STATISTICA PER FISICI

1. Calcolo delle probabilità
2a. Statistica frequentista
2b. Statistica bayesiana
3. Likelihood
4. Fondo e segnale
5. Metodi di best fit
6. Metodi Bootstrap
7. Metodi ricorsivi e filtri
8. Unfolding

Alberto Rotondi
Università di Pavia

What is Statistics ?

- ***a problem of probability calculus*: if $p = 1/2$ for having head in tossing a coin, what is the probability to have in 1000 coin tosses less than 450 heads?**

- ***the same problem in statistics*: if in 1000 coin tosses 450 heads have been obtained, what is the estimate of the true head probability?**

Statistical error: $s \approx \sigma$

$$\mu \pm \sigma = 500.0 \pm 15.8 \simeq 500 \pm 16 = [484, 516]$$

$$x \pm s = 450.0 \pm 15.7 \simeq 450 \pm 16 = [434, 466]$$

Physics and Statistics

- **Higgs mass (PDG 2000):**

$$m > 95.3\ GeV, CL = 95\%$$

- W **mass:**

$$m_W = 80.419 \pm 0.056\ GeV$$

These are
confidence intervals

The hystorical path

	FREQUENTISTS	BAYESIANS
1763		Thomas Bayes writes a fundamental paper. Bayesian age
1900	Karl Pearson proposes the χ^2 test	
1910	Robert Fisher invents Maximum Likelihood	
1937	The J. Neyman frequentist interval estimate	
1940	The Hypothesis testing of Pearson. Frequentist age The Popper scheme Frequentist teaching	
1990		rediscovering of the bayesian works of Jeffreys, De Finetti and Jaynes
now	the debate is open: see on Confidence Limits	the CERN Workshop (Geneva 2000) neo-Bayesian age?

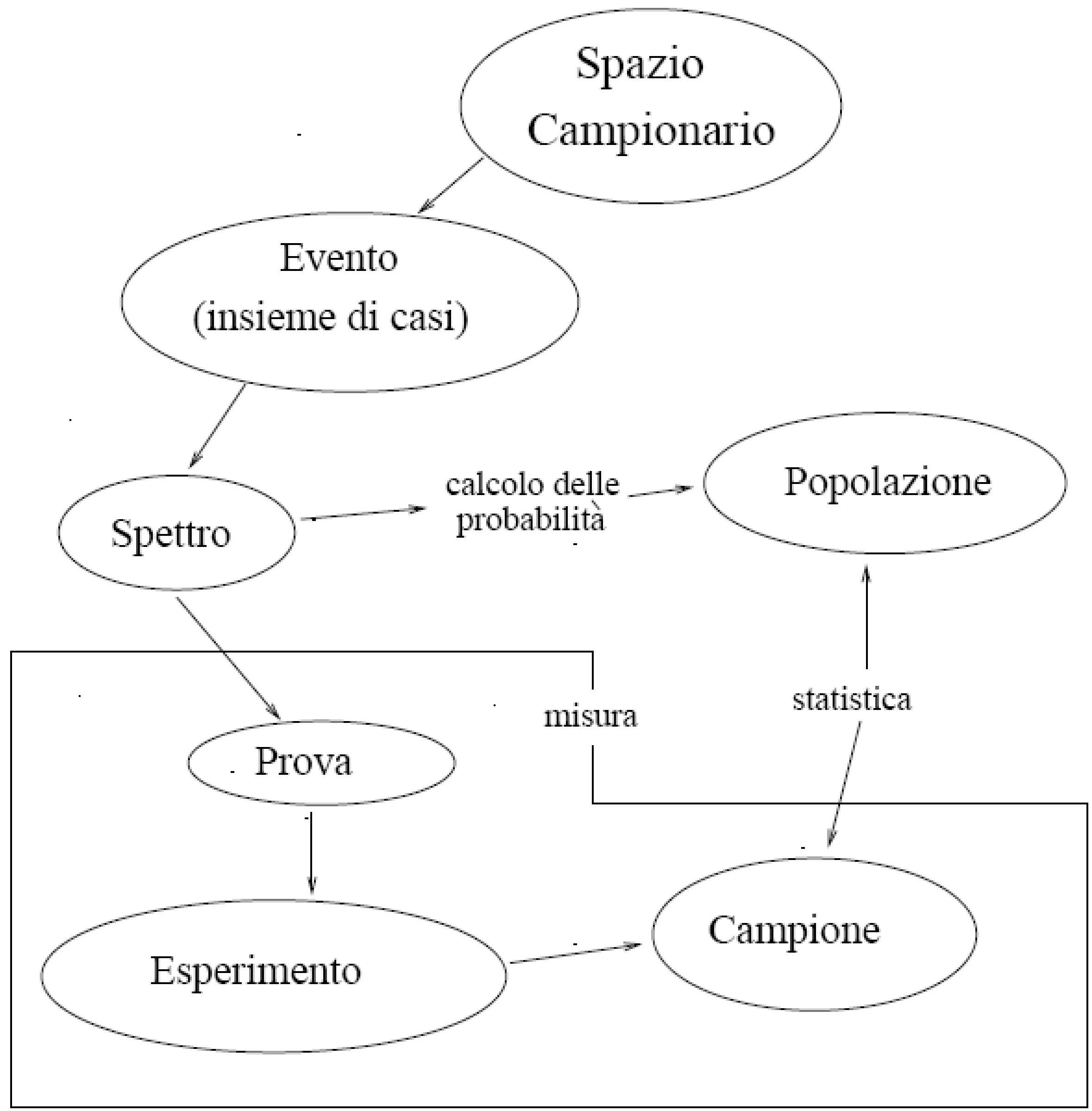
Whorkshop on confidence limits
CERN, 2000

PHYSTAT2003 USA

PHYSTAT2005 Durham

PHYSTAT2007 CERN

James, Cowan, Barlow, Feldman, Cousins, Zech, Roe, Lyons
D'Agostini, Punzi, Giunti
some statisticians


Spazio
Campionario
Evento
(insieme di casi)
Spettro
calcolo delle
probabilità
Popolazione
misura
statistica
Prova
Esperimento
Campione

Il concetto di probabilità

è un atto soggettivo, basato sull'esperienza.

- **Probabilità soggettiva**
  La probabilità è il grado di convinzione *(degree of belief)* soggettivo circa il verificarsi di un evento. La probabilità soggettiva deve essere *coerente* (assiomi di Kolmogorov)

- **Probabilità a priori o classica**
  N è il numero totale di casi, **n** il numero di casi favorevoli per i quali si realizza l'evento A, la probabilità a priori di A è data da:

$$P(A) = \frac{n}{N} \tag{34}$$

- **Probabilità frequentista** **m** è il numero di prove in cui si è verificato l'evento A su un totale di **M** prove, la probabilità di A è data da:

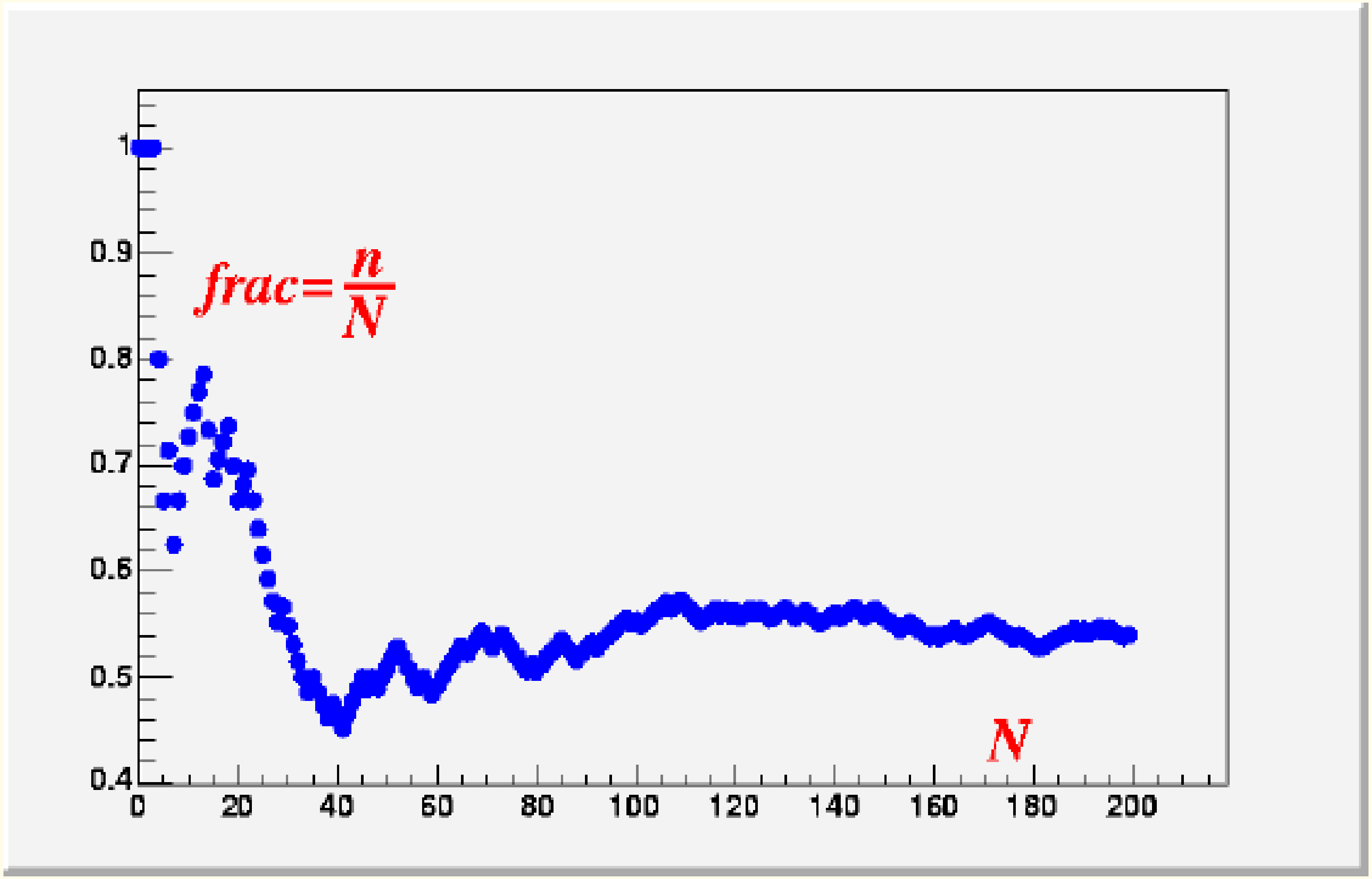
$$P(A) = \lim_{M\to\infty} \frac{m}{M} \ . \tag{35}$$

In molti casi l'esperienza mostra che

$$\lim_{M\to\infty} \frac{m}{M} = \frac{n}{N} \qquad \text{(dall'esperienza !)} \ . \tag{36}$$

legge empirica del caso o dei grandi numeri

Legge empirica del caso



Done during a class: a die is cast several times; success if the outcome is less than 4

Probabilità assiomatica (Kolmogorov)

formalmente la probabilità è una unica entità matematica

Una funzione $P(A) : S \to \mathcal{R}$ **tale che :**

$$P(A) \geq 0 \ ; \tag{39}$$

$$P(S) = 1 \ ; \tag{40}$$

per ogni famiglia finita o numerabile $A_1, A_2, \ldots$ **di insiemi di** $\mathcal{F}$**, tra loro mutuamente disgiunti:**

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \ \ \text{se } A_i \cap A_j = \emptyset \ \ \forall i \neq j \ , \tag{41}$$

viene detta probabilità.

Probabilità composta

Si dice probabilità composta e si indica con

$$P(A \cap B) \quad \text{oppure con} \quad P(AB)$$

la probabilità che si verifichino contemporaneamente gli eventi A *e* **B**

Probabilità condizionata

A: insieme degli assi
B: insieme dei quadri

$$P(A|B) = \frac{\textbf{casi favorevoli all'asso di quadri}}{\textbf{casi favorevoli al seme di quadri}} = \frac{1}{13} = \frac{1}{52}\Big/\frac{13}{52} = \frac{P(A \cap B)}{P(B)} \; .$$

$$P(B|A) = \frac{\textbf{casi favorevoli all'asso di quadri}}{\textbf{casi favorevoli ad un asso}} = \frac{1}{4} = \frac{1}{52}\Big/\frac{4}{52} = \frac{P(B \cap A)}{P(A)} \; .$$

$P(B|A)$ **è la probabilità che si verifichi l'evento B essendosi verificato l'evento A:**

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{se} \quad P(A) > 0 \; .$$

la definizione di probabilità condizionata è in accordo con gli assiomi generali di Kolmogorov

Eventi indipendenti

Due eventi A e B sono detti indipendenti se e solo se

$$P(A \cap B) = P(A)P(B) \ .$$

Segue che per eventi indipendenti $P(A|B) = P(A)$ **e** $P(B|A) = P(B)$**.**

Eventi incompatibili

$$A \cap B = \emptyset \ .$$

Si ha allora:

$$P(A \cap B) = 0 \ , \qquad P(A|B) = P(B|A) = 0 \ .$$

Leggi della somma e del prodotto

- **eventi incompatibili:**

$$P(A \text{ oppure } B) \equiv P(A \cup B) = P(A) + P(B)$$

- **eventi indipendenti:**

$$P(A \text{ e } B) \equiv P(A \cap B) \equiv P(AB) = P(A) \cdot P(B) \ .$$

Partizione dell'evento certo

Se gli insiemi $B_i\,(i=1,2,..n)$ sono mutuamente disgiunti ed esaustivi di S,

$$\bigcup_{i=1}^{n} B_i = S, \qquad B_i \cap B_k = \emptyset \qquad \forall i,k \ , \qquad (44)$$

per ogni insieme $A \in S$:

$$\begin{aligned} P(A) &= P[A \cap (B_1 \cup B_2 \cup \ldots \cup B_n)] \\ &= P[(A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)] \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i) \ . \end{aligned}$$

Esempio **Una malattia H colpisce in un anno il 10% degli uomini e il 5% delle donne. Se la popolazione di $10\,000$ persone è composta dal 45% di uomini e dal 55% di donne, trovare in numero N atteso di ammalati**

$$P(H) = 0.45 \cdot 0.10 + 0.55 \cdot 0.05 = 0.0725 \ .$$

$$N = 10\,000 \cdot 0.0725 = 725 \ .$$

Teorema di Bayes

$$P(B_k|A)P(A) = P(A|B_k)P(B_k)$$

se gli eventi B_k sono disgiunti ed esaustivi di S,

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

e quindi $P(B_k|A)$ si può scrivere come:

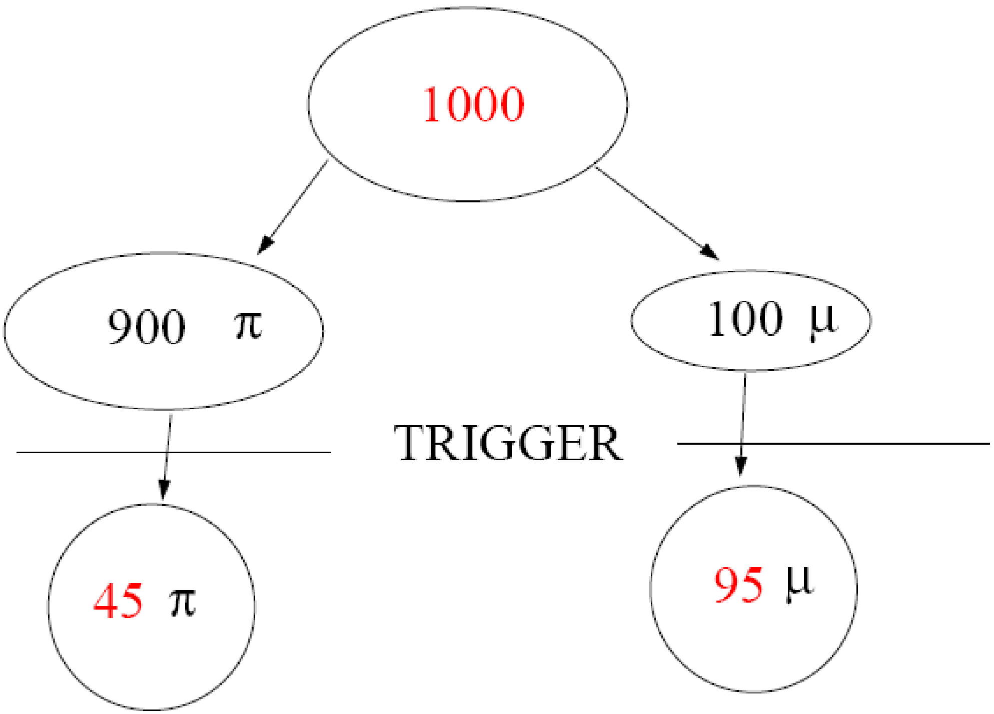
$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \,, \quad P(A) > 0 \,.$$

Esempio il problema del trigger
Un trigger ha $\varepsilon(\pi) = 0.05$ e $\varepsilon(\mu) = 0.95$. Se il fascio ha 90% di π e 10% di μ trovare efficienza ed arricchimento.

$$\begin{aligned} P(\mu|T) &= \frac{P(T|\mu)P(\mu)}{P(T|\mu)P(\mu) + P(T|\pi)P(\pi)} \\ &= \frac{0.95 \cdot 0.10}{0.95 \cdot 0.10 + 0.05 \cdot 0.90} = 0.678 \end{aligned}$$

efficienza $= 14\%$, **arricchimento** $= 68\%$

Teorema di Bayes.
Metodi grafici



$$P(\text{trigger}) = \frac{95 + 45}{1000} = 0.14$$

$$P(\mu | T) = \frac{95}{95 + 45} = 0.678$$

Uso bayesiano della formula di Bayes

Si estende l'utilizzo della formula di Bayes anche quando le probabilità iniziali sono soggettive (ipotesi)

$$P(H_k|\mathrm{dati}) = \frac{P(\mathrm{dati}|H_k)P(H_k)}{\sum\limits_{i=1}^{n} P(\mathrm{dati}|H_i)P(H_i)} \; .$$

il calcolo può essere ripetuto in modo iterativo:

$$P_n(H_k|E) = \frac{P(E_n|H_k)P_{n-1}(H_k)}{\sum\limits_{i=1}^{n} P(E_n|H_i)P_{n-1}(H_i)} \; ,$$

Uso bayesiano della formula di Bayes

probabilità a posteriori $P(H|n = \ldots \bullet)$ a partire da probabilità a priori tutte uguali per estrazioni consecutive di n biglie nere da un'urna contenente 5 biglie bianche e nere

$$P(E|H_1) = 0 \ , P(E|H_2) = 1/5$$

$$P(E|H_3) = 2/5 \ldots P(E|H_6) = 1$$

$$P_0(H_k|E) \equiv P(H_k) = 1/6 \qquad \text{arbitrario !!!}$$

$$P_n(H_k|E) = \frac{P(E|H_k)P_{n-1}(H_k|E)}{\sum_{i=1}^{n} P(E|H_i)P_{n-1}(H_i|E)} \ ,$$

ipotesi **tipo di urna**	H_1 ○○○○○	H_2 ○○○○●	H_3 ○○○●●	H_4 ○○●●●	H_5 ○●●●●	H_6 ●●●●●
$P(H_i)$ **a priori**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**
$P(H_i \mid n = 1 \bullet)$	**0**	**0.07**	**0.13**	**0.20**	**0.27**	**0.33**
$P(H_i \mid n = 5 \bullet)$	**0.**	**0.00**	**0.01**	**0.05**	**0.23**	**0.71**
$P(H_i \mid n = 10 \bullet)$	**0.**	**0.00**	**0.00**	**0.00**	**0.11**	**0.89**

Due strade

- **approccio frequentista:**
  non si assumono probabilità di tipo soggettivo per le ipotesi.
  Non si determinano mai probabilità di ipotesi del tipo

$$P(\mathrm{H}|\mathrm{dati}) \qquad \mathrm{NO!!!!}$$

- **approccio bayesiano:**
  si determinano probabilità del tipo

$$\begin{aligned} P(\mathrm{H}|\mathrm{dati}) &\propto P(\mathrm{dati}|\mathrm{H})P(\mathrm{H}) \\ &\propto \text{verosimiglianza x prob. soggettiva} \\ &\propto \text{likelihood x prior} \end{aligned}$$

che dipendono sia dai dati ottenuti durante le prove sia dalle probabilità di ipotesi iniziali assunte in modo arbitrario.

Il caso di Sally Clark, I

Sally Clark, a British woman, was accused in 1998 of having killed her first child at 11 weeks of age, then conceived another child and allegedly killed him at 8 weeks of age.

The defense claimed that these were two cases of sudden infant death syndrome.

The prosecution had expert witness Sir Roy Meadow testify that the probability of two children in the same family dying from sudden infant death syndrome is about 1 in 73 million.

Based only on this probability statement, Mrs Clark was convicted in 1999.

The Royal Statistical Society, in a press release, pointed out the mistake.

A higher court later quashed Sally Clark's conviction but on other grounds, on 29 January 2003.

- Clark's release in January 2003 prompted the Attorney General to order a review of hundreds of other cases.[1] Two other women convicted of murdering their children, Donna Anthony and Angela Cannings, had their convictions overturned and were released from prison. Trupti Patel, who was also accused of murdering her three children, was acquitted in June 2003. In each case, Roy Meadow had testified about the unlikelihood of multiple cot deaths in a single family.
- Meadow was struck off the medical register by the General Medical Council in 2005 for serious professional misconduct, but he was reinstated in 2006 after he appealed. In June 2005, Alan Williams, the pathologist who conducted the postmortem examinations on both the Clark babies, was banned from Home Office pathology work and coroners' cases for three years after the General Medical Council found him guilty of "serious professional misconduct" in the Clark case.[13] This decision was upheld by the High Court in November 2007.[14]
- According to her family, Clark was unable to recover from the effects of her conviction and imprisonment.[15] After her release, her husband said she would "never be well again".[1] She was unable to read John Batt's book about her case, *Stolen Innocence: A Mother's Fight for Justice*.[1]
- By all accounts a broken woman, Clark was found dead in her home in Hatfield Peverel in Essex on 16 March 2007.[1][3] It was originally thought that she had died of natural causes,[8][16] but an inquest ruled that she had died of acute alcohol intoxication, though the coroner stressed that there was no evidence that she had intended to commit suicide.[3]

SALFORD ADVERTISER THURSDAY JUNE 13, 2002

SALFORD UNI MAN SAYS SALLY CLARK CONVICTION MAY BE WRONG

Maths professor challenges double baby murder case

A SALFORD University Maths professor will challenge evidence used to convict a solicitor of murdering her two baby sons at a conference on cot-deaths next week.

Prof Ray Hill, from Eccles, head of the university's Applied and Discrete Mathematics Research Unit said statistical evidence used to convict Sally Clark, from Wilmslow, in October 2000, was not only quoted out of context and unfairly used to imply guilt, but was actually wrong.

Watching the trial on the TV he became furious and told us: "I shouted at the screen 'that figure's wrong!' They took an estimated figure for the likelihood of one cot death and then just squared it to get this one-in-73 million chance. That's not allowed unless you're sure the events are independent. A bookie wouldn't give you those odds."

He has now studied the Confidential Enquiry into Stillbirths and Deaths in Infancy (CESDI) report, which gives detailed figures on the number of deaths from 1993-1996.

He said: "It seems the chances of two cot deaths in the same family are much higher than the prosecution led the jury to believe."

Prof Hill has written to several national newspapers and is working with Sally Clark's defence team on the campaign to free her.

He will present his full criticism of the evidence at a Developmental Physiology Conference on cot deaths organised by Leicester University on June 28.

The Criminal Cases Review Commission has been looking at the case and is expected to report within the next few weeks. With their report imminent, Sally Clark's defence team and family do not feel it is appropriate to comment.

For more information on the Sally Clark campaign visit www.sallyclark.org.uk

Evidence challenge: Prof Ray Hill

Il caso di Sally Clark, II

The argument can be analyzed using conditional probability. let us call:

- E= the observed evidence;
- I= the accused is innocent;

Hence:

- $P(E|I)$ is the probability of the *evidence* if the accused is innocent; this is the quoted 1/73 millions.
- $P(I|E)$ is the probability that the accused is innocent given the evidence.

The prosecutor wrongly confuses P(I|E) with P(E| I), two rather different concepts. We will need Bayes theorem to compute one from the other.

Il caso di Sally Clark, III

Back to Sally Clark case. The Bayes formula is needed to compute $P(I \mid E)$. We will see it is not so easy... We have called:

E = the observed evidence, I = Sally is innocent

$$P(I \mid E) = \frac{P(E \mid I)P(I)}{P(E \mid I)P(I) + P(E \mid \bar{I})P(\bar{I})}$$

$P(I)$ can be estimated from the frequency of those crimes in the population. In those years England and Walls reported 30 children killed by mother on 640000 births.
$P(I) = 1 - 30/640000 = 1 - 5 \cdot 10^{-5}$. Also $P(E \mid \bar{I}) \approx 1$. Thus:

$$P(I \mid E) \approx \frac{10^{-8} \cdot 1}{10^{-8} \cdot 1 + 1 \cdot 10^{-5}} \approx 10^{-3}$$

Small but not so small.

Variabile aleatoria

Una funzione $X(a)$

$$X : S \to (-\infty, +\infty)\,, \quad \text{ovvero} \quad X(a) = x$$

è una variabile aleatoria quando l'insieme degli elementi a per i quali

$$X(a) \le x$$

è un evento per ogni $x \in \mathcal{R}$.

La variabile aleatoria è definita come una funzione

ESTRAI E LEGGI(biglia) ≤ numero intero

ATTENZIONE!!!

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$g(X) = \frac{1}{\sqrt{2\pi}} e^{-X^2/2}$$

$$-2\ln g(X) = +2\ln\sqrt{2\pi} + X^2 \cong \chi^2$$

Variabili aleatorie indipendenti

Se gli eventi $\{X_i \in A_i\}$, $(i = 1, 2, \ldots, n)$ sono indipendenti per ogni scelta possibile degli intervalli $A_i \in \mathcal{R}_x$, allora

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots,\} = \Pi_i P\{X_i \in A_i\}$$

e le variabili X_i sono stocasticamente indipendenti

Spettro

codominio reale della v.a $X(a)$.

- **discreto: insieme numerabile (v.a. discrete)**
- **continuo: valori su tutto $\mathcal{R}$ o su unioni di intervalli di $\mathcal{R}$ (v.a. continue)**

Funzione cumulativa

$F(x) = P\{X \leq x\}$ in lettere maiuscole

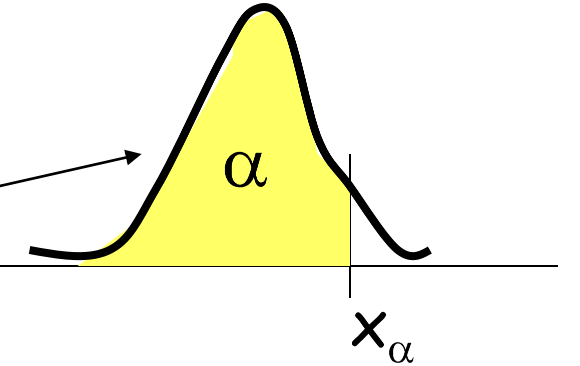
è la probabilità che X assuma un valore non superiore ad un valore assegnato x.

Esempio: lancio di un dado, $X = 1, 2, 3, 4, 5, 6$:

$$F(3.4) = P\{X \leq 3.4\} = P\{X \leq 3\} = F(3)$$

proprietà

- $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$.
  da cui $F(x_2) \geq F(x_1)$.
- $F(x)$ **è continua a destra**



Quantile

$$P\{X \leq x_\alpha\} = F(x_\alpha) \geq \alpha \ .$$

Densità di probabilità

variabili discrete

Data una v.a. X discreta, la funzione $p(x)$, che vale

$$p(x_i) = P\{X = x_i\}$$

per i valori x_i dello spettro di X e $p(x) = 0$ al di fuori di essi, è detta densità di probabilità

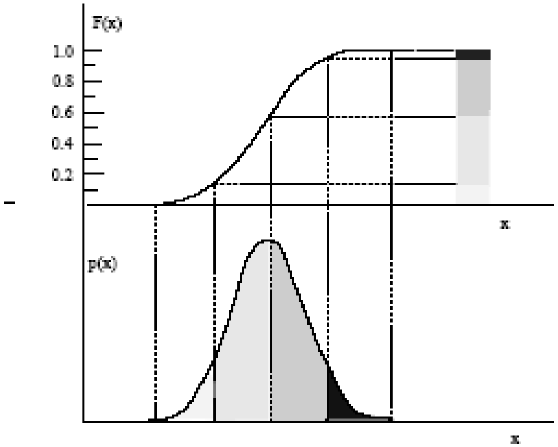
$$\begin{aligned} \sum_i p(x_i) &= \sum_i P\{X = x_i\} \\ &= P\left(\bigcup_i \{X = x_i\}\right) = P(S) = 1 \end{aligned}$$

$$\begin{aligned} P(A) &= P\{X(a) \in \mathcal{R}_0\} \\ &= \sum_{x_i \in \mathcal{R}_0} P\{X = x_i\} = \sum_{x_i \in \mathcal{R}_0} p(x_i) \end{aligned}$$

funzione cumulativa o di ripartizione

$$P\{X \le x_k\} = F(x_k) = \sum_{i=1}^{k} p(x_i) \ .$$

Densità di probabilità



variabili continue

$$p_X(\Delta x_k) \equiv p(x_k)\Delta x_k = P\{x_k \le X \le x_{k+1}\}$$

definita dalla cumulativa

$$F(x) = \int_{-\infty}^{x} p(t)\,\mathrm{d}t \ .$$

$$\begin{aligned} P\{x_k \le X \le x_{k+1}\} &= F(x_{k+1}) - F(x_k) \\ &= \int_{x_k}^{x_{k+1}} p(x)\,\mathrm{d}x \ , \end{aligned}$$

$$\int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1 \ .$$

$$\sum_k p(x_k) \ \to \ \int p(x)\,\mathrm{d}x \ , \qquad p(x_k) \ \to \ p(x_k)\,\mathrm{d}x \ ,$$

Media e Varianza
Istogrammi e Densità

$$m = \sum_{k=1}^{C} f_k x_k \ ,$$

$$\mu = \sum_{k=1}^{\infty} p_k x_k \to \int x p(x)\,\mathrm{d}x \ ,$$

$$s_\mu^2 = \sum_{k=1}^{C} f_k (x_k - \mu)^2 \ ,$$

$$s_m^2 = \frac{N}{N-1} \sum_{k=1}^{C} f_k (x_k - m)^2 \ ,$$

$$\sigma^2 = \sum_{k=1}^{\infty} p_k (x_k - \mu)^2 \to \int (x-\mu)^2 p(x)\,\mathrm{d}x \ ,$$

$$D_n = \sum_{k=1}^{C} f_k (x_k - \mu)^n \ ,$$

$$\Delta_n = \sum_{k=1}^{\infty} p_k (x_k - \mu)^n \to \int (x-\mu)^n p(x)\,\mathrm{d}x \ .$$

Operatori

Valori sperimentali ricavati dai dati:

$$m_x,\ \ m(x),\ \ \langle x\rangle\ ,\ \ \ s_x^2,\ \ s^2(x),\ \ s_x,\ \ s(x)\ ,$$

operatore su una variabile aleatoria:

$$\langle X\rangle\ ,\quad E[X],\qquad \mathrm{Var}[X]\ .$$

risultato di un operatore (valore vero)

$$\langle X\rangle = \mu_x = \mu\ ,\qquad \mathrm{Var}[X] = \sigma_x^2 = \sigma^2$$

$$\sigma[X] = \sigma_x = \sigma$$

Campione Casuale

L'insieme delle N variabili

$$(X_1, X_2, \ldots, X_N)$$

indipendenti ed aventi densità di probabilità $p(x)$ viene detto

campione casuale
di dimensione N

estratto dalla popolazione di densità $p(x)$.

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_N\} = \prod_{i=1}^{N} P\{X_i = x_i\}$$
$$= \prod_{i=1}^{N} p(x_i) \ .$$

primo campione	$x'_1, \ x'_2, \ \ldots, \ x'_N$	$\rightarrow m_1, \ s_1^2$,
secondo campione	$x''_1, \ x''_2, \ \ldots, \ x''_N$	$\rightarrow m_2, \ s_2^2$,
terzo campione	$x'''_1, \ x'''_2, \ \ldots, \ x'''_N$	$\rightarrow m_3, \ s_3^2$,
$\ldots$	$\ldots$	
tutti i campioni	$X_1, \ X_2, \ \ldots, \ X_N$	$\rightarrow M, \ \ S^2$;

Statistica

$$T = t(X_1, X_2, \ldots, X_N)$$

è una v. a. che non contiene alcun parametro incognito

Media e varianza campionarie sono due *statistiche*

Stimatore

è una statistica che converge ad un parametro

$$M = \sum_i \frac{X_i}{N}\ , \qquad S^2 = \sum_i \frac{(X_i - \mu)^2}{N}\ .$$

Media e varianza campionarie sono due *stimatori*

Criteri di convergenza

limite frequentista

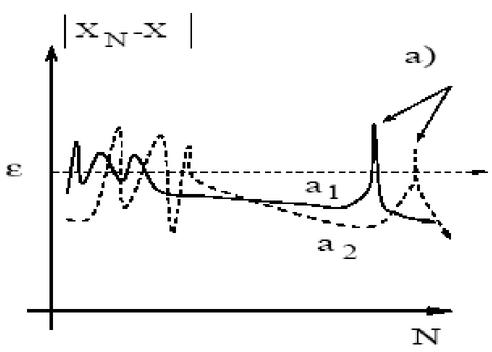
$$lim \frac{1}{N} \sum_k x_k n_k = lim \sum_k n_k f_k = \sum_k x_k p_k = \mu$$

limite in probabilità

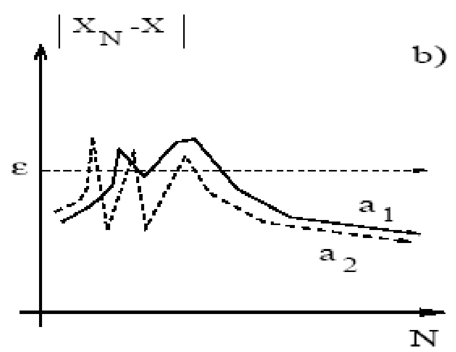
$$\lim_{N\to\infty} P\{|X_N(a) - X(a)| \geq \epsilon\} = 0 \ ,$$
$$\lim_{N\to\infty} P\{|X_N(a) - X(a)| \leq \epsilon\} = 1$$

limite quasi certo

$$P\{\lim_{N\to\infty} |X_N(a) - X(a)| \leq \epsilon\} = 1 \ .$$



limite in probabilità



limite quasi certo

Criteri di convergenza

convergenza in legge

$$\lim_{N\to\infty} F_{X_N}(x) = F_X(x)$$

per ogni punto x in cui $F_X(x)$ è continua.

$$\boxed{\textbf{quasi certa} \to \textbf{in prob.} \to \textbf{in legge}}$$

Kolmogorov: se

$$\Sigma_N \frac{\mathrm{Var}[X_N]}{N^2} < +\infty \ .$$

allora la convergenza è quasi certa. Valida pertutti gli stimatori statistici usuali

Convergenza di stimatori

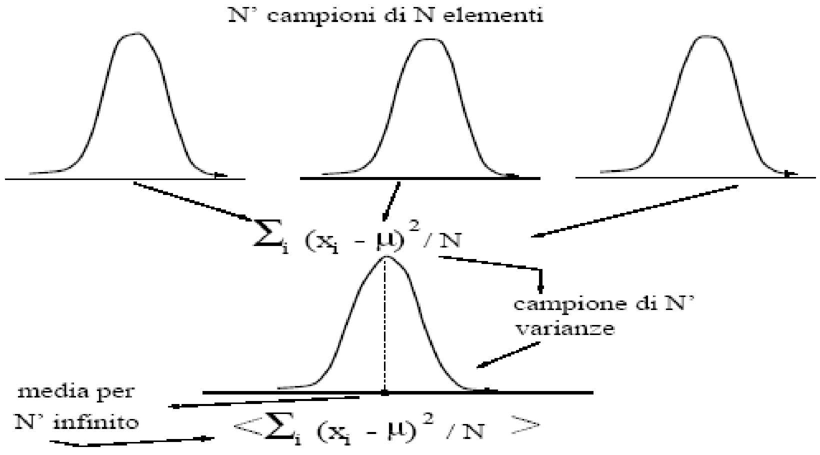
È definita come convergenza in probabilità

stimatore consistente

$$\lim_{N\to\infty} P\left\{|T_N(X)-\mu|>\epsilon\right\}=0$$

media di uno stimatore

$$\langle T_N(X)\rangle \equiv \langle \frac{1}{N}\Sigma_i\,(X_i-\mu)^2\rangle$$



stimatore corretto

$$\langle T_N(X)\rangle = \sigma^2\ .$$

nome	densità	media	deviazione standard	commenti
binomiale	$\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}np$		$\sqrt{np(1-p)}$	**successi in prove indipendenti con probabilità costante**
gaussiana	$\frac{\exp[-(x-\mu)^2/2\sigma^2]}{\sqrt{2\pi}\sigma}$	μ	σ	**combinazione lineare di variabili indipendenti**
chi-quadrato	$\frac{(\chi^2)^{\frac{\nu}{2}-1}\exp(-\frac{\chi^2}{2})}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}$	ν	$\sqrt{2\nu}$	**modulo di un vettore gaussiano**
poissoniana	$\frac{\mu^x}{x!}e^{-\mu}$	μ	$\sqrt{\mu}$	**conteggi**
esponenziale	$\lambda e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$	**tempi tra variabili poissoniane**
gamma	$\frac{\lambda^k}{\Gamma(k)}t^{k-1}e^{-\lambda t}$	$\frac{k}{\lambda}$	$\frac{\sqrt{k}}{\lambda}$	**somma di k variabili esponenziali negative**
uniforme	$\frac{1}{\Delta} \quad (0 \le x \le \Delta)$	$\frac{\Delta}{2}$	$\frac{\Delta}{\sqrt{12}}$	**variabili cumulative**

Teorema Limite Centrale

Sia Y combinazione lineare di N variabili aleatorie X_i

$$Y_N = \sum_{i=1}^{N} a_i X_i$$

dove a_i sono coefficienti costanti

a) **le variabili X_i sono tra loro indipendenti**

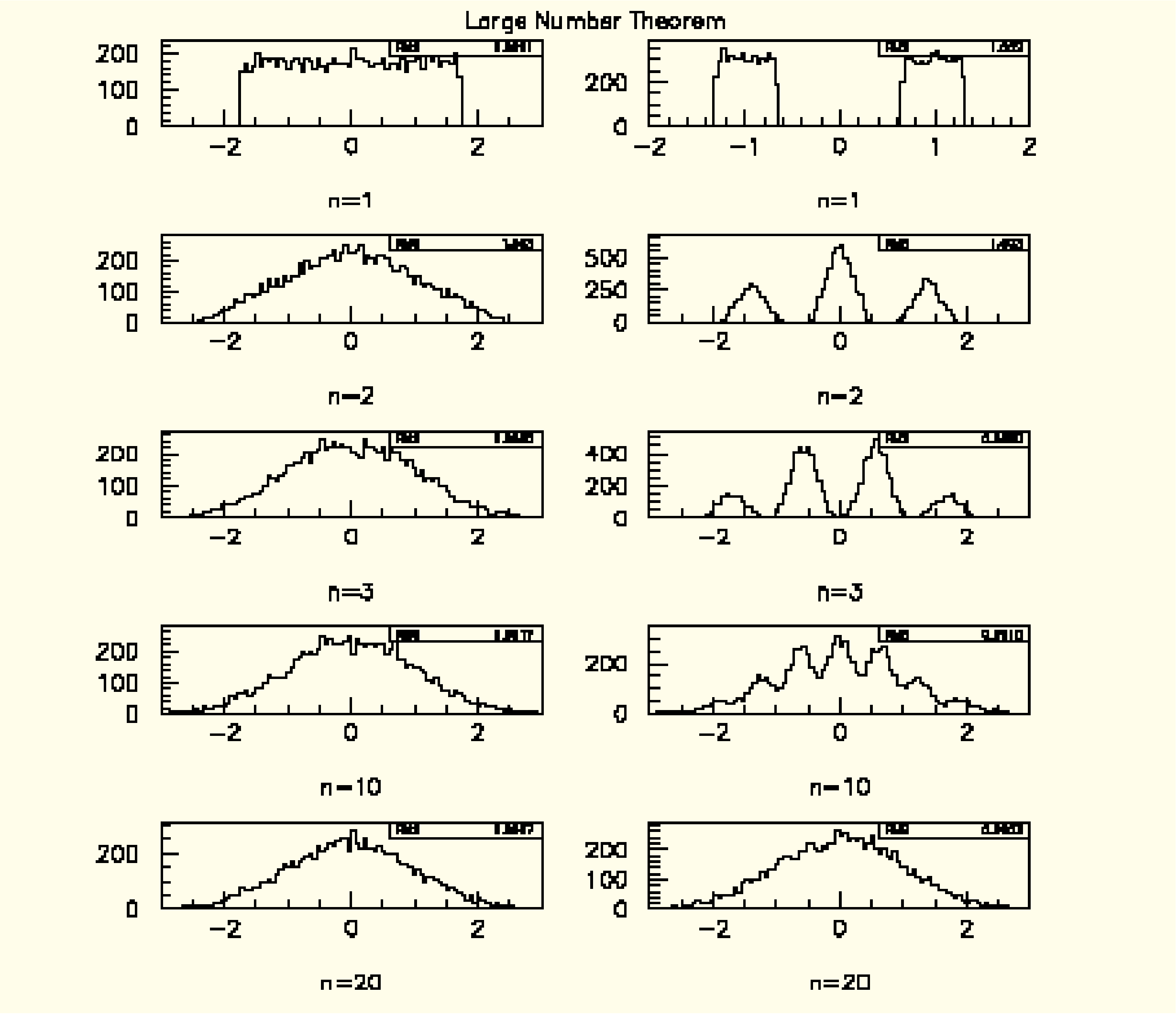
b) **le variabili X_i hanno varianza finita;**

c) **le varianze (o deviazioni standard) sono tutte dello stesso ordine di grandezza:**

$$\frac{\sigma[X_i]}{\sigma[X_j]} = O(1) \quad \text{per ogni} \quad i, j$$

allora, per $N \to \infty$

$$P\left\{\frac{Y_N - \langle Y_N \rangle}{\sigma[Y_N]} \leq x\right\} \xrightarrow{N \gg 1} \Phi(x) \ ,$$


Large Number Theorem
n=1
n=1
n=2
n=2
n=3
n=3
n=10
n=10
n=20
n=20

Processi Poissoniani

processo stocastico in cui una sorgente genera eventi

a) discreti, in modo tale che in un intervallo infinitesimo di tempo $\mathrm{d}t$ si generi *al più* un evento

$$p = \lambda \, \mathrm{d}t \ll 1 \, , \quad n = \Delta t / \mathrm{d}t \gg 1 \, , \quad \mu = np = \lambda \Delta t$$

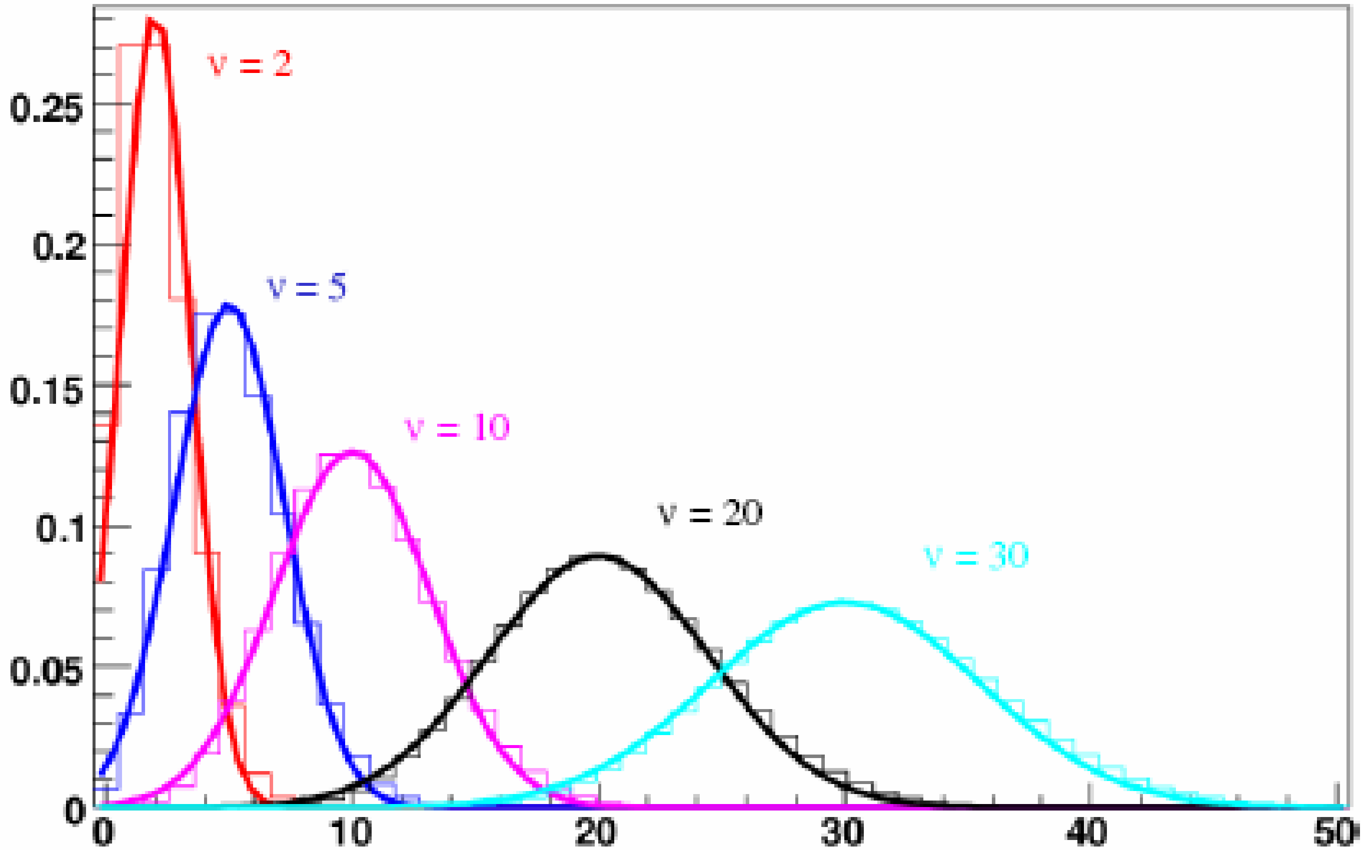
b) con probabilità per unità di tempo λ costante e uguale per tutti gli eventi generati.

c) indipendenti tra loro

Processo Stocastico Poissoniano:
una sorgente *stabile* (λ costante) emette eventi *discreti* ($\lambda \, \mathrm{d}t \ll 1$) ed *indipendenti*.

$$P\{X(\Delta t) = x\} = p_x(\Delta t) = \frac{(\lambda \Delta t)^x}{x!} e^{-\lambda \Delta t} \, .$$

- For large ν a Gaussian approximation is sufficiently accurate



Processi Poissoniani
tempi di arrivo

tempo T tra due eventi (intertempo)

zero eventi fino a t $\times$ un evento in $(t, t + \mathrm{d}t)$

$$p_0(t) = e^{-\lambda t}$$

legge esponenziale negativa:

$$e(t)\,\mathrm{d}t = p_0(t)\lambda\,\mathrm{d}t = \lambda e^{-\lambda t}\,\mathrm{d}t\ , \qquad t \geq 0\ .$$

$$\mu = \int t\,e(t)\,\mathrm{d}t = \frac{1}{\lambda}\,, \qquad \sigma^2 = \int \left(t - \frac{1}{\lambda}\right)^2 e(t)\,\mathrm{d}t = \frac{1}{\lambda^2}\ .$$

nel discreto:

$$\begin{aligned} \ln g(n) &= \ln[p(1-p)^{n-1}] = \ln p + (n-1)\ln(1-p) \\ &\to \ln p - (n-1)p\ , \end{aligned}$$

e quindi:

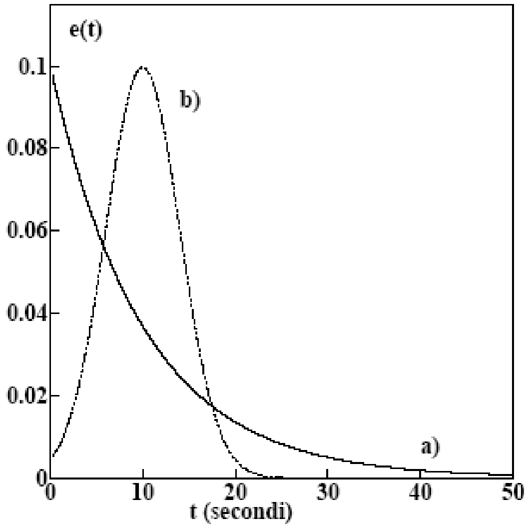
$$g(n) \stackrel{p \ll 1}{\longrightarrow} p e^{-(n-1)p}\ .$$

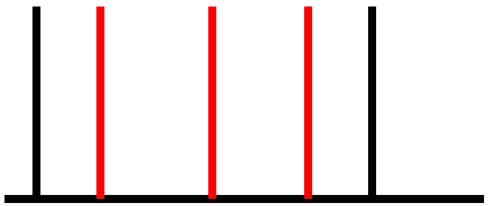
detta legge geometrica.

Processi Poissoniani
legge esponenziale

eventi ravvicinati nel tempo sono molto più probabili di eventi separati da lunghi intervalli di attesa

a) legge esponenziale b) gaussiana di media 10 secondi





e deviazione standard arbitraria di 4 secondi; rappresentazione intuitiva (ma errata!) dei fenomeni stocastici

- effetti nei contatori
- calamità naturali
- incidenti

Poissonian counting
Fundamental theorem

Let's count a Poisson variable N with mean λ with a detector of efficiency ε. The registered number of counts n follows the distribution

$$P(n|N)P(N) = \frac{\mathrm{e}^{-\lambda}\lambda^N}{N!} \; \frac{N!}{n!(N-n)!}\,\varepsilon^n(1-\varepsilon)^{N-n}$$

By using the new variables

$$\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda\varepsilon}\mathrm{e}^{-\lambda(1-\varepsilon)}$$

$$m = N - n$$

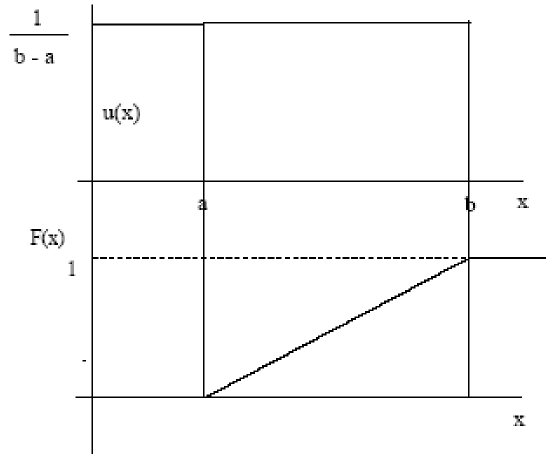
$$\lambda^N = \lambda^{N-n}\lambda^n \equiv \lambda^m\lambda^n$$

one has

$$P(n|N)P(N) = \frac{\mathrm{e}^{-\lambda\varepsilon}(\lambda\varepsilon)^n}{n!} \; \frac{\mathrm{e}^{-\lambda(1-\varepsilon)}\lambda^m(1-\varepsilon)^m}{m!}$$

The number of counts n is still an independent Poisson variable with mean $\lambda\varepsilon$!
(also the lost counts m with mean $\lambda(1-\varepsilon)$)

La densità uniforme



$$u(x) = \begin{cases} \dfrac{1}{b-a} & \textbf{per } \; a \le x \le b \\ 0 & \textbf{per } \; x < a \;, x > b \end{cases}$$

E' facile vedere che:

$$P\{a \le X \le b\} = \int_a^b u(x)\,\mathrm{d}x = 1$$

$$\mu \;=\; \frac{1}{b-a} \int_a^b x \;\mathrm{d}x = \frac{b+a}{2} \;,$$

$$\sigma^2 \;=\; \frac{1}{b-a} \int_a^b \left(x - \frac{b+a}{2}\right)^2 \,\mathrm{d}x = \frac{(b-a)^2}{12}$$

se $b - a = \Delta$

$\mu - \Delta/2 \le x \le \mu + \Delta/2,\ \sigma = \Delta/\sqrt{12}$.

La densità uniforme
teoremi

Teorema 1:

Se $a \leq X \leq b$**, e** $X \sim U(a,b)$

$$P\{x_1 \leq X \leq x_2\} = \frac{1}{b-a}\int_{x_1}^{x_2} \mathrm{d}x = \frac{x_2 - x_1}{b-a}\,. \qquad (1)$$

Viceversa, se vale la (1), allora
$X \sim U(a,b)$

Teorema 2:

Se X **ha densità p(x) continua, la variabile aleatoria cumulativa**

$$\boxed{C(X) = \int_{-\infty}^{X} p(x)\,\mathrm{d}x}$$

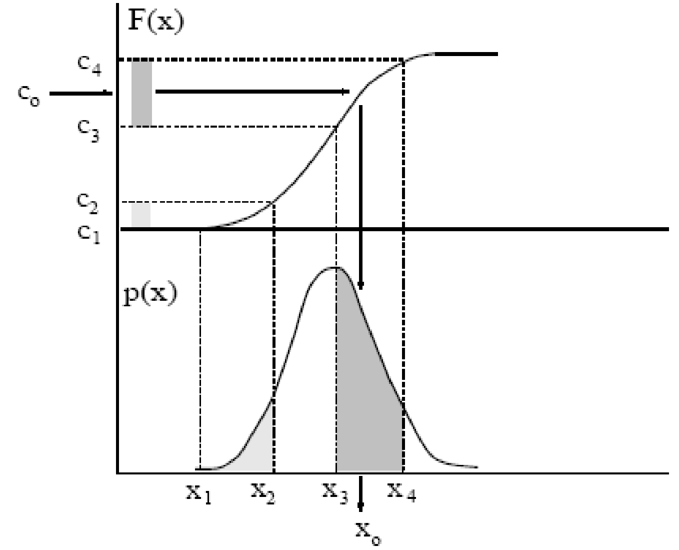
è uniforme in $[0,1]$**, cioè** $C \sim U(0,1)$**.**

Infatti:

$$\begin{aligned} P\{c_1 \leq C \leq c_2\} &= P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} p(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{x_2} p(x)\,\mathrm{d}x - \int_{-\infty}^{x_1} p(x)\,\mathrm{d}x = C_2 - C_1 \end{aligned}$$

da cui $C \sim U(0,1)$**, poiché soddisfa la (1) con** $(b-a) = 1$**.**

La densità uniforme
tecniche di simulazione



la variabile cumulativa
è sempre uniforme, qualunque sia la distribuzione d'origine
Allora se $X \sim p(x)$ **e**

$$X = F^{-1}(C)$$

estraggo C **da una RANDOM** $\sim U(0,1)$
ed ottengo un
campionamento artificiale di X **!!!**

$$X = F^{-1}(\mathtt{RANDOM}) \ .$$

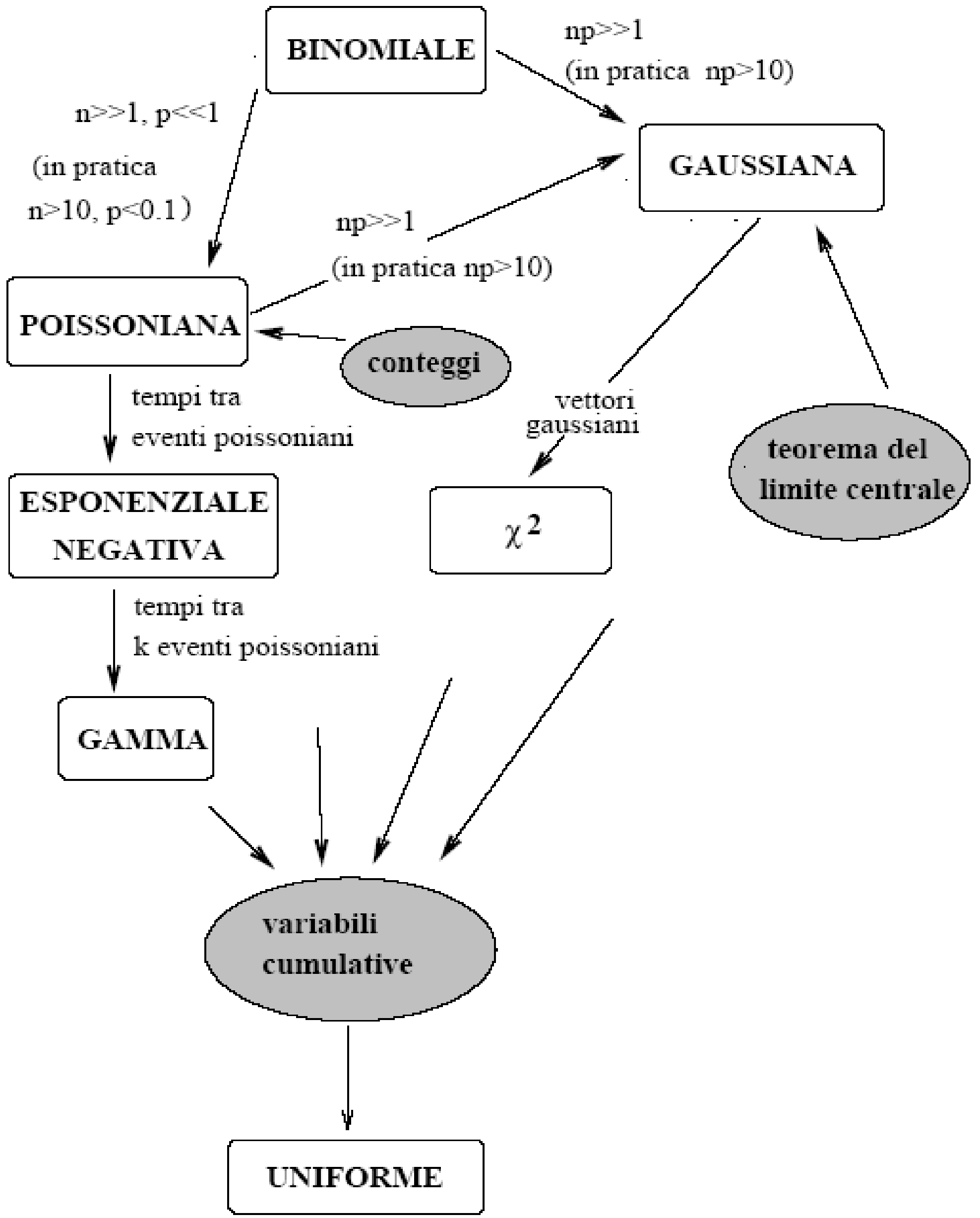
Per variabili discrete: Se C **è una variabile uniforme,**

$$P\{X = x_k\} = P\{x_{k-1} < X \leq x_k\}$$
$$= F(x_k) - F(x_{k-1}) = P\{F(x_{k-1}) < C \leq F(x_k)\} \ .$$

Allora:

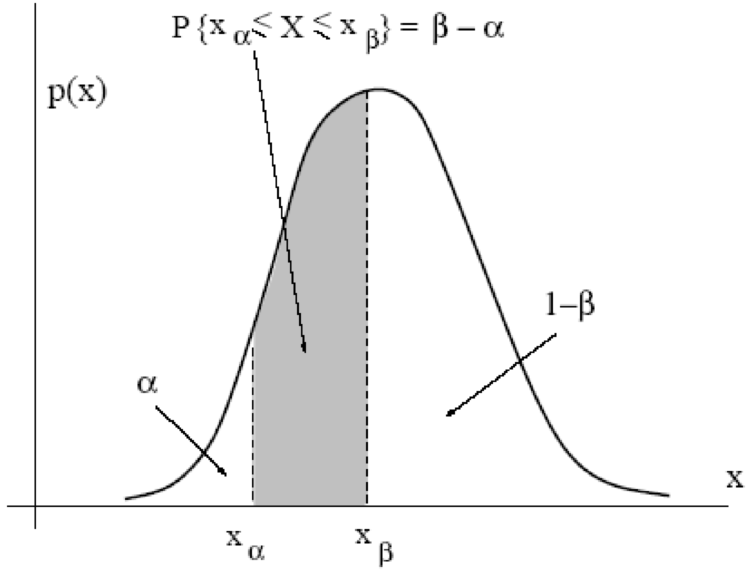
$$\{X = x_k\} \quad \text{se} \quad \{F(x_{k-1}) < \mathtt{RANDOM} \leq F(x_k)\}$$
$$(\text{se } k = 1, \ F(x_0) = 0) \ .$$

Lo schema



BINOMIALE
np>>1
(in pratica np>10)
GAUSSIANA
n>>1, p<<1
(in pratica
n>10, p<0.1)
np>>1
(in pratica np>10)
POISSONIANA
conteggi
tempi tra
eventi poissoniani
vettori
gaussiani
teorema del
limite centrale
ESPONENZIALE
NEGATIVA
χ^2
tempi tra
k eventi poissoniani
GAMMA
variabili
cumulative
UNIFORME

Verifica di ipotesi (modello)



Rappresentazione grafica del livello di probabilità $\beta - \alpha$.

$$P\{x_\alpha \le X \le x_\beta\} = \beta - \alpha \ .$$

Se X ha una densità simmetrica:

$$P\left\{-t_{1-\alpha/2} \le \frac{X-\mu}{\sigma} \le t_{1-\alpha/2}\right\} = 1-\alpha \ .$$

In molti casi t_α sono i quantili della curva normale standard o di Student

$$1-\alpha = 0.90,\ 0.95,\ 0.99,\ 0.999$$

$$\rightarrow \quad t_{1-\alpha/2} = 1.64,\ 1.96,\ 2.58,\ 3.29$$

Calcolo delle probabilità per più variabili

$$P(AB) \equiv P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} \ ,$$

Se esiste una funzione $p(x,y) \geq 0$ tale che

$$P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x,y) \, \mathrm{d}x \, \mathrm{d}y$$

$p(x,y)$ **è la densità congiunta di** X **e** Y

Più in generale, la probabilità che $(X,Y) \in A$ **è data da:**

$$P\{(X,Y) \in A\} = \int_A p(x,y) \, \mathrm{d}x \, \mathrm{d}y \ ,$$

Calcolo delle probabilità per più variabili

$$\int\int p(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1 \ ,$$

$$\begin{aligned}
\langle X\rangle &= \int\int x\,p(x,y)\,\mathrm{d}x\,\mathrm{d}y = \mu_x \ ,\\
\langle Y\rangle &= \int\int y\,p(x,y)\,\mathrm{d}x\,\mathrm{d}y = \mu_y \ ,\\
\mathrm{Var}[X] &= \int\int (x-\mu_x)^2\,p(x,y)\,\mathrm{d}x\,\mathrm{d}y = \sigma_x^2 \ ,\\
\mathrm{Var}[Y] &= \int\int (y-\mu_y)^2\,p(x,y)\,\mathrm{d}x\,\mathrm{d}y = \sigma_y^2 \ ,
\end{aligned}$$

Se X e Y sono stocasticamente indipendenti:

$$P\{x_1 \le X \le x_2, y_1 \le Y \le y_2\} =$$

$$P\{x_1 \le X \le x_2\}P\{y_1 \le Y \le y_2\} \ ,$$

ovvero:

$$\int_{x_1}^{x_2}\int_{y_1}^{y_2} p(x,y)\ \mathrm{d}x\,\mathrm{d}y = \int_{x_1}^{x_2} p_X(x)\,\mathrm{d}x \int_{y_1}^{y_2} p_Y(y)\ \mathrm{d}y$$

in questo caso la densità congiunta vale:

$$p(x,y) = p_X(x)\,p_Y(y)$$

(se X e Y sono indipendenti)

Quantità nuove

Densità marginale

$P\{X \in A\} \equiv \int_A p_X(x)\,\mathrm{d}x = \int_A \mathrm{d}x \int_{-\infty}^{+\infty} p(x,y)\;\mathrm{d}y$,

da cui:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x,y)\;\mathrm{d}y \ .$$

Teorema:
Due variabili aleatorie (X,Y), di densità congiunta $p(x,y)$, sono stocasticamente indipendenti se e solo se

$$p(x,y) = g(x)\,h(y) \ .$$

Densità condizionata

$$p(y|x_0) = \frac{p(x_0,y)}{p_X(x_0)} = \frac{p(x_0,y)}{\int_{-\infty}^{+\infty} p(x_0,y)\;\mathrm{d}y} \ .$$

Le densità condizionate sono normalizzate:

$$\int p(y|x_0)\,\mathrm{d}y = \frac{\int p(x_0,y)\,\mathrm{d}y}{\int p(x_0,y)\;\mathrm{d}y} = 1 \ .$$

Densità marginale e condizionata

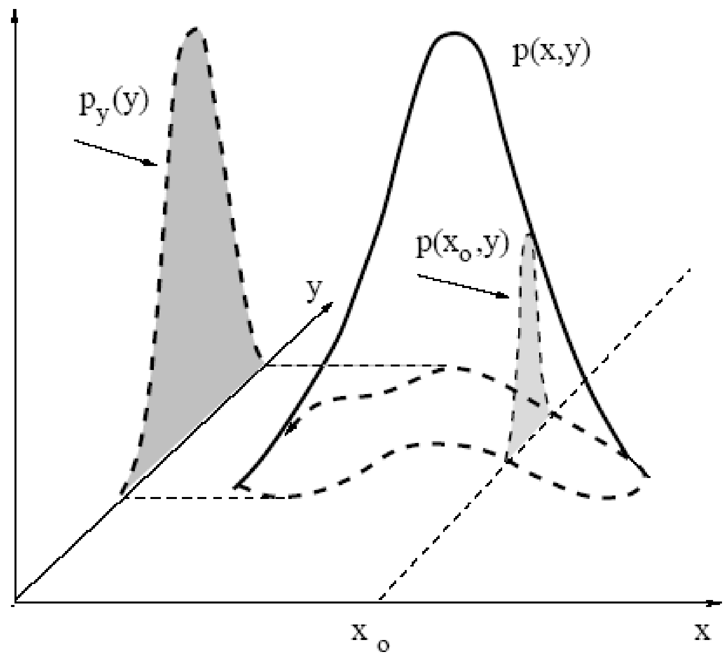
Probabilità composte:

$$p(x,y) = p_Y(y)\,p(x|y) = p_X(x)\,p(y|x)$$

Quando le variabili sono indipendenti:

$$p(x|y) = p_X(x)\ , \quad p(y|x) = p_Y(y)\ ,$$

densità marginale $p_Y(y)$ di y e funzione $p(x_0, y)$,



che normalizzata rappresenta la densità condizionata di Y in corrispondenza del valore x_0.

Teorema di Bayes

$$p(x,y) = p_Y(y)\, p(x|y) = p_X(x)\, p(y|x)$$

da cui

$$p(x|y) = \frac{p(y|x)\, p_X(x)}{p_Y(y)}$$

ovvero

$$p(x|y) = \frac{p(y|x)\, p_X(x)}{\int p(y|x)\, p_X(x)\, \mathrm{d}x}$$

passaggio bayesiano:

$$p(\mu;x) = \frac{p(x;\mu)\, p_\mu(\mu)}{\int p(x;\mu)\, p_\mu(\mu)\, \mathrm{d}x}$$

ovvero

$$p(\mu;x) = \frac{\text{likelihood} \times \text{prior}}{\text{normalization}}$$

Covarianza

$$\begin{aligned}\mathrm{Var}[X+Y] &= \int [(x-\mu_x)+(y-\mu_y)]^2 p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\ \mathrm{Cov}[X,Y]\ ,\end{aligned}$$

dove

$$\mathrm{Cov}[X,Y] = \int\int (x-\mu_x)(y-\mu_y)\,p(x,y)\,\mathrm{d}x\,\mathrm{d}y \equiv \sigma_{xy}\ .$$

è la covarianza.
Per variabili discrete:

$$\mathrm{Cov}[X,Y] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)p_{XY}(x,y)\ ,$$

Per un insieme di dati sperimentali:

$$s_{xy} = \frac{1}{N-1}\sum_i (x_i - m_x)(y_i - m_y)\ ,$$

Come operatore:

$$\mathrm{Cov}[X,Y] = \langle (X-\mu_x)(Y-\mu_y)\rangle$$

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \langle (X-\mu_x)(Y-\mu_y)\rangle \\ &= \langle XY\rangle - \mu_x\langle Y\rangle - \mu_y\langle X\rangle + \mu_x\mu_y \\ &= \langle XY\rangle - \mu_x\mu_y\ ,\end{aligned}$$

Covarianza e correlazione

Per variabili indipendenti:

$$\begin{aligned}
\mathrm{Cov}[X,Y] &= \int\int (x-\mu_x)(y-\mu_y)\, p(x,y)\, \mathrm{d}x\, \mathrm{d}y \\
&= \int (x-\mu_x)\, p_X(x)\, \mathrm{d}x \int (y-\mu_y)\, p_Y(y)\, \mathrm{d}y \\
&= [\int x\, p_X(x)\, \mathrm{d}x - \mu_x \int p_X(x)\, \mathrm{d}x] \\
&\quad\; [\int y\, p_Y(y)\, \mathrm{d}y - \mu_y \int p_Y(y)\, \mathrm{d}y] \\
&= \mu_x - \mu_x + \mu_y - \mu_y = 0
\end{aligned}$$

Se X_i sono tra loro indipendenti:

$$\mathrm{Var}\left[\sum_i X_i\right] = \sum_i \mathrm{Var}[X_i] \ .$$

Teorema di Cauchy-Schwartz:

$$|\,\mathrm{Cov}[X,Y]| \leq \sigma[X]\,\sigma[Y]$$

Coefficiente di correlazione:

$$\rho_{xy} \equiv \rho[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sigma[X]\,\sigma[Y]} \ .$$

gode della proprietà

$$-1 \leq \rho_{xy} \leq 1 \ ,$$

Covarianza e correlazione

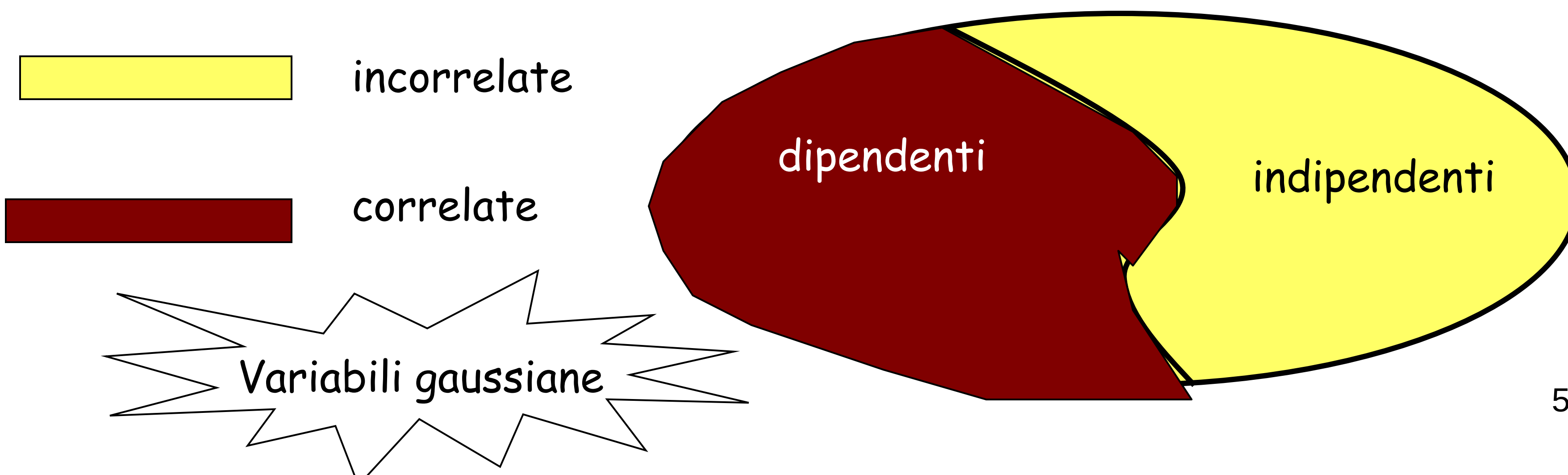
Variabili dipendenti come

$X = U + V$ **e** $Y = U - V, \quad U, V \sim U(0,1)$

sono dipendenti, ma...

$$
\begin{aligned}
\mathrm{Cov}[X,Y] &= \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \\
&= \langle (U + V - \langle U + V \rangle)(U - V - \langle U - V \rangle) \rangle \\
&= \langle (U + V - 1)(U - V) \rangle = \langle U^2 - V^2 - U + V \rangle \\
&= \langle U^2 \rangle - \langle V^2 \rangle - \langle U \rangle + \langle V \rangle = 0 \ ,
\end{aligned}
$$

sono incorrelate!!



Densità gaussiana bidimensionale

Si ottiene, se $\rho \neq \pm 1$:

$$g(u, v; 0, 1) = \frac{1}{2\pi\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] \; .$$

Per variabili non standard:

$$g(x, y) \; = \; \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}\gamma(x,y)} \; ,$$

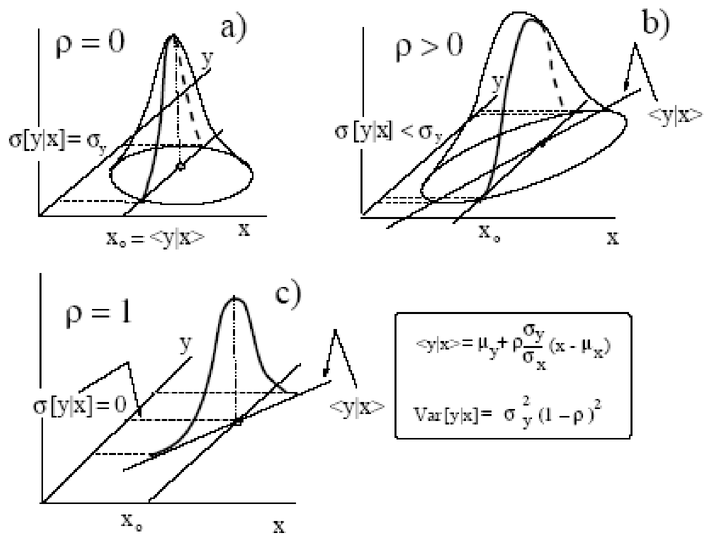
$$\gamma(x, y) \; = \; \frac{1}{1-\rho^2} \times \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]$$

Indipendenza di variabili gaussiane

Condizione necessaria e sufficiente affinché due variabili aleatorie congiuntamente gaussiane (normali) siano indipendenti, è che il loro coefficiente di correlazione lineare sia nullo.

Densità gaussiana bidimensionale Ellissi di concentrazione

Intersezione con un piano:



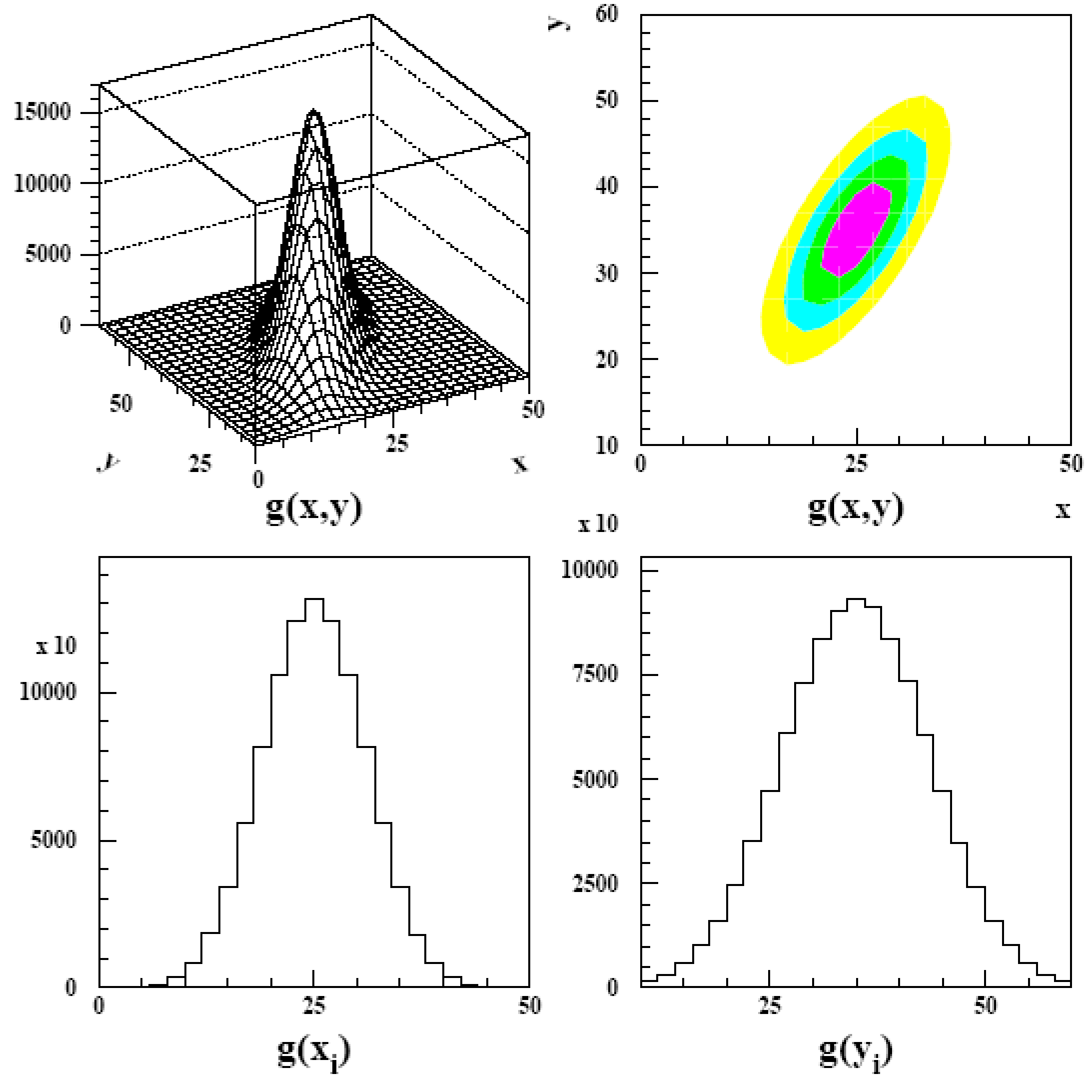
$$u^2 - 2\rho uv + v^2 = \textbf{costante} \ ,$$

ovvero

$$\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} = \textbf{costante}$$

Se $\rho = \pm 1$,

$$\frac{(x-\mu_x)}{\sigma_x} \pm \frac{(y-\mu_y)}{\sigma_y} = \textbf{costante} \ .$$


15000
10000
5000
0
50
y
25
0
25
50
x
g(x,y)
y
60
50
40
30
20
10
0
25
50
g(x,y)
x
x 10
10000
5000
0
0
25
50
g(x_i)
x 10
10000
7500
5000
2500
0
25
50
g(y_i)

Gaussiana multivariata:

$$g(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}}|V|^{-1/2} \times$$
$$\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\dagger V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$
$$|V| \equiv \det|V| \neq 0$$

Gradi di libertà

$$H\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\mu}$$

L'esponente della gaussiana si trasforma in:

$$\gamma(\boldsymbol{X}) \equiv (\boldsymbol{X}-\boldsymbol{\mu})^\dagger V^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) = \boldsymbol{Z}^\dagger H^\dagger V^{-1} H \boldsymbol{Z}$$
$$= \boldsymbol{Z}^\dagger \boldsymbol{Z} = \sum_{i=1}^{n} Z_i^2$$

La densità chi quadrato (χ^2)

È la distribuzione di Q, dove X sono n variabili gaussiane indipendenti:

$$Q = \sum_{i=1}^{n} X_i^2$$

$$p_n(q)\, \mathrm{d}q = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{1}{2}q} q^{\frac{1}{2}(n-2)} \mathrm{d}q \ , \qquad q \geq 0 \ ,$$

dove $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x}\, \mathrm{d}x$
notazione dei fisici:

$$p_\nu(\chi^2)\, \mathrm{d}\chi^2 \equiv p(\chi^2;\nu)\, \mathrm{d}\chi^2 = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}\, \mathrm{d}\chi^2 \ ,$$

$$\langle Q \rangle \ = \ \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \int_0^\infty x(x)^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}\, \mathrm{d}x = \nu$$

$$\mathrm{Var}[Q] \ = \ \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \int_0^\infty (x-\mu)^2\, (x)^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}\, \mathrm{d}x = 2\nu$$

La densità chi quadrato (χ^2)

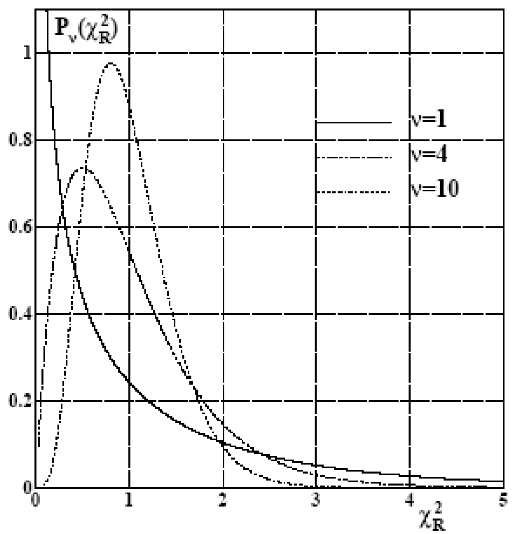
Spesso si usa la distribuzione χ^2 ridotto della variabile

$$Q_R(\nu) = \frac{Q(\nu)}{\nu} \ ,$$

$$P\{Q_R(\nu) \geq \chi^2_R(\nu)\} = \int_{\chi^2_R(\nu)}^{\infty} \frac{\nu^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \, x^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu x}{2}\right) \, dx$$

$$\langle Q_R(\nu) \rangle = 1 \ , \qquad \mathrm{Var}[Q_R(\nu)] = \frac{2}{\nu}$$

Densità χ^2 ridotto per vari



gradi di libertà.

χ^2 Degrees of Freedom

If

$$H\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\mu}$$

the Q variable in general is

$$Q \equiv (\boldsymbol{X} - \boldsymbol{\mu})^\dagger V^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{Z}^\dagger H^\dagger V^{-1} H \boldsymbol{Z}$$
$$= \boldsymbol{Z}^\dagger \, \boldsymbol{Z} = \sum_{i=1}^{n} Z_i^2$$

Therefore, $Q \sim \chi^2(n)$.

When $|V| = 0$, the general theorem on the quadratic forms says that it is possible to find $n - p$ new independent variables such as

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n-p} T_i^2$$

In conclusion, when

$$Q = (\boldsymbol{X} - \boldsymbol{\mu})^\dagger W \, (\boldsymbol{X} - \boldsymbol{\mu}) \ ,$$

one must verify if $|W| = 0$. In this case one has to find new variable such as

$$(\boldsymbol{X} - \boldsymbol{\mu})^\dagger W \, (\boldsymbol{X} - \boldsymbol{\mu}) = \sum_{i=1}^{n-p} T_i^2 \ .$$

The important principle is that these new variables have not to be found, to reduce the DoF is enough! $\nu = n - p$ (points - equations)