f. : $x_\alpha = F^{-1}(\alpha)$

Il quantile più usato è quello che corrisponde ad $\alpha=0.5$ e che è detto mediana:

$$0.5 = \int_{-\infty}^{\text{mediana}} f(x) dx$$

Esempio di p.d.f. e c.d.f.



p.d.f. e corrispondente c.d.f.

Funzioni di Variabili Casuali

Sia $Y=\phi(X)$ una funzione della vc X che sappiamo avere pdf $f(x)$. Vogliamo la p.d.f. $g(y)$ della vc Y .

$P(Y < y) = P(\phi(X) < y)$ e se $\phi(X)$ è **crescente** $\rightarrow P(Y < y) = P[X < \phi^{-1}(y)]$
 $G(y) = F[\phi^{-1}(y)]$ G e F cdf delle vc Y e X .

Derivando rispetto ad y

$$g(y) = f[\phi^{-1}(y)] \frac{d\phi^{-1}}{dy} = f[\phi^{-1}(y)] \frac{dx}{dy} \text{ infatti}$$
$$x = \phi^{-1}(y) \text{ e che } \frac{d\phi^{-1}(y)}{dx} = (d\phi/dx)^{-1} = dx/dy$$

Se la funzione $\phi(y)$ è **decrecente** \rightarrow

$$P(Y < y) = P[X > \phi^{-1}(y)] = 1 - P[X < \phi^{-1}(y)]$$

$$G(y) = 1 - F[\phi^{-1}(y)]$$

Derivando entrambi i membri rispetto ad y

Funzioni di Variabili Casuali

$g(y) = - f(\Phi^{-1}(y)) d\Phi^{-1}/dy$ e quindi combinando assieme si ha:

$$g(y) = f(x) | dx/dy |$$

Il termine in valore assoluto si dice **jacobiano** della trasformazione

Il risultato si generalizza al caso di una funzione di n vc e jacobiano della trasformazione diventa il determinante di una matrice $n \times n$

Sia $Y = X^2$ e sia X una vc distribuita con p.d.f. $f(x) = 1$ per $0 < x < 1$ e 0 altrove

$$dy/dx = 2x; \quad g(y) = 1/2x$$

Poichè $x = \sqrt{y} \rightarrow g(y) = 1/(2\sqrt{y})$ con $0 < x < 1$

Si verifica facilmente che : $\int_0^1 \frac{1}{2\sqrt{y}} dy = 1$

Valori di Aspettazione

Sia X vc con p.d.f. $f(x)$. Si dice valore di aspettazione della vc X la quantità:

$$E[x] = \int_{-\infty}^{+\infty} x f(x) dx = \mu \quad \text{è detta anche media aritmetica}$$

Se $a(x)$ funzione di X allora $E[a] = \int_{-\infty}^{+\infty} a(x) f(x) dx$

Momento centrale secondo : $E[(x - E[x])^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \sigma^2 = V[x]$

È chiamato varianza $V_{[x]}$. Si può vedere facilmente che:

$$V_{[x]} = \overline{x^2} - \mu^2$$

La varianza è data dalla media dei quadrati meno il quadrato della media

Valori di Aspettazione

La radice quadrata della varianza è detta deviazione standard σ

Esempio: particelle distribuite in modo casuale in un quadrato di lato a con lati paralleli agli assi. Calcolare la posizione media, la varianza e la σ

$f(x) = k$ costante $\int_0^a k dx = 1$ da cui $k=1/a$ e quindi

$$\bar{x} = \int_0^a \frac{1}{a} x dx = \frac{a}{2}$$

$$\overline{x^2} = \int_0^a \frac{1}{a} x^2 dx = \frac{a^2}{3}$$

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2} = a/\sqrt{12} \qquad \sigma_x^2 = \frac{a^2}{3} - \left(\frac{a}{2}\right)^2 = \frac{a^2}{12}$$

Matrice di Covarianza

Se si hanno due vc X e Y con pdf congiunta $f(x, y)$ oltre alle varianze delle singole X e Y bisogna considerare anche la correlazione tra queste variabili.

La correlazione V_{xy} è detta **covarianza** ed è definita così:

$$V_{xy} = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x\mu_y$$

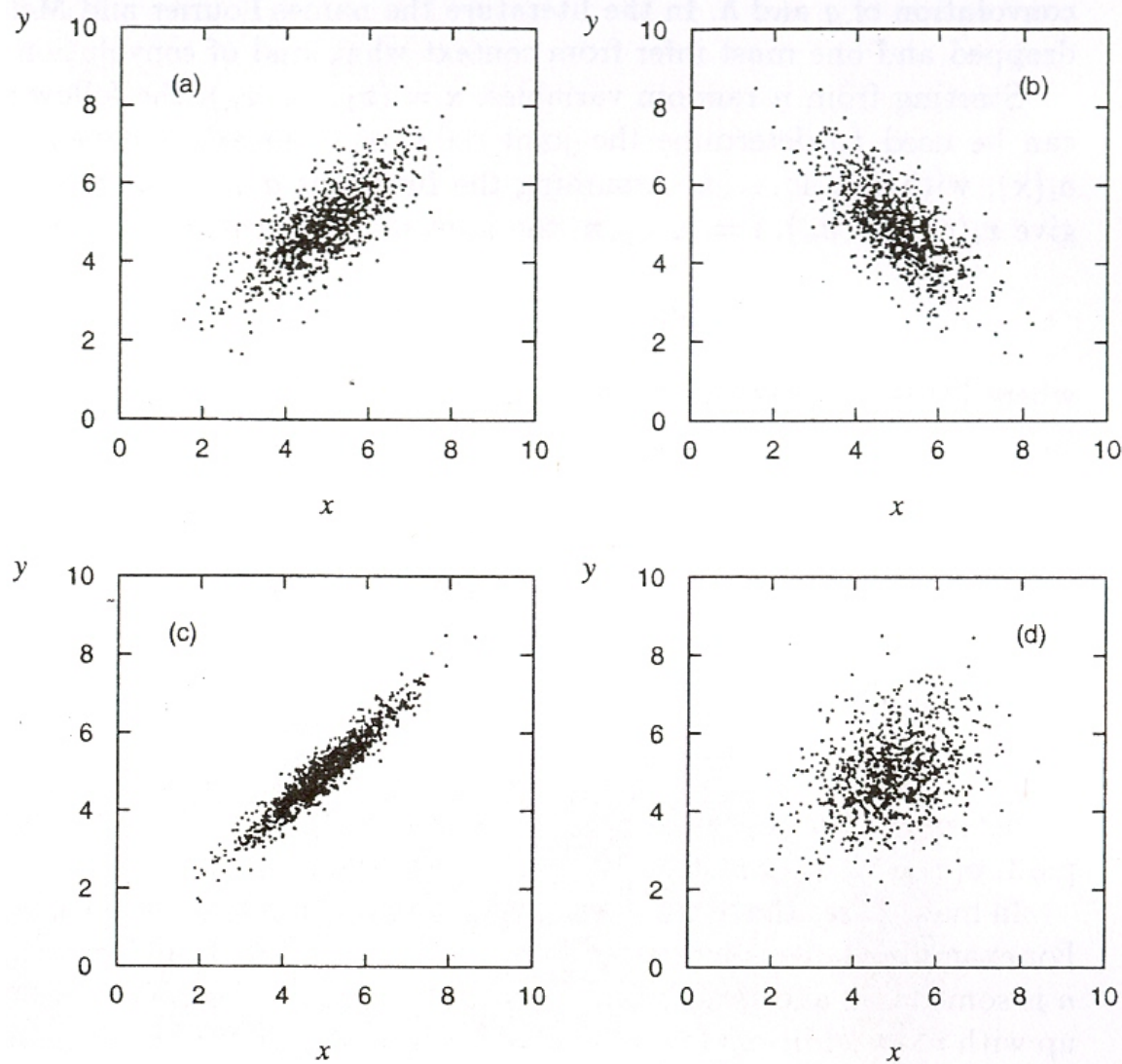
$$V_{xy} = \int_{-\infty}^{+\infty} xy f(x, y) dx dy - \mu_x\mu_y$$

La matrice di covarianza ha elementi diagonali positivi (le varianze di X e Y) Ed è una matrice simmetrica ($V_{xy} = V_{yx}$). È detta anche **matrice degli errori**

Si definisce **coefficiente di correlazione** (lineare) la quantità : $\rho_{xy} = \frac{V_{xy}}{\sigma_x\sigma_y}$

La correlazione è tra -1 e 1 . Se X e Y sono indipendenti allora $\rho_{xy} = 0$ ma non necessariamente vale il viceversa : solo scorrelazione al primo ordine ma la correlazione potrebbe esistere a livello superiore. Se $\rho_{xy} > 0$, correlazione positiva: aumentando X aumenta Y altrimenti negativa

Coefficiente di Correlazione



Distribuzione Binomiale

Esperimento con due sole uscite : A con probabilità p e B con probabilità $q=1-p$. Si pensi al lancio della moneta dove $p = 0.5$ oppure ad un campione di dati dove c è il segnale (che state cercando) ed il fondo (che volete scartare). È chiaramente una variabile discreta.

La probabilità di avere n segnali consecutivi su N esperimenti è data da $p^n q^{N-n}$

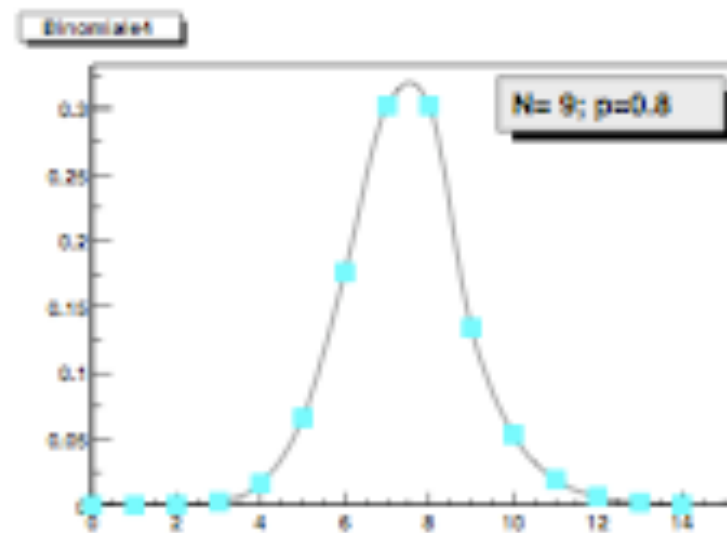
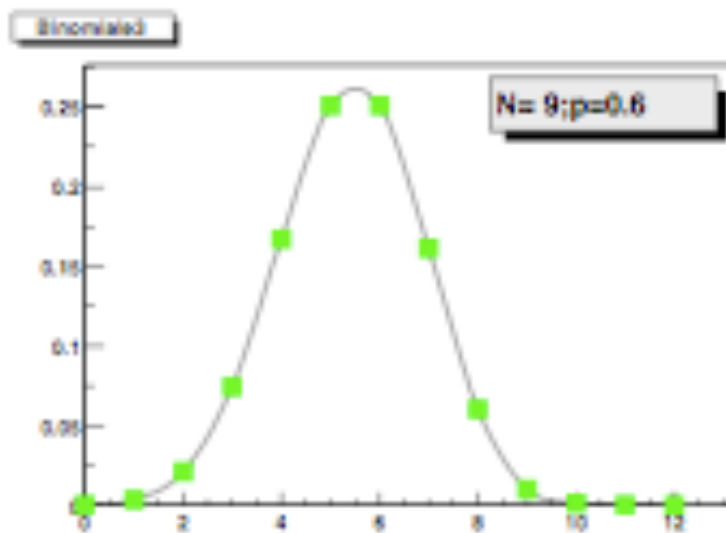
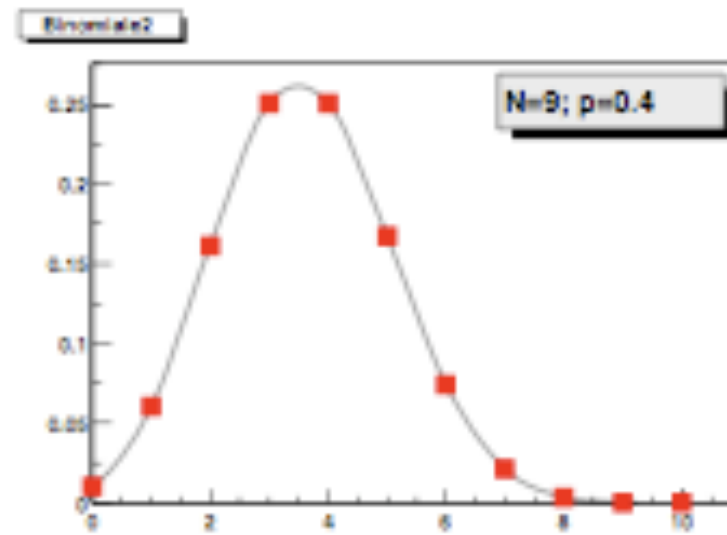
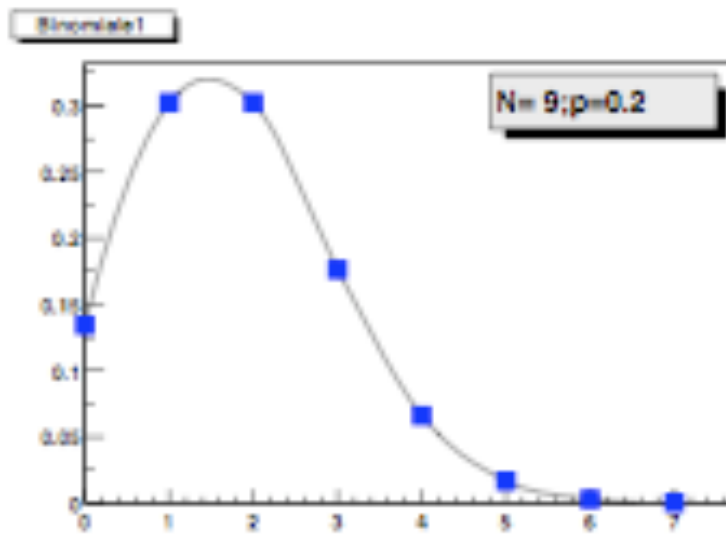
Se non siamo interessati all'ordine con cui ho i segnali allora si ha:

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad n = 0, 1, \dots, N$$

Valore di aspettazione di n :
$$E[n] = \sum_{n=0}^{\infty} n \frac{N!}{n!(N-n)!} p^n q^{N-n} = Np$$

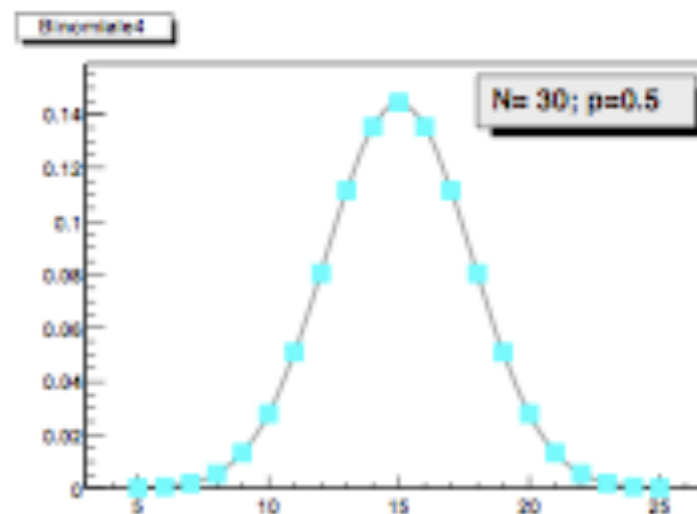
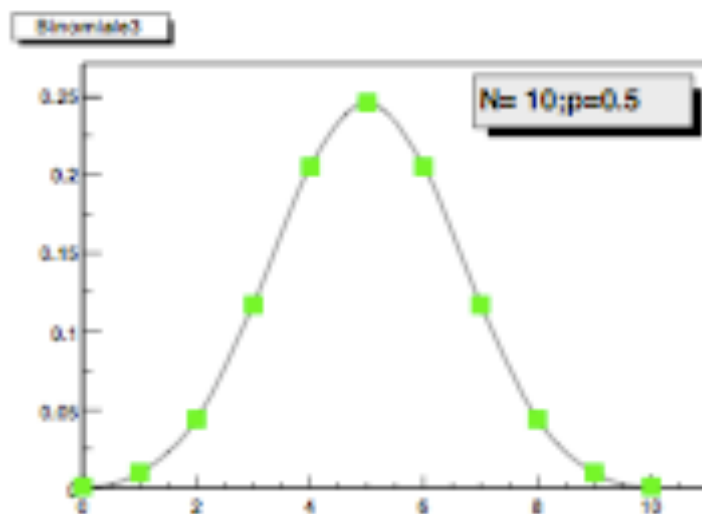
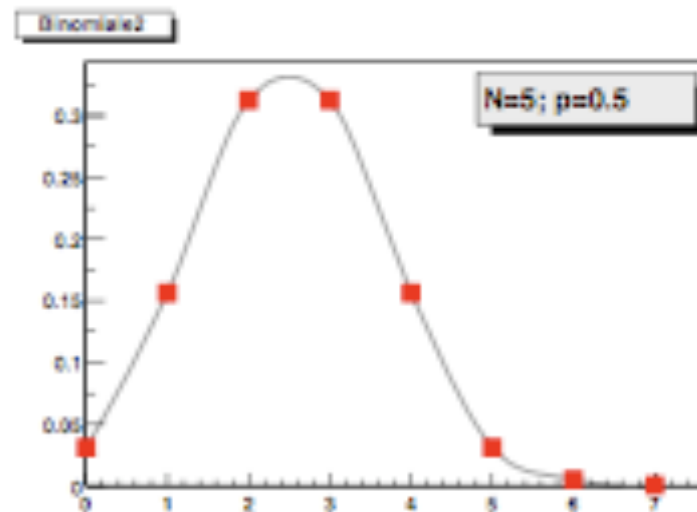
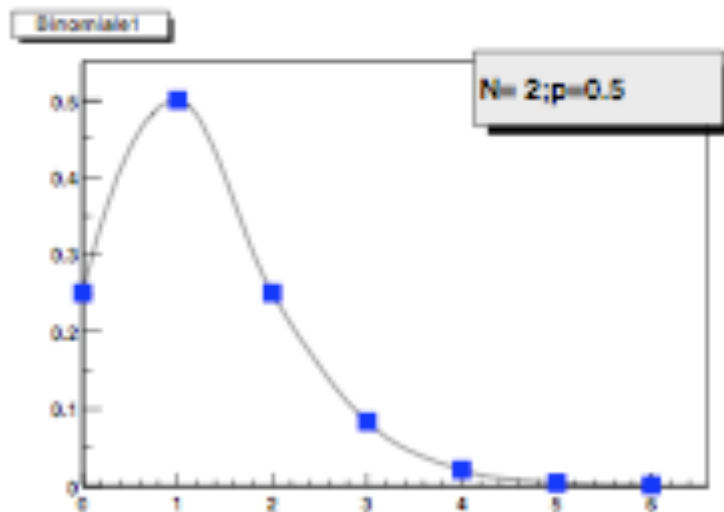
Varianza di n :
$$V[n] = E[n^2] - (E[n])^2 = Npq$$

Esempi di Distribuzione Binomiale



Numero N di esperimenti costante e diversi valori della probabilità p

Esempi di Distribuzione Binomiale



Numero N di esperimenti variabile e identico valore della probabilità p

Distribuzione di Poisson

Se la dimensione del campione N è molto grande, p la probabilità di vedere un segnale è molto piccolo (evento raro) ma Np è una quantità costante allora la distribuzione binomiale diventa quella di Poisson:

$$f(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$$

Il valore di aspettazione di n è: $E[n] = \sum_{n=0}^{\infty} n \frac{\nu^n}{n!} \exp[-\nu] = \nu$

mentre la varianza è: $V[n] = \sum_{n=0}^{\infty} (n - \nu)^2 \cdot \frac{\nu^n}{n!} \cdot \exp[-\nu] = \nu$

Questa distribuzione è tra le più adottate per descrivere fenomeni naturali (tipo il decadimento radioattivo). È usata quando un fenomeno è raro ma con la statistica a disposizione è finita (non nulla) la probabilità di osservarlo.

Distribuzione di Poisson

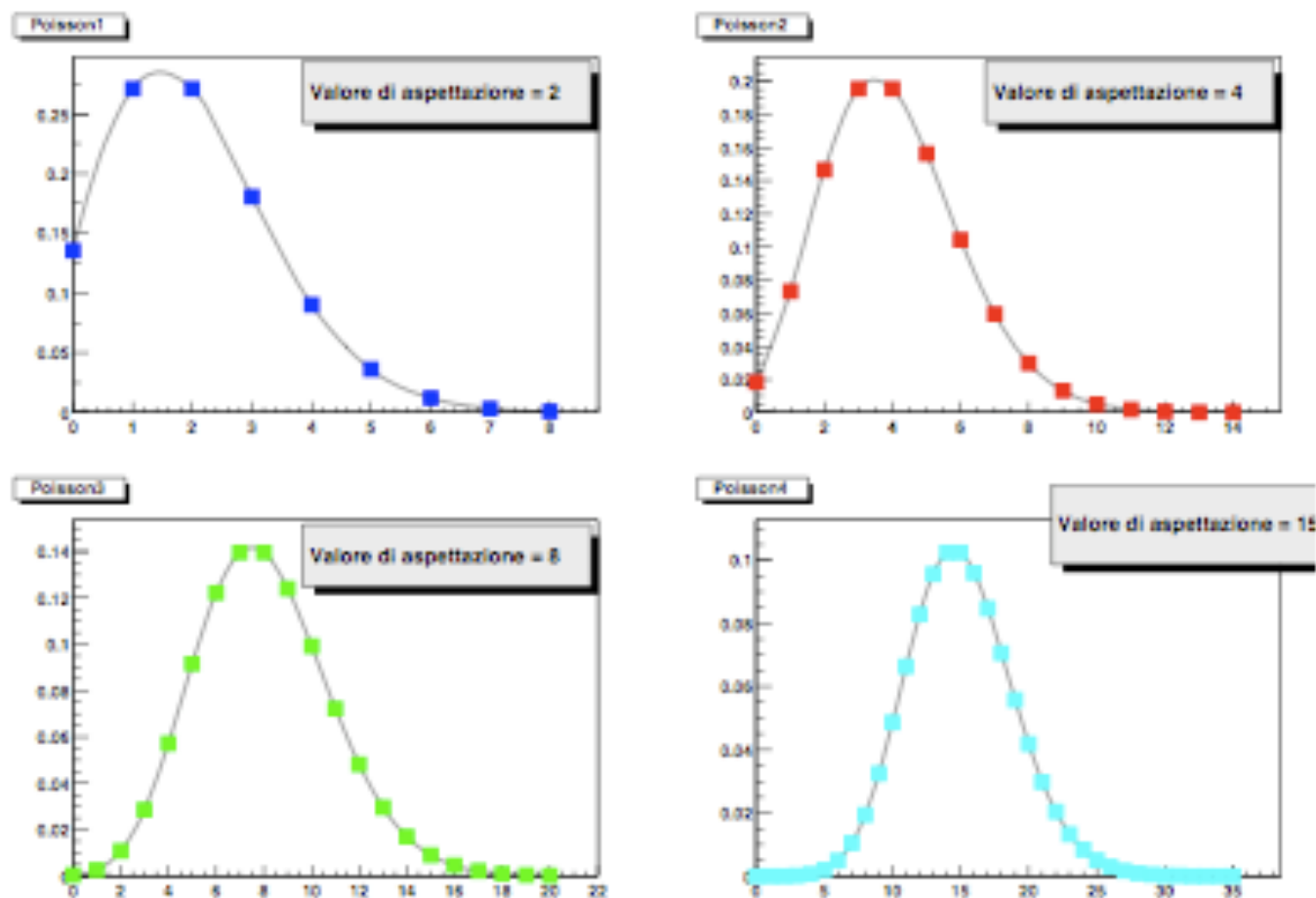


Figure 3.3: Distribuzioni poissoniane con diversi valori di aspettazione.

Distribuzione di Poisson

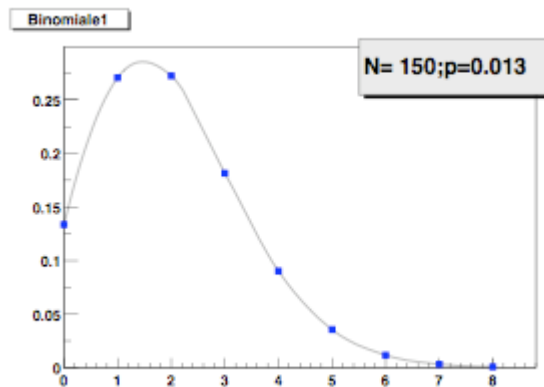
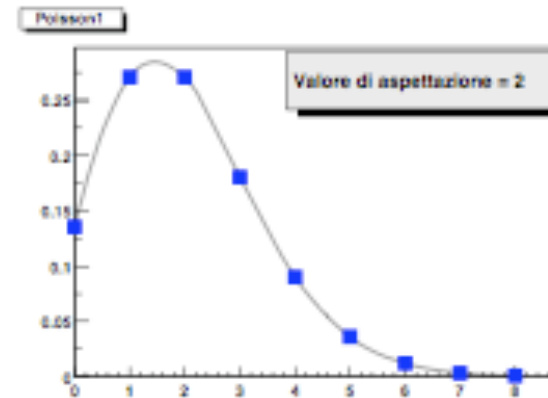


Figure 3.4: Distribuzione binomiale con $N = 150$ e $p = 0.013$



- ❑ In questa distribuzione binomiale (a sinistra) $Np = 1.95$. Confrontare questa distribuzione con quella poissoniana (a destra) con $\nu = 2$.
 - ❑ Con N molto grande e p molto piccolo in modo che Np resti un numero di eventi finito e osservabile, allora la distribuzione binomiale diventa quella di Poisson.
 - ❑ La distribuzione poissoniana è tra quelle più adottate nella descrizione di fenomeni naturali (come nei decadimenti radioattivi, ecc)
-

Distribuzione Gaussiana

Una v.c. X segue una distribuzione gaussiana se la sua pdf è

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

dove i due parametri μ e σ sono il valore di aspettazione e la deviazione standard

$$E[x] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] dx = \mu$$

$$V[x] = \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] dx = \sigma^2$$

La distribuzione gaussiana è spesso scritta così $N(\mu, \sigma^2)$. La gaussiana standard è quella con $\mu=0$ e $\sigma = 1$, cioè

$$N(0,1) = \phi(x) = f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{x^2}{2} \right]$$

La cdf della gaussiana standard è $\Phi(x) = \int_{-\infty}^x \phi(z) dz$

Distribuzione Gaussiana

Data una variabile Y con distribuzione gaussiana $N(\mu, \sigma^2)$, allora la variabile

$X = (Y - \mu)/\sigma$ segue una distribuzione gaussiana standard $N(0,1)$.

Inoltre le corrispondenti cdf $F(y)$ e $\Phi(x)$ sono uguali.

I valori di $\Phi(x)$ ed i quantili $x_\alpha = \Phi^{-1}(\alpha)$ sono tabulati (oggi si trova tutto facilmente in rete)

Calcolare la probabilità che una variabile gaussiana con media uguale a 10 e varianza uguale a 100 assuma un valore compreso tra 8 e 16:

$$p(8 \leq X \leq 10) = \int_{\frac{8-10}{10}}^{\frac{16-10}{10}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \simeq 0.305$$

La distribuzione gaussiana è di gran lunga la più importante (come vedremo) tra tutte le distribuzioni.

http://socr.ucla.edu/htmls/SOCR_Distributions.html

Legge dei Grandi Numeri

Sia $E[x] = \mu$ il valore di aspettazione di una vc X e supponiamo di stimarla utilizzando un campione finito di n misure.

Sia \bar{x}_n la media aritmetica (campionaria) di queste n misure. Posso usare la media campionaria per fare stime attendibili di μ ?

Sì, è possibile utilizzando questa legge dei grandi numeri:

È possibile determinare un intero positivo n tale che prendendo un campione casuale di dimensioni maggiore o uguale ad n di una vc X distribuita con valore di aspettazione μ , la media (campionaria) differisca da μ per una quantità piccola a piacere.

Questa legge è alla base della **inferenza statistica**

Teorema Limite Centrale

Siano date n vc x_1, x_2, \dots, x_n , indipendenti, ognuna con media μ_i e varianza σ_i^2 , il teorema limite centrale stabilisce che :

la vc $x = \sum_{i=1}^n x_i$ per grandi n tende ad essere distribuita secondo una curva gaussiana di valore medio $\mu = \sum_i \mu_i$ e varianza $\sigma^2 = \sum_i \sigma_i^2$

La cosa veramente importante di questo teorema è che quanto stabilito è vero indipendentemente dal tipo di distribuzione delle vc x_i . Con tante vc comunque distribuite l'effetto cumulativo è quello di spingere verso una distribuzione di tipo gaussiano.

Si pensi alle molteplici sorgenti indipendenti di errore nelle misure sperimentali che alla fine portano a distribuzioni di tipo gaussiano.

In diversi casi con statistiche limitate e per semplificare le distribuzioni si assumono gaussiane anche quando non lo sono (l'esempio di code non gaussiane)

Generazione di Numeri Casuali

Il calcolatore essendo una macchina deterministica non può generare una sequenza di veri numeri casuali. È possibile però che la macchina utilizzando particolari algoritmi sia in grado di produrre una sequenza di numeri che ai fini pratici per i quali sono generati si possano considerare casuali.

Questi numeri sono detti **pseudo-casuali**. Vediamo un esempio di algoritmo di generazione (detto *generatore lineare congruenziale*):

$$r_i = \text{mod}(a * r_{i-1} + c, m)$$

La funzione $\text{mod}(x, y) = x - \text{int}(x/y) * y$ ritorna il resto della divisione tra x e y

Prendiamo per esempio (illustrativo) $c=1$, $a=4$ e $m=9$.

Iniziamo con $r_1 = 3$ [questo valore iniziale è detto **seed (seme)**], allora $r_2 = 4$, $r_3 = 8$, $r_4 = 6$, ecc, $r_{10} = 3$, $r_{11} = 4$, ecc, ecc. La nostra sequenza è lunga 9 (come m che rappresenta il **periodo** dopo il quale la sequenza viene rigenerata).

m deve essere il più grande intero possibile. In una macchina a 32 bit il massimo intero è $2^{32} - 1$

Generazione di Numeri Casuali

Anche il parametro a va scelto in modo opportuno ($a = 16807$). Si noti che avendo fissato i parametri (a e c), l'algoritmo ripete la sequenza se gli viene dato lo stesso seed altrimenti produce un'altra sequenza.

La riproducibilità della sequenza, il costo computazionale e la lunghezza del periodo sono aspetti molto importanti dell'algoritmo di generazione.

Esistono molti generatori di numeri pseudo-casuali. Assumendo che `Random()` generi una distribuzione uniforme di numeri tra 0 e 1, se voglio generare una distribuzione uniforme di numeri ϕ tra 0 e 2π allora faccio $\phi = 2\pi \text{Random}()$

Noi vogliamo ora vedere come sia possibile a partire da distribuzioni uniformi generare numeri, che chiamiamo d'ora in poi casuali, che siano distribuiti secondo un altro tipo di distribuzione.

Metodo della Trasformata

Una vc r sia distribuita in modo uniforme. Noi vogliamo determinare una funzione $x(r)$ la quale segua una distribuzione $f(x)$. Sia $g(r)$ la pdf della variabile r .

La probabilità di avere un valore tra r e $r + dr$ (cioè $g(r) dr$) deve essere uguale alla probabilità di avere un valore tra $x(r)$ e $x(r) + dx(r)$ (cioè $f(x) dx$):

$$f(x) dx = g(r) dr$$

Integrando passo alle cdf : $F(x(r)) = G(r)$

$$F(x(r)) = \int_{-\infty}^{x(r)} f(x') dx' = \int_{-\infty}^r g(r') dr' = r$$

Se la distribuzione è uniforme allora **$G(r) = r$** ($0 \leq r \leq 1$).

Se riesco a risolvere analiticamente questa relazione ottengo $x(r)$. Questo è possibile solo in qualche caso.

Distribuzione Esponenziale

Consideriamo un esempio in cui si riesce a risolvere analiticamente la relazione appena vista. È questo il caso della distribuzione esponenziale:

$$\int_0^{x(r)} \frac{1}{\xi} e^{-x'/\xi} dx' = r$$

Questa equazione si sa integrare ed ha soluzione: $x(r) = -\xi \log(1 - r)$

Se una variabile r è uniforme in $[0,1]$, lo è anche la variabile $1-r$. Quindi:

$$x(r) = -\xi \log r$$

Se genero uniformemente r tra 0 e 1, i valori $x(r)$ saranno distribuiti secondo una curva esponenziale (in queste distribuzioni ξ rappresenta il valore di aspettazione di x mentre ξ^2 è la varianza σ^2 di x). Nel caso di decadimenti di risonanze studiate nel loro sistema di decadimento, ξ rappresenta la vita media delle risonanze.

Algoritmo di Box-Muller

- ❑ Consideriamo due variabili casuali y_1 e y_2 funzioni di due variabili casuali x_1 e x_2 distribuite queste ultime uniformemente in $[0,1]$.
- ❑ Vogliamo determinare due funzioni delle variabili x_1 e x_2 che permettano di ottenere variabili y_1 e y_2 indipendenti tra di loro distribuite in modo gaussiano standard $N(0,1)$. La pdf congiunta $g(y_1, y_2)$ deve essere:

$$g(y_1, y_2) \equiv \frac{1}{2\pi} e^{-\frac{y_1^2}{2}} e^{-\frac{y_2^2}{2}} = \frac{1}{2\pi} e^{-(y_1^2 + y_2^2)/2}$$

- ❑ La pdf congiunta $g(y_1, y_2)$ deve essere legata alla pdf congiunta $f(x_1, x_2)$ dalla relazione:

$$g(y_1, y_2) = f(x_1, x_2) |\det J|$$

dove J è la matrice jacobiana:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Algoritmo di Box-Muller

□ Si ha perciò: $g(y_1, y_2) = f(x_1, x_2) |det J| = \left| \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1} \right|$

ed essendo $f(x_1, x_2) = 1$ si ha:

$$\frac{1}{2\pi} e^{-(y_1^2 + y_2^2)/2} = \left| \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1} \right|$$

□ Si può dimostrare che le trasformazioni cercate sono:

$$\begin{aligned} y_1 &= \sqrt{-2 \ln x_2} \sin(2\pi x_1) \\ y_2 &= \sqrt{-2 \ln x_2} \cos(2\pi x_1) \end{aligned}$$

□ Estrahendo due numeri a caso x_1 e x_2 da una distribuzione uniforme in $[0, 1]$, le variabili y_1, y_2 seguono una distribuzione $N(0,1)$. Su queste trasformazioni si basa l'algoritmo di *Box-Muller*.

Algoritmo di Box-Muller

- ❑ La $g(y_1, y_2)$ è una funzione radiale che è comodo esprimere in forma polare (R, θ) con θ uniforme in $[0, 2\pi]$. Quindi $\theta = 2\pi x_1$ e

$$\begin{aligned} y_1 &= R \cos \theta \\ y_2 &= R \sin \theta \end{aligned} \quad \text{con} \quad R \equiv \sqrt{y_1^2 + y_2^2} \text{ e } \theta = \tan^{-1}\left(\frac{y_2}{y_1}\right)$$

- ❑ Da queste si ha anche $R = \sqrt{-2 \ln x_2}$
- ❑ Estratti le due variabili x_1 e x_2 , si ottengono sia θ che R e quindi y_1 e y_2 che seguono $N(0,1)$
- ❑ Il calcolo di funzioni trigonometriche hanno scarsa efficienza computazionale. Una variante del metodo di Box-Muller che non fa uso di funzioni trigonometriche è il metodo polare di *Marsaglia*

Metodo Polare di Marsaglia

- ❑ Generiamo due variabili x_1 e x_2 uniformemente distribuite in $[-1, +1]$ e accettiamo solo i punti che si trovano all'interno del cerchio unitario:

$$s = x_1^2 + x_2^2 \leq 1$$

- ❑ Allora le seguenti due variabili y_1 e y_2 sono indipendenti e distribuite in modo normale standard:

$$y_1 = x_1 \sqrt{\frac{-2\ln(s)}{s}}$$
$$y_2 = x_2 \sqrt{\frac{-2\ln(s)}{s}}$$

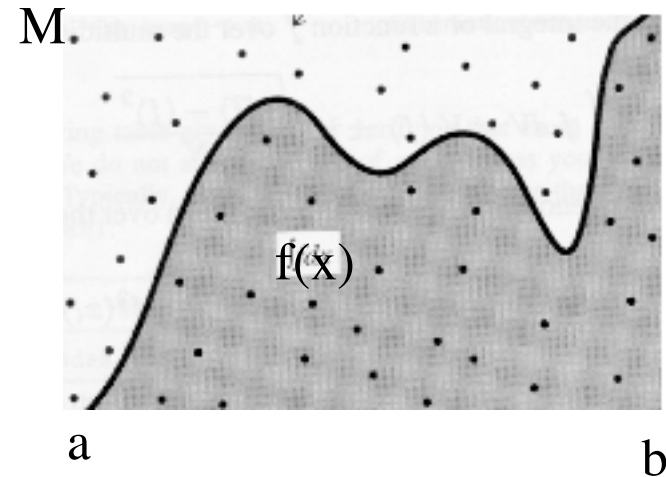
- ❑ Si noti che se y_1 è una variabile distribuita in modo $N(0,1)$, allora la variabile $z_1 = \mu + \sigma y_1$ è una variabile distribuita secondo una gaussiana di media μ e deviazione standard σ

Metodo dell'Accettazione-Reiezione

Quando non si può applicare il metodo della trasformata, allora si può utilizzare il metodo dell'accettazione-reiezione (hit and miss).

Vogliamo generare numeri casuali compresi tra a e b distribuiti secondo una pdf $f(x)$.

Indichiamo con M il massimo valore della funzione $f(x)$ in $[a, b]$.



1. generiamo due numeri casuali r_1 e r_2 uniformemente distribuiti tra $[0, 1]$ e poniamo : $x_i = a + (b-a) \cdot r_1$ e per questo x_i prendiamo $y_i = M \cdot r_2$
2. se $y_i > f(x_i)$ si ritorna al punto 1 (il punto y_i viene rigettato)
3. se $y_i \leq f(x_i)$, allora x_i viene accettato. I punti accettati sono distribuiti come $f(x)$ per costruzione.

Metodo dell'Accettazione-Reiezione

Questo metodo se la pdf $f(x)$ fosse stretta avrebbe bassa efficienza (e quindi alto costo computazionale) perché la gran parte dei punti generati starebbe fuori della funzione $f(x)$ e quindi verrebbero rigettati.

Il metodo può essere migliorato inscrivendo la funzione $f(x)$, invece che nel rettangolo visto prima, in una curva $g(x)$ che rappresenta una pdf di cui io so generare eventi. Per esempio potrei accorgermi che la funzione $f(x)$ è inscrivibile all'interno di una curva gaussiana.

Allora genero punti in accordo alla gaussiana e li accetto o rigetto come fatto in precedenza. In questo caso l'efficienza può aumentare di molto!

Stimatori Puntuali

Supponiamo di aver fatto n misure sperimentali x_1, x_2, \dots, x_n di una variabile X . Io in generale non conosco la distribuzione esatta di questa variabile. Talvolta però riconosco che la distribuzione della variabile X appartiene ad una famiglia parametrica : $f(x) = f(x | \theta) \theta \in \Theta$

Noi quindi dobbiamo trovare un modo per stimare il parametro (o i parametri) θ . Per esempio potrei accorgermi che i dati sono distribuiti come un esponenziale e allora il parametro da stimare sarebbe ξ .

Il processo che porta alla determinazione del parametro θ si dice fit del parametro.

Gli stimatori puntuali stimano il valore del parametro con una incertezza.

Gli stimatori intervallari stimano l'intervallo in cui si trova il valore del parametro con una predefinita probabilità.

Una generica funzione t delle misure fatte è detta statistica; $t = t(x_1, \dots, x_n)$

Stimatori Puntuali

Una statistica concepita per stimare un parametro è detto stimatore. Un esempio di statistica può essere $x_m = (x_1 + x_2 + \dots + x_n)/n$. Questa statistica la uso per stimare la media aritmetica delle misure. Quindi questa statistica è lo stimatore della media aritmetica.

Io però posso avere tanti stimatori della media. Per esempio di tutte le n misure potrei togliere i 4 valori più alti e i quattro più bassi e pigliare la media aritmetica delle misure restanti.

Potrei decidere di usare solo le misure che appaiono in ordine pari oppure fare la media delle prime dieci misure.

Tutti questi sarebbero stimatori della media. Ma come faccio a decidere che uno stimatore è preferibile agli altri ?

Proprietà degli Stimatori Puntuali

Uno stimatore lo consideriamo buono se è consistente, non distorto ed efficiente.

Stimatore Consistente : è uno stimatore che tende al valore vero quando il numero delle misure tende all'infinito

Stimatore non distorto : è uno stimatore in cui il valore di aspettazione coincide col valore vero

Stimatore efficiente: è lo stimatore che stima il parametro con la minima varianza

Negli esempi visti prima lo stimatore della media aritmetica è quello che è consistente, non distorto ed efficiente

Stimatore della Media

Lo stimatore della media aritmetica è non distorto. Infatti

$$E[\bar{x}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

dal momento che : $E[x_i] = \iint x_i f(x_1) f(x_2) \dots f(x_n) dx_1 dx_2 \dots dx_n = \mu$

La varianza di questo stimatore è : $V = \frac{\sigma^2}{n}$

La deviazione standard sul valore medio stimato è $\frac{\sigma}{\sqrt{n}}$

ed è chiamata **errore standard della media**

Stimatore della Varianza

Supponiamo di avere n misure sperimentali x_1, \dots, x_n . Il valore di aspettazione della variabile μ può essere noto prima di aver fatto le misure (cosa rara). In questo caso la varianza è stimata da $\frac{1}{n} \sum (x_i - \mu)^2$

Questo stimatore è consistente e non distorto.

Se il valore di aspettazione non è noto a priori, allora la varianza la calcoliamo

rispetto alla media aritmetica: $V = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \sum (x_i^2 - \bar{x}^2)$

Questo stimatore è distorto (non lo è asintoticamente). Noi possiamo costruire un altro stimatore della varianza non distorto: $\frac{1}{n-1} \sum (x_i - \bar{x})^2$

Questa è la varianza del campione. È una stima consistente e non distorta. Il termine $n-1$ è dovuto al fatto che i dati sono già stati utilizzati per stimare la media aritmetica rispetto alla quale è calcolata la varianza. C'è quindi un grado di libertà in meno.