Dottorato di Ricerca in Fisica

XIV Ciclo (Anno Accademico 1998—99)

Statistica Applicata alla Fisica

Benvenuti

"Where shall I begin, please, your Majesty?" he asked.
"Begin at the beginning," the King said, gravely,
"and go on till you come to the end: then stop."
(Alice in Wonderland)

"The time has come," the Walrus said,
"to talk of many things."
(Through the Looking Glass)

Richiami di teoria della probabilità - I

- Fenomeni casuali; spazio dei risultati; spazio degli eventi. Evento impossibile; evento certo. Frequenza assoluta e relativa in N prove.
- Somma logica (unione) e prodotto logico (intersezione) di eventi casuali. Eventi condizionati.
- Definizione assiomatica di probabilità:

$$
\begin{aligned}
&\Pr(E) \geq 0 \\
&\Pr(S) = 1 \\
&\Pr(E_1 \cup E_2 \cdots) = \Pr(E_1) + \Pr(E_2) + \cdots \\
&\qquad (\forall E_i, E_k \ : \ E_i \cap E_k = \emptyset)
\end{aligned}
$$

- Definizione empirica di probabilità:

$$\Pr(E) = \lim_{N \to \infty} f(E)$$

- Definizione di limite statistico, $\lim_{N \to \infty} x = X$:

$$\forall \epsilon, \delta > 0 \quad \longrightarrow \quad \exists M :$$

$$N > M \ \Rightarrow \ \Pr\Big(|x - X| \geq \epsilon \Big) \leq \delta$$

Richiami di teoria della probabilità - II

- Legge della probabilità totale:

$$\Pr(E \cup F) = \Pr(E) + \Pr(F) - \Pr(E \cap F)$$

- Legge della probabilità composta:

$$\begin{aligned}\Pr(E \cap F) &= \Pr(E)\,\Pr(F|E) \\ &= \Pr(F)\,\Pr(E|F)\end{aligned}$$

- Indipendenza statistica di uno o più eventi casuali
- Teorema di Bayes:

$$\Pr(A_i|E) = \frac{\Pr(A_i) \cdot \Pr(E|A_i)}{\sum_j \left[\Pr(A_j) \cdot \Pr(E|A_j)\right]}$$

Scuola di Dottorato in Fisica

Esercizio

Lanciamo un dado per due volte, e consideriamo i seguenti eventi casuali:

- A, consistente nell'ottenere un punteggio dispari sul primo dado;
- B, consistente nell'ottenere un punteggio dispari sul secondo dado;
- C, consistente nell'ottenere un punteggio complessivo dispari.

Quali di questi eventi sono tra loro statisticamente indipendenti?

Soluzione

- A è indipendente da B;
- A è indipendente da C:

$$\Pr(C|A) = \Pr\left(\overline{B}\right) = \frac{1}{2} = \Pr(C)$$

- B è indipendente da C (per la stessa ragione);
- I tre eventi non sono indipendenti (Il realizzarsi di A e di B rende impossibile il realizzarsi di C).

Scuola di Dottorato in Fisica

Esercizio

Lanciamo una moneta per tre volte, e consideriamo i seguenti eventi casuali:

- A: consistente nell'ottenere testa al primo lancio;
- B: consistente nell'ottenere testa al secondo lancio;
- C: consistente nell'ottenere solo due teste, che siano consecutive.

Quali di questi eventi sono tra loro statisticamente indipendenti?

Soluzione

$$S \equiv \{TTT, TTC, TCT, CTT, \\ TCC, CCT, CTC, CCC\}$$

$$\begin{cases} A \equiv \{TTT, TTC, TCT, TCC\} \\ B \equiv \{TTT, TTC, CTT, CTC\} \\ C \equiv \{TTC, CTT\} \end{cases}$$

$$\begin{cases} \Pr(A) = \frac{1}{2} \\ \Pr(B) = \frac{1}{2} \\ \Pr(C) = \frac{1}{4} \end{cases}$$

$$\begin{cases} A \cap B \equiv \{TTT, TTC\} \\ A \cap C \equiv \{TTC\} \\ B \cap C \equiv \{TTC, CTT\} \end{cases}$$

$$\begin{cases} \Pr(A \cap B) = \frac{1}{4} = \Pr(A) \cdot \Pr(B) \\ \Pr(A \cap C) = \frac{1}{8} = \Pr(A) \cdot \Pr(C) \\ \Pr(B \cap C) = \frac{1}{4} \neq \Pr(B) \cdot \Pr(C) \end{cases}$$

A e B sono indipendenti; A e C sono indipendenti; ma B e C non sono indipendenti.

Esercizio

Due fisici esaminano indipendentemente un film che registra il passaggio di un fascio di particelle attraverso una camera a bolle, alla ricerca di decadimenti rari: il primo ne trova $n_A = 95$ ed il secondo $n_B = 98$. Se gli eventi trovati da entrambi sono $n_{AB} = 94$, si chiede (supponendo di poter approssimare ovunque le probabilità con le frequenze rispettive):

1. Il numero più probabile N di eventi;
2. L'efficienza complessiva ε, ossia la probabilità che tra i due osservatori almeno uno trovi un evento.

Soluzione

$$\begin{cases} \Pr(A) = \dfrac{n_A}{N} \\ \Pr(B) = \dfrac{n_B}{N} \end{cases}$$

$$\Pr(A \cap B) = \frac{n_{AB}}{N} = \Pr(A) \cdot \Pr(B) = \frac{n_A}{N} \cdot \frac{n_B}{N}$$

$$N = \frac{n_A \cdot n_B}{n_{AB}}$$

$$\begin{aligned} \varepsilon &\equiv \Pr(A \cup B) \\ &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\ &= \frac{n_A + n_B - n_{AB}}{N} \\ &= \frac{(n_A + n_B - n_{AB}) \cdot n_{AB}}{n_A \cdot n_B} \end{aligned}$$

Scuola di Dottorato in Fisica

Esercizio

Si progetta un esperimento per lo studio di eventi (segnale) individuati da un rivelatore che reagisce anche ad altri eventi (di fondo).

Si ammettono note la probabilità del segnale, quella del fondo, e l'efficienza del rivelatore (sia per il segnale che per il fondo); quale è la probabilità che un *evento osservato* appartenga al segnale?

Soluzione

Sono note Pr(S), Pr(F), Pr(R|S) e Pr(R|F); si chiede Pr(S|R), e si può usare il teorema di Bayes:

$$\Pr(S|R) = \frac{\Pr(R|S) \cdot \Pr(S)}{\Pr(R|S) \cdot \Pr(S) + \Pr(R|F) \cdot \Pr(F)}$$

Richiami di teoria della probabilità - III

- Variabili casuali, continue e discrete
- Probabilità e densità di probabilità:

Variabile discreta	Variabile continua
$p_i \geq 0$	$f(x) \geq 0$
$\sum_i p_i = 1$	$\int_{-\infty}^{+\infty} f(x)\,dx = 1$
	$\Pr(x \in [x_0, x_0 + dx]) = f(x_0)\,dx$
$\Pr(x_1 \leq x \leq x_2) = \sum_{x \in [x_1, x_2]} p_i$	$\Pr(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x)\,dx$

- Funzione di distribuzione: $F(x) = \sum_{x_i \leq x} p_i = \int_{-\infty}^{x} f(t)\,dt$

Stime di tendenza centrale e di dispersione

- Campioni e popolazione
- Valor medio e varianza di un campione:

$$\bar{x} = \frac{1}{N}\sum_i x_i = \sum_k f_k\, x_k$$

$$s^2 = \frac{1}{N}\sum_i \left[x_i - \bar{x}\right]^2 = \sum_k f_k \left[x_k - \bar{x}\right]^2$$

$$= \frac{1}{N}\sum_i x_i^2 - \left(\bar{x}\right)^2$$

- Speranza matematica della variabile casuale x:

$$E(x) = \sum_k p_k\, x_k$$

$$= \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

- Varianza della popolazione:

$$\mathrm{Var}(x) = E\left\{\left[x - E(x)\right]^2\right\}$$

$$\mathrm{Var}(x) = \sum_k p_k \left[x_k - E(x)\right]^2$$

$$= \int_{-\infty}^{+\infty} \left[x - E(x)\right]^2 f(x)\, dx$$

Scuola di Dottorato in Fisica

Speranza matematica delle combinazioni lineari di variabili casuali

$$z = ax + by$$

$$E(z) = \sum_{ik} P_{ik}(ax_i + by_k)$$

$$= a\sum_i x_i \sum_k P_{ik} + b\sum_k y_k \sum_i P_{ik}$$

$$= aE(x) + bE(y)$$

che si generalizza immediatamente per induzione completa a

$$z = \sum_i a_i\, x_i \qquad \Rightarrow \qquad E(z) = \sum_i a_i\, E(x_i)$$

Varianza delle combinazioni lineari di variabili casuali indipendenti

$$
\begin{aligned}
z &= ax + by \\
P_{ik} &= p_i q_k \\
E(x) = E(y) &= 0 \quad \Rightarrow \quad E(z) = 0 \\
\mathrm{Var}(z) &= \sum\nolimits_{ik} P_{ik}(ax_i + by_k)^2 \\
&= \sum\nolimits_{ik} p_i q_k (a^2 x_i^2 + b^2 y_k^2 + 2ab x_i y_k) \\
&= a^2 \sum\nolimits_i p_i x_i^2 \sum\nolimits_k q_k + \\
&\qquad + b^2 \sum\nolimits_k q_k y_k^2 \sum\nolimits_i p_i + \\
&\qquad + 2ab \sum\nolimits_i p_i x_i \sum\nolimits_k q_k y_k \\
&= a^2 \mathrm{Var}(x) + b^2 \mathrm{Var}(y)
\end{aligned}
$$

Che si generalizza prima a due variabili casuali indipendenti aventi speranza matematica non nulla e poi alla somma di un numero qualsiasi di variabili casuali tutte indipendenti:

$$
z = \sum\nolimits_i a_i x_i \qquad \Rightarrow \qquad \mathrm{Var}(z) = \sum\nolimits_i a_i^2 \, \mathrm{Var}(x_i)
$$

Scuola di Dottorato in Fisica

Richiami di teoria della probabilità - IV

$$
\Pr\Big(|x - E(x)| \ge \epsilon\Big) = \sum_{|x_i - E(x)| \ge \epsilon} p_i
$$

$$
\begin{aligned}
\mathrm{Var}(x) &= E\left\{\Big[x - E(x)\Big]^2\right\} \\
&= \sum\nolimits_i p_i \Big[x_i - E(x)\Big]^2 \\
\mathrm{Var}(x) &\ge \sum_{|x_i - E(x)| \ge \epsilon} p_i \Big[x_i - E(x)\Big]^2 \\
&\ge \sum_{|x_i - E(x)| \ge \epsilon} p_i \, \epsilon^2 \; = \; \epsilon^2 \sum_{|x_i - E(x)| \ge \epsilon} p_i
\end{aligned}
$$

Da qui si ottiene la diseguaglianza di Bienaymé–Čebyšef: ponendo $\mathrm{Var}(x) = \sigma^2$,

$$
\Pr\Big(|x - E(x)| \ge \epsilon\Big) \le \frac{\sigma^2}{\epsilon^2}
$$

e, ponendo $\epsilon = k\sigma$,

$$
\Pr\Big(|x - E(x)| \ge k\sigma\Big) \le \frac{1}{k^2}
$$

Altre stime di tendenza centrale

- Moda ($\hat{x}$): $f(\hat{x}) \geq f(x)$
- Mediana ($\tilde{x}$): $F(\tilde{x}) = 0.5$
- Media geometrica (g): $g^N = \prod_{i=1}^{N} x_i$
- Media armonica (h): $\frac{1}{h} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{x_i}$
- Media quadratica (q): $q = \sqrt{\overline{x^2}}$
- Quantili:
  - ⋆ Quartili: $Q_1 \cdots Q_3$
  - ⋆ Decili: $D_1 \cdots D_9$
  - ⋆ Percentili: $P_1 \cdots P_{99}$

Altre stime di dispersione

- Range (R): $R = x_{\text{max}} - x_{\text{min}}$
- Semidispersione (r): $r = \frac{R}{2}$
- Errore medio (ϵ): $\epsilon = E\Big[\,|x - E(x)|\,\Big]$
- Intervallo $Q_3 - Q_1$
- Intervallo $P_{90} - P_{10}$

I momenti - I

- Momenti rispetto all'origine:

$$\begin{aligned}\lambda_k &= E\left(x^k\right)\\ &= \sum_i p_i\, x_i^k\\ &= \int_{-\infty}^{+\infty} x^k f(x)\, dx\end{aligned}$$

- Momenti rispetto alla media:

$$\begin{aligned}\mu_k &= E\left\{\left[x - E(x)\right]^k\right\}\\ &= \sum_i p_i \left[x_i - E(x)\right]^k\\ &= \int_{-\infty}^{+\infty} \left[x - E(x)\right]^k f(x)\, dx\end{aligned}$$

- **I momenti possono non esistere** per popolazioni composte da un numero infinito di individui.
- Momenti rispetto all'origine e rispetto alla media per un campione: definiti analogamente, usando f_i al posto delle p_i e $\bar{x}$ al posto di $E(x)$.

Scuola di Dottorato in Fisica

I momenti - II

- Corollari:
  - ⋆ $\mu_1 = E\left[x - E(x)\right] \equiv 0$
  - ⋆ Per distribuzioni simmetriche rispetto alla media: $\mu_{2k+1} \equiv 0$
  - ⋆ $\mu_2 = \sigma^2 = \lambda_2 - (\lambda_1)^2$
- Coefficiente di asimmetria (*skewness*): $\sigma = \dfrac{\mu_3}{\sigma^3}$
- Coefficiente di curtòsi (*kurtosis*): $\gamma = \dfrac{\mu_4}{\sigma^4} - 3$

La funzione caratteristica - I

- Funzione caratteristica:

$$\phi_x(t) = E\left(e^{itx}\right) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx$$

(o trasformata di Fourier della f).

- La funzione caratteristica esiste sempre.
- $f(x) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{+\infty} e^{-ixt}\, \phi_x(t)\, dt$
(trasformazione inversa di Fourier).
- Solo se tutti i momenti esistono,

$$\phi_x(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\, \lambda_k$$

$$\left.\frac{d^k \phi_x(t)}{dt^k}\right|_{t=0} = (i)^k\, \lambda_k$$

La funzione caratteristica - II

- Per la somma di N variabili casuali tutte statisticamente indipendenti:

$$y = \sum_{k=1}^{N} x_k$$

$$f(y)\, dy = \prod_{k=1}^{N} f_k(x_k)\, dx_k$$

$$\phi_y(t) = \int_{-\infty}^{+\infty} e^{ity} f(y)\, dy = \int_{-\infty}^{+\infty} \prod_{k=1}^{N} e^{itx_k} f_k(x_k)\, dx_k = \prod_{k=1}^{N} \phi_k(t)$$

- Per una trasformazione lineare $y = ax + b$:

$$\phi_y(t) = E\left(e^{ity}\right) = E\left[e^{it(ax+b)}\right] = e^{itb} E\left(e^{itax}\right) = e^{itb} \phi_x(at)$$

La funzione generatrice dei momenti

- Funzione generatrice dei momenti:

$$M_x(t) = E\left(e^{tx}\right) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx$$

- La funzione generatrice esiste solo se **tutti** i momenti esistono. In questo caso risulta anche

$$M_x(t) = \sum_{k=0}^{\infty} \frac{(t)^k}{k!}\,\lambda_k$$

$$\left.\frac{d^k M_x(t)}{dt^k}\right|_{t=0} = \lambda_k$$

- Per una trasformazione lineare $y = ax + b$:

$$M_y(t) = e^{tb} M_x(at)$$

- La funzione generatrice della somma di più variabili casuali indipendenti è uguale al prodotto delle loro funzioni generatrici.
- Funzione caratteristica, funzione generatrice e momenti sono *caratterizzanti*.

Scuola di Dottorato in Fisica

La funzione caratteristica delle variabili casuali discrete

Per una variabile discreta x introduciamo la nuova variabile complessa $z = e^{it}$: la funzione caratteristica è

$$\phi_x(t) = \sum\nolimits_k p_k\, e^{itx_k}$$

e rispetto alla z si può scrivere

$$\phi_x(z) = \sum\nolimits_k p_k\, z^{x_k} = E\left(z^x\right)$$

Per la somma di più variabili indipendenti, le $\phi(z)$ si moltiplicano:

$$t = x + y$$

$$\phi_t(z) = \sum\nolimits_{jk} \Pr(x_j) \Pr(y_k)\, z^{x_j + y_k} = \sum\nolimits_j \Pr(x_j)\, z^{x_j} \cdot \sum\nolimits_k \Pr(y_k)\, z^{y_k} = \phi_x(z) \cdot \phi_y(z)$$

e similmente per la somma di più di due variabili indipendenti:

$$S = \sum\nolimits_k x_k \qquad \Rightarrow \qquad \phi_S(z) = \prod\nolimits_k \phi_{x_k}(z)$$

Somma di un *numero casuale di variabili casuali* (discrete e indipendenti) - I

Per la somma di N variabili (con N prefissato) provenienti dalla stessa popolazione:

$$S = \sum_{j=1}^{N} x_j \qquad e \qquad \phi_S(z) = \left[\phi_x(z)\right]^N$$

Se anche N è una variabile casuale discreta con funzione caratteristica ϕ_N,

$$\phi_N(z) = \sum\nolimits_N \Pr(N)\, z^N$$

risulterà

$$\begin{aligned}
\Pr(S) &= \sum\nolimits_N \Pr(N) \Pr(S|N) \\
\phi_S(z) &= \sum\nolimits_S \Pr(S) \cdot z^S \\
&= \sum\nolimits_S z^S \cdot \sum\nolimits_N \Pr(N) \Pr(S|N) \\
&= \sum\nolimits_N \Pr(N) \cdot \sum\nolimits_S \Pr(S|N)\, z^S \\
&= \sum\nolimits_N \Pr(N) \left[\phi_x(z)\right]^N
\end{aligned}$$

Somma di un *numero casuale di variabili casuali* (discrete e indipendenti) - II

In definitiva:

$$\phi_S(z) = \phi_N\left[\phi_x(z)\right]$$

Se N non è propriamente una variabile casuale e può assumere un unico valore

$$N = N_O$$

si ricade nel caso precedente:

$$\begin{aligned}
\phi_N(z) &= z^{N_O} \\
\phi_S(z) &= \phi_N\left[\phi_x(z)\right] = \left[\phi_x(z)\right]^{N_O}
\end{aligned}$$

Cambiamento di variabile casuale - I

- Se la corrispondenza tra x e y è biunivoca:
  - ⋆ $y = y(x)$ è strettamente monotona
  - ⋆ Esiste la funzione inversa $x = x(y)$
  - ⋆ Se $y(x)$ è derivabile lo è anche $x(y)$; risulta $y'(x) \neq 0$ e

$$x'(y) = \frac{1}{y'[x(y)]} \qquad (\neq 0)$$

- L'invarianza della probabilità su un intervallo infinitesimo, ovvero la condizione $f(x)\,dx \equiv g(y)\,dy$, implica

$$dx \longrightarrow dy = |y'(x)|\,dx$$

$$\begin{aligned} g(y) &= \frac{f(x)}{|y'(x)|} \\ &= \frac{f[x(y)]}{|y'[x(y)]|} \\ &= f[x(y)] \cdot |x'(y)| \end{aligned}$$

Scuola di Dottorato in Fisica

Cambiamento di variabile casuale - II

- Se la corrispondenza non è biunivoca bisogna vedere caso per caso. Se ad esempio

$$y = x^2 \qquad \Rightarrow \qquad x = \sqrt{y}$$

un particolare valore per la y corrisponde a due eventualità esclusive per la x; perciò

$$g(y)\,dy = \left[f(-\sqrt{y}) + f(\sqrt{y})\right]\,dx$$

$$\begin{aligned} g(y) &= \left[f\left(-\sqrt{y}\right) + f\left(\sqrt{y}\right)\right] \cdot |x'(y)| \\ &= \frac{f\left(-\sqrt{y}\right) + f\left(\sqrt{y}\right)}{2\sqrt{y}} \end{aligned}$$

Cambiamento di variabile casuale - III

Consideriamo il cambiamento di variabile casuale

$$x \longrightarrow y = F(x) = \int_{-\infty}^{x} f(t)\, dt \qquad (y \in [0,1])$$

da cui

$$y'(x) = F'(x) = f(x)$$

La corrispondenza è biunivoca se risulta ovunque

$$f(x) > 0$$

In tal caso

$$y'(x) > 0$$

$$g(y) = \frac{f(x)}{|y'(x)|} \equiv 1$$

ed abbiamo così dimostrato che la funzione di distribuzione di qualsiasi variabile casuale che possa assumere tutti i valori di un intervallo continuo è a sua volta distribuita uniformemente in $[0,1]$.

Variabili casuali bidimensionali - I

- Funzione di frequenza congiunta:

$$f(x,y)$$

- Funzione di distribuzione congiunta:

$$F(x,y) = \int_{-\infty}^{x} du \int_{-\infty}^{y} dv\, f(u,v)$$

- Funzioni di frequenza marginali:

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

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

- Funzioni di distribuzione marginali:

$$G(x) = \int_{-\infty}^{x} g(t)\, dt = F(x, +\infty)$$

$$H(y) = \int_{-\infty}^{y} h(t)\, dt = F(+\infty, y)$$

- Condizione di normalizzazione:

$$F(+\infty, +\infty) = 1$$

Variabili casuali bidimensionali - II

Per definizione:

$$f(x,y)\,dx\,dy = g(x)\,dx \cdot \pi(y|x)\,dy$$
$$= h(y)\,dy \cdot \pi(x|y)\,dx$$

- Probabilità condizionate:

$$\pi(y|x) = \frac{f(x,y)}{g(x)}$$
$$\pi(x|y) = \frac{f(x,y)}{h(y)}$$

- Indipendenza statistica:

$$\pi(x|y) \equiv g(x)$$
$$\pi(y|x) \equiv h(y)$$
$$f(x,y) = g(x) \cdot h(y)$$

Due variabili casuali continue sono statisticamente indipendenti se (e solo se) la loro densità di probabilità congiunta è fattorizzabile.

Esempio: decadimento debole della Λ^0

$$\Lambda^0 \longrightarrow p + \pi^- \qquad e \qquad \Lambda^0 \longrightarrow n + \pi^0$$

Il processo può essere descritto da due variabili casuali: c (carica del nucleone) e t (tempo di vita della Λ^0).

Dalla teoria e dagli esperimenti: la legge di decadimento è la stessa per qualunque stato finale (*esponenziale* con la stessa vita media τ); *ed il* branching ratio è indipendente dal tempo di vita, *e* vale

$$\frac{\Pr\left(\Lambda^0 \longrightarrow p + \pi^-\right)}{\Pr\left(\Lambda^0 \longrightarrow n + \pi^0\right)} = 2$$

Quindi:

$$g(1) = g(1|t) = \frac{2}{3}$$
$$g(0) = g(0|t) = \frac{1}{3}$$
$$g(c) = g(c|t) = \frac{c+1}{3}$$
$$h(t) = h(t|0) = h(t|1) = \frac{1}{\tau}\,e^{-\frac{t}{\tau}}$$
$$f(c,t) = g(c)\cdot h(t) = \frac{c+1}{3\tau}\,e^{-\frac{t}{\tau}}$$

Esempio: decadimento debole K^0_{e3}

$$K^0 \longrightarrow e^- + \pi^+ + \bar{\nu}_e \qquad e \qquad K^0 \longrightarrow e^+ + \pi^- + \nu_e$$

Variabili casuali: s (∓ 1, segno dell'elettrone) e t (tempo di vita). *Dalla teoria*: la funzione di frequenza, nell'ipotesi $\Delta Q = \Delta S$ ed una volta definito

$$N(t,s) = e^{-\lambda_1 t} + e^{-\lambda_2 t} + 2\, s \cos(\omega t)\, e^{-\frac{\lambda_1+\lambda_2}{2} t}$$

(λ_1 e λ_2 sono gli inversi delle vite medie del K^0_1 e del K^0_2, e ω è la loro differenza di massa) è

$$f(t,s) = N(t,s) \Big/ \left\{ \sum_s \int_0^{+\infty} N(t,s)\, dt \right\}$$

(supposto di osservare tutti i decadimenti a qualsiasi tempo di vita con efficienza 1): e non è fattorizzabile.

$$h(t) = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \left(e^{-\lambda_1 t} + e^{-\lambda_2 t} \right)$$

$$g(s) = \frac{1}{2} + \frac{4\, s\, \lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2 + 4\omega^2}$$

$$h(t|s) = \frac{f(t,s)}{g(s)}$$

$$g(s|t) = \frac{f(t,s)}{h(t)}$$

Variabili casuali bidimensionali - III

- Valor medio (o speranza matematica) di una funzione $\psi(x,y)$:

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

- Momenti rispetto all'origine:

$$\lambda_{mn} = E\left[x^m y^n\right]$$

- Momenti rispetto alla media:

$$\mu_{mn} = E\left[(x - \lambda_{10})^m (y - \lambda_{01})^n\right]$$

Variabili casuali bidimensionali - IV

- Risulta:
  - ⋆ $\lambda_{00} \equiv 1$
  - ⋆ $\lambda_{10} = E(x)$
  - ⋆ $\lambda_{01} = E(y)$
  - ⋆ $\mu_{20} = E\left\{\left[x - E(x)\right]^2\right\} = Var(x)$
  - ⋆ $\mu_{02} = E\left\{\left[y - E(y)\right]^2\right\} = Var(y)$
  - ⋆ $\mu_{11} = E\left\{\left[x - E(x)\right]\left[y - E(y)\right]\right\}$
- μ_{11} si chiama anche covarianza di x e y, e si indica col simbolo $Cov(x, y)$.

Scuola di Dottorato in Fisica

Variabili casuali bidimensionali - V

- Funzione caratteristica:

$$\phi_{xy}(u, v) = E\left[e^{i(ux+vy)}\right]$$

$$\left.\frac{\partial^{m+n}\phi_{xy}}{\partial u^m\, \partial v^n}\right|_{\substack{u=0\\v=0}} = (i)^{m+n}\lambda_{mn}$$

- Funzione generatrice:

$$M_{xy}(u, v) = E\left[e^{(ux+vy)}\right]$$

$$\left.\frac{\partial^{m+n}M_{xy}}{\partial u^m\, \partial v^n}\right|_{\substack{u=0\\v=0}} = \lambda_{mn}$$

Variabili casuali bidimensionali - VI

- Per un cambiamento di variabili casuali:

$$u = u(x, y) \qquad e \qquad v = v(x, y)$$

- ⋆ Se la corrispondenza è biunivoca, esistono le funzioni inverse:

$$x = x(u, v) \qquad e \qquad y = y(u, v)$$

- ⋆ Se esistono le derivate parziali prime della x e della y rispetto alla u ed alla v, esiste anche non nullo il determinante Jacobiano

$$\frac{\partial(x, y)}{\partial(u, v)} = \det \begin{Vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{Vmatrix}$$

dotato della proprietà che

$$\frac{\partial(x, y)}{\partial(u, v)} = \left[\frac{\partial(u, v)}{\partial(x, y)} \right]^{-1}$$

Variabili casuali bidimensionali - VII

In tal caso, dalla invarianza della probabilità per il cambiamento di variabili

$$f(x, y)\,dx\,dy = g(u, v)\,du\,dv$$

si ottiene

$$g(u, v) = f\left[x(u, v), y(u, v)\right] \cdot \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$$

Covarianza e correlazione - I

$$\begin{aligned}\mathrm{Cov}(x,y) &= E\Big\{[x-E(x)]\,[y-E(y)]\Big\}\\ &= E(xy) - E(x)\,E(y)\end{aligned}$$

- Condizione necessaria perchè due variabili siano statisticamente indipendenti è che la loro covarianza sia nulla; in tal caso infatti

$$E(xy) = E(x)E(y)$$

- La condizione non è sufficiente: se ad esempio

$$y = x^2$$

ne risulta

$$\mathrm{Cov}(x,y) = E(x^3) - E(x)\,E(x^2)$$

nulla per tutte le distribuzioni simmetriche anche se le variabili sono correlate.

Scuola di Dottorato in Fisica

Covarianza e correlazione - II

Per una combinazione lineare $z = ax + by$ di variabili casuali qualsiasi, per la covarianza vale la

$$\sigma_z^2 = a^2\,\sigma_x^2 + b^2\,\sigma_y^2 + 2\,a\,b\;\mathrm{Cov}(x,y)$$

Infatti, per valori medi nulli ($E(x) = E(y) = 0$),

$$\begin{aligned}E(z) &= 0\\ \sigma_z^2 &= E(z^2)\\ &= \textstyle\sum_{ik} P_{ik}\left(a^2x_i^2 + b^2y_k^2 + 2\,abx_iy_k\right)\\ &= a^2E(x^2) + b^2E(y^2) + 2\,abE(xy)\\ &= a^2\sigma_x^2 + b^2\sigma_y^2 + 2\,ab\,\mathrm{Cov}(x,y)\end{aligned}$$

che si generalizza immediatamente alla combinazione lineare di due variabili a valor medio qualunque.

Covarianza e correlazione - III

- Il coefficiente di correlazione lineare è:

$$r = r_{xy} = \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y}$$

  - ⋆ Adimensionale
  - ⋆ Compreso tra -1 e $+1$; infatti, definendo $z = \sigma_y x \pm \sigma_x y$ si ricava:

$$\begin{aligned} \mathrm{Var}(z) &= \sigma_y^2 \sigma_x^2 + \sigma_x^2 \sigma_y^2 \pm 2\sigma_x\sigma_y \,\mathrm{Cov}(x,y) \\ &= 2\left[\sigma_x^2 \sigma_y^2 \pm \sigma_x\sigma_y \,\mathrm{Cov}(x,y)\right] \\ &\geq 0 \end{aligned}$$

    da cui la tesi.

- Per le combinazioni lineari, $z = ax + by$:

$$\mathrm{Var}(ax+by) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2\,a\,b\,r\,\sigma_x\sigma_y$$

Scuola di Dottorato in Fisica

Covarianza e correlazione - IV

Il coefficiente di correlazione lineare vale ± 1 se le due variabili sono legate da una relazione lineare:

$$\begin{aligned} y &= ax + b \\ E(y) &= a\,E(x) + b \\ \mathrm{Var}(y) &= a^2\,\mathrm{Var}(x) \\ E(xy) &= E(ax^2 + bx) \\ &= a\,E(x^2) + b\,E(x) \\ \mathrm{Cov}(x,y) &= E(xy) - E(x)\,E(y) \\ &= a\left\{E(x^2) - \left[E(x)\right]^2\right\} \\ &= a\;\mathrm{Var}(x) \\ r &= \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\;\mathrm{Var}(y)}} \\ &= \frac{a}{|a|} = \pm 1 \end{aligned}$$

- Non è necessariamente vero l'inverso; punti su una retta parallela all'asse x hanno r indefinito ($r = \frac{0}{0}$) pur essendo perfettamente allineati.
- Se le variabili sono legate da una funzione non lineare, non raggiunge i valori estremi ± 1

Il caso *N*-dimensionale - I

- Densità di probabilità congiunta

$$f(\underline{X})$$

- Funzione di distribuzione

$$F(\underline{X}) = \int_{-\infty}^{x_1} dx_1 \cdots \int_{-\infty}^{x_N} dx_N \, f(\underline{X})$$

- Funzioni marginali
- Densità di probabilità condizionate
- Indipendenza statistica di gruppi di variabili casuali
- Funzione caratteristica

$$\phi_{\underline{X}}(\underline{T}) = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_N \, f(\underline{X}) \, e^{i\underline{T}\cdot\underline{X}}$$

- Cambiamento di variabili:

$$x_i \longrightarrow y_k(x_1, \ldots, x_N)$$

$$g(\underline{Y}) = f\Big(x_1(\underline{Y}), \ldots, x_N(\underline{Y})\Big) \left| \frac{\partial \underline{X}}{\partial \underline{Y}} \right|$$

Il caso *N*-dimensionale - II

Per combinazioni lineari di variabili casuali

$$y = \sum_{k=1}^{N} a_i \, x_i$$

risulta:

$$E(y) = \sum_{k=1}^{N} a_i E(x_i)$$

$$\mathrm{Var}(y) = \sum_{k=1}^{N} a_i^2 \, \mathrm{Var}(x_i) + \sum_{\substack{i,k \\ i \neq k}} a_i a_k \, \mathrm{Cov}(x_i, x_k)$$

$$= \underline{A}^T \underset{\sim}{V} \underline{A}$$

dove $\underset{\sim}{V}$ *è la matrice delle covarianze, di elementi*

$$V_{ij} = E\Big[(x_i - \mu_i) \, (x_j - \mu_j) \Big]$$

($\underline{\mu}$ *è il vettore dei valori medi delle* x_i); $\underset{\sim}{V}$ *è simmetrica,*
e inoltre

$$V_{ii} = \mathrm{Var}(x_i) \qquad e \qquad V_{ij} = \mathrm{Cov}(x_i, x_j)$$

La distribuzione uniforme - I

$$\begin{cases} f(x) = 0 & x < a, \quad x > b \\ f(x) = \dfrac{1}{b-a} & a \le x \le b \end{cases}$$

$$\begin{cases} F(x) = 0 & x < a \\ F(x) = \dfrac{x-a}{b-a} & a \le x \le b \\ F(x) = 1 & x > b \end{cases}$$

$$E(x) = \frac{a+b}{2}$$

$$\mathrm{Var}(x) = \frac{(b-a)^2}{12}$$

La distribuzione uniforme - II

$$\begin{aligned} M_x(t) &= \int_{-\infty}^{+\infty} f(x)\, e^{tx}\, dx \\ &= \int_a^b \frac{e^{tx}}{b-a}\, dx \\ &= \frac{1}{t(b-a)} \left[e^u\right]_{ta}^{tb} \\ &= \frac{e^{tb} - e^{ta}}{t(b-a)} \end{aligned}$$

$$\phi_x(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

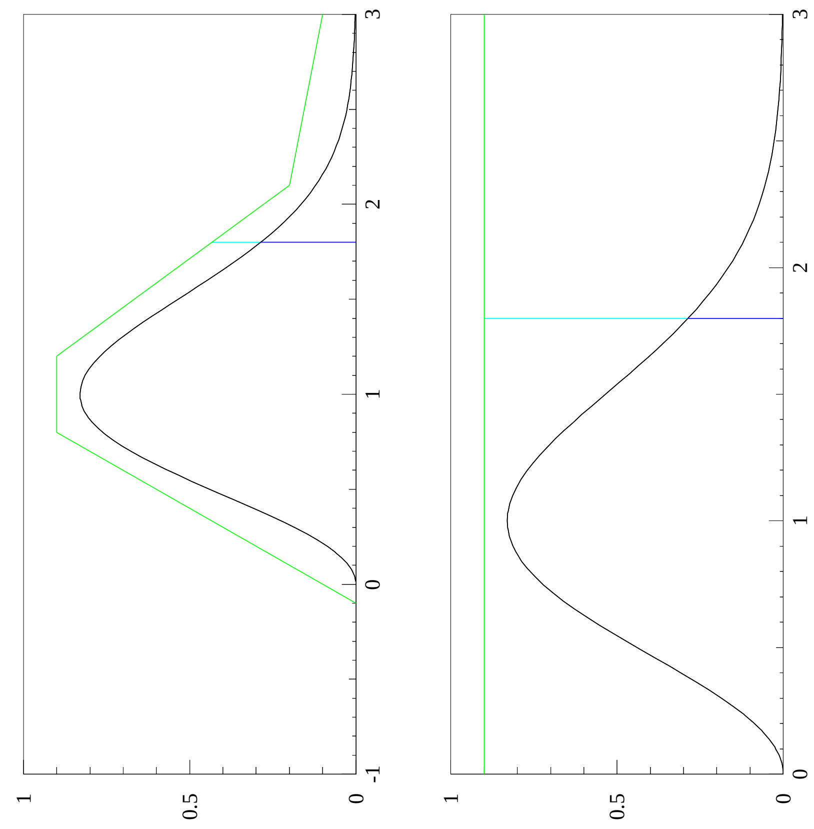
La distribuzione uniforme - III

- Esempio: il decadimento del π^0

$$dP \quad \propto \quad d\Omega \quad \propto \quad d(\cos\theta)\, d\phi$$

Quindi $\cos\theta$ e ϕ sono variabili casuali statisticamente indipendenti e con distribuzione uniforme.

- Applicazione: numeri casuali con distribuzione data:
  - ⋆ Inversione della funzione di distribuzione;
  - ⋆ Il metodo dei rigetti.

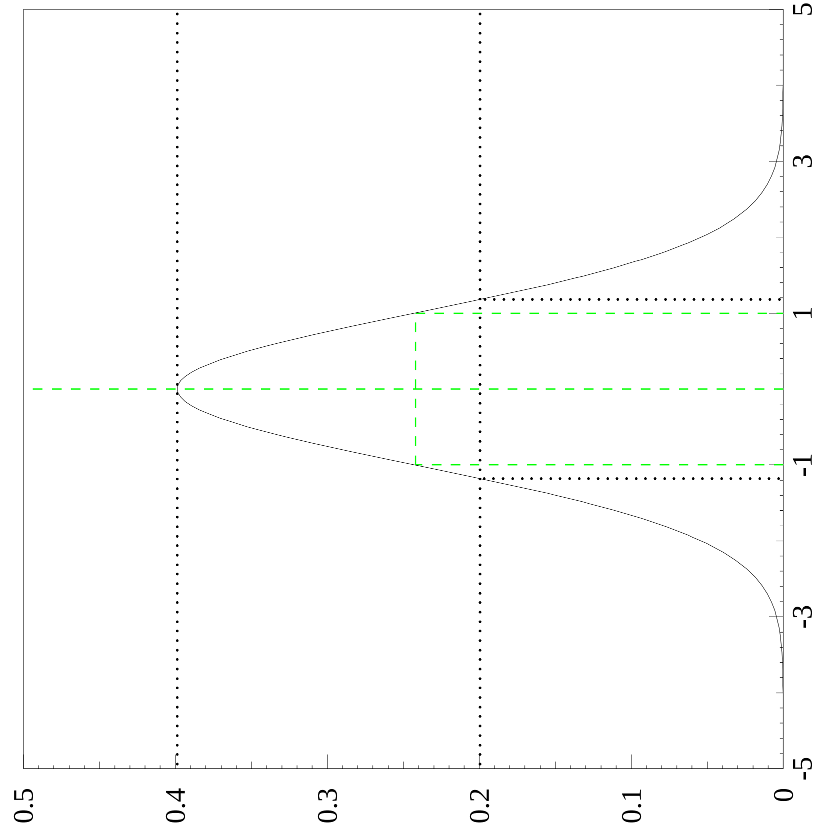


La distribuzione normale - I

$$f(x) \equiv N(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\begin{aligned}
M_x(t) &= \int_{-\infty}^{+\infty} e^{tx}\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{t(x-\mu)} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{\left[\frac{\sigma^2 t^2}{2} - \frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right]} dx \\
&= e^{\left(t\mu + \frac{\sigma^2 t^2}{2}\right)} \int_{-\infty}^{+\infty} N(x;\mu+\sigma^2 t,\sigma)\, dx \\
&= e^{\left(t\mu + \frac{\sigma^2 t^2}{2}\right)}
\end{aligned}$$

$$\phi_x(t) = e^{\left(it\mu - \frac{\sigma^2 t^2}{2}\right)}$$



La distribuzione normale - II

$$M_x(t) = e^{\left(t\mu + \frac{\sigma^2 t^2}{2}\right)}$$

$$\overline{M}_x(t) = e^{-t\mu} \cdot M_x(t) = e^{\frac{\sigma^2 t^2}{2}}$$

$$\frac{d}{dt} M_x(t) = M_x(t) \cdot (\mu + \sigma^2 t)$$

$$\frac{d}{dt} \overline{M}_x(t) = \overline{M}_x(t) \cdot \sigma^2 t$$

$$\frac{d^2}{dt^2} \overline{M}_x(t) = \overline{M}_x(t) \cdot (\sigma^4 t^2 + \sigma^2)$$

$$E(x) = \lambda_1 = \left. \frac{dM_x(t)}{dt} \right|_{t=0} = \mu$$

$$Var(x) = \mu_2 = \left. \frac{d^2 \overline{M}_x(t)}{dt^2} \right|_{t=0} = \sigma^2$$

Scuola di Dottorato in Fisica

La distribuzione normale - III

Per la distribuzione normale:

$$\mu_{2k+1} = 0$$

$$\mu_{2k} = \frac{(2k)!}{2^k k!} \sigma^{2k}$$

Per la variabile standardizzata, $u = \dfrac{x - E(x)}{\sigma}$:

$$E(u) = 0$$

$$Var(u) = 1$$

$$M_u(t) = e^{\frac{t^2}{2}}$$

$$\phi_u(t) = e^{-\frac{t^2}{2}}$$

La distribuzione normale - IV

TEOREMA: *combinazioni lineari di variabili casuali* normali statisticamente indipendenti tra loro seguono anch'esse la distribuzione normale.

$$y = \sum_k a_k x_k$$

$$\phi_{x_k}(t) = e^{\left(it\mu_k - \frac{\sigma_k^2 t^2}{2}\right)}$$

$$\xi_k = a_k x_k$$

$$\phi_{\xi_k}(t) = \phi_{x_k}(a_k t) = e^{\left(ita_k\mu_k - \frac{a_k^2\sigma_k^2 t^2}{2}\right)}$$

$$\phi_y(t) = \prod_k \phi_{\xi_k}(t)$$

$$= e^{\left[it\left(\sum_k a_k\mu_k\right) - \frac{t^2}{2}\left(\sum_k a_k^2\sigma_k^2\right)\right]}$$

$$= e^{\left(it\mu - \frac{t^2\sigma^2}{2}\right)}$$

con

$$\mu = \sum_k a_k \mu_k \qquad e \qquad \sigma^2 = \sum_k a_k^2 \sigma_k^2$$

La distribuzione normale bidimensionale

In due dimensioni,

$$f(x,y) =$$

$$= \frac{e^{\left\{-\frac{1}{2(1-r^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2r\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}}{2\pi\,\sigma_x\,\sigma_y\sqrt{1-r^2}}$$

In funzione delle variabili standardizzate

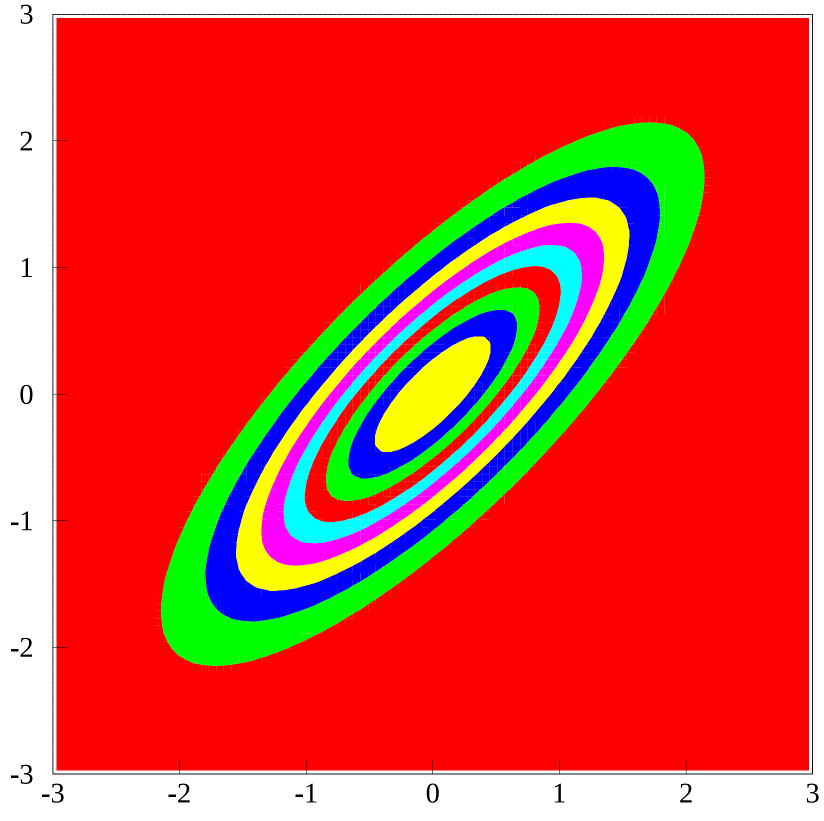
$$u = \frac{x - E(x)}{\sigma_x} \qquad e \qquad v = \frac{y - E(y)}{\sigma_y}$$

$$f(u,v) = \frac{e^{-\frac{1}{2}\frac{u^2-2ruv+v^2}{1-r^2}}}{2\pi\sqrt{1-r^2}}$$

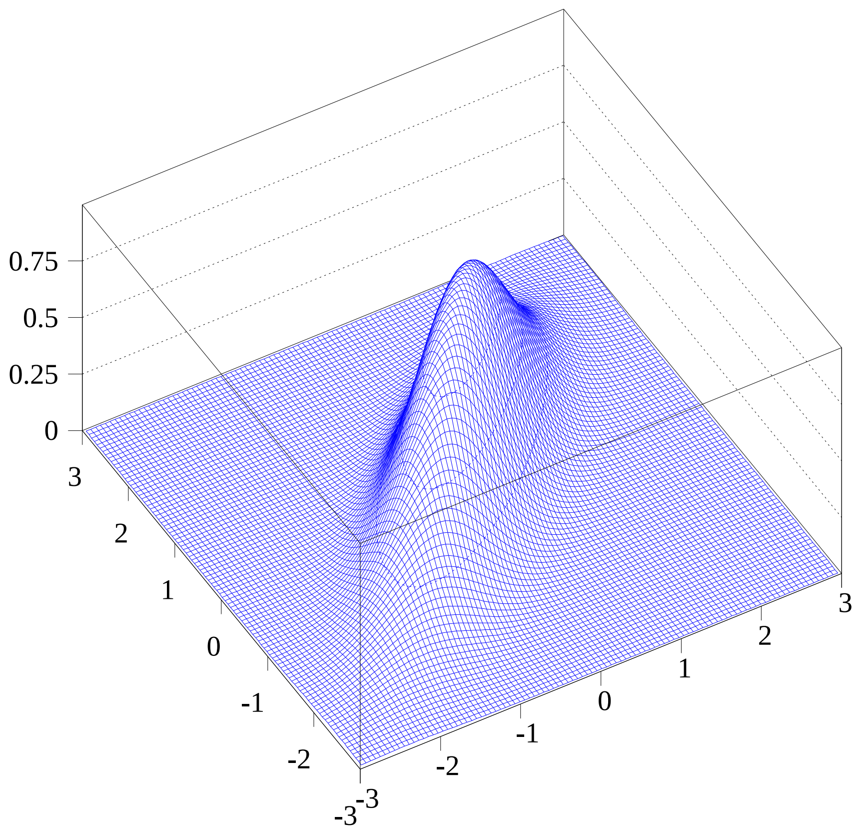
- $r = 0$ è condizione necessaria e sufficiente perchè x ed y siano indipendenti.
- Le distribuzioni marginali sono gaussiane:

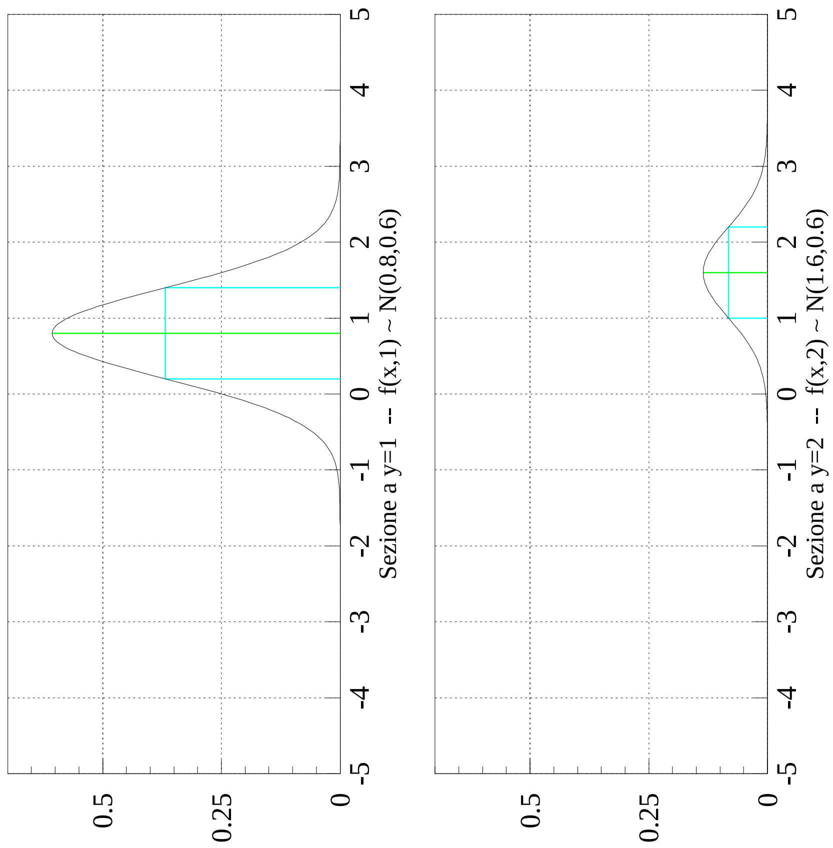
$$\begin{cases} g(x) = N(x;\mu_x,\sigma_x) \\ h(y) = N(y;\mu_y,\sigma_y) \end{cases}$$

- Le densità condizionate sono anch'esse gaussiane; ad esempio, $f(x|y=y_0)$.



r = 0.8





La distribuzione normale N-dimensionale - I

Nel caso più generale la densità di probabilità di una distribuzione Gaussiana N-dimensionale è

$$f(x_1, x_2, \ldots, x_N) = K e^{-\frac{1}{2} H(x_1, x_2, \ldots, x_N)}$$

ove H è una forma quadratica nelle variabili standardizzate

$$t_i = \frac{x_i - E(x_i)}{\sigma_i}$$

nella quale i coefficienti dei termini quadratici sono tra loro tutti uguali; le t_i generalmente non sono indipendenti (e quindi H contiene anche i termini rettangolari del tipo $t_i \cdot t_k$).

K è poi un fattore di normalizzazione che vale

$$K = \sqrt{\frac{\Delta}{(2\pi)^N}}$$

in cui Δ è il determinante della matrice (simmetrica) dei coefficienti della forma quadratica.

La distribuzione normale N-dimensionale - II

- La integrabilità di f implica che le ipersuperfici di equazione $H = \text{cost.}$ siano tutte al finito, e siano quindi iperellissi nello spazio N-dimensionale dei parametri.
- Le funzioni di distribuzione marginali e condizionate di qualsiasi sottoinsieme delle variabili sono ancora normali.
- Esiste sempre un cambiamento di variabili $x_i \to y_k$ che cambia H nella forma canonica (senza termini rettangolari); in tal caso

$$\Delta = \left[\prod_{k=1}^{N} \mathrm{Var}(y_k) \right]^{-1}$$

$$f(y_1, \ldots, y_k) =$$

$$= \frac{1}{\sqrt{\prod_{k=1}^{N} [2\pi \, \mathrm{Var}(y_k)]}} \, e^{-\frac{1}{2} \sum_{k=1}^{N} \frac{[y_k - E(y_k)]^2}{\mathrm{Var}(y_k)}}$$

- Le y_k sono statisticamente indipendenti (e normali).

La distribuzione di Cauchy (o di Breit–Wigner) - I

$$f(x; \theta, d) = \frac{1}{\pi d} \frac{1}{1 + \left(\frac{x-\theta}{d}\right)^2} \qquad (d > 0)$$

$$F(x; \theta, d) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\theta}{d}\right)$$

- Nessuno dei momenti esiste (nemmeno la media)
- θ è la mediana della distribuzione, e d la larghezza a metà altezza (FWHM)
- Funzione caratteristica:

$$\phi_x(t; \theta, d) = e^{\left(i\theta t - |t| \, d\right)}$$

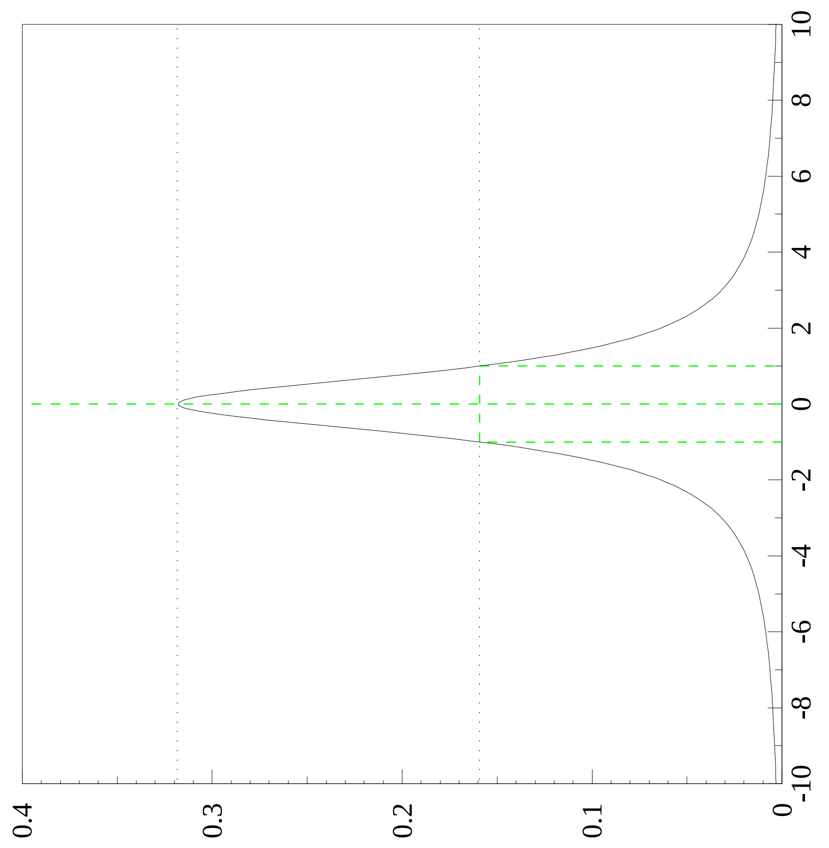
- Variabile standardizzata:

$$u = \frac{x-\theta}{d} \qquad \Longleftrightarrow \qquad x = ud + \theta$$

$$f(u) = \frac{1}{\pi \, (1 + u^2)}$$

$$F(u) = \frac{1}{2} + \frac{1}{\pi} \arctan u$$

$$\phi_u(t) = e^{-|t|}$$



La distribuzione di Cauchy (o di Breit—Wigner) - II

La distribuzione troncata ammette i momenti: limitandoci alla variabile standardizzata,

$$f(u|-K \leq u \leq K) =$$

$$= \begin{cases} 0 & |u| > K \\ \dfrac{1}{2\arctan K}\,\dfrac{1}{(1+u^2)} & |u| \leq K \end{cases}$$

(discontinua in $u = \pm K$);

$$E(u|-K \leq u \leq K) = 0$$

$$Var(u|-K \leq u \leq K) = \frac{K}{\arctan K} - 1$$

La distribuzione di Cauchy (o di Breit–Wigner) - III

Se le x_k sono N variabili casuali indipendenti che seguono la distribuzione di Cauchy con mediana θ_k e FWHM d_k, una loro generica combinazione lineare

$$y = \sum_{k=1}^{N} a_k x_k$$

segue la stessa distribuzione. Infatti

$$\phi_{x_k}(t) = e^{i\theta_k t - |t|\, d_k}$$

e, definendo $\xi_k = a_k x_k$,

$$\phi_{\xi_k}(t) = \phi_{x_k}(a_k t) = e^{i a_k \theta_k t - |t| \cdot |a_k|\, d_k}$$

$$\phi_y(t) = \prod_{k=1}^{N} \phi_{\xi_k}(t) = e^{i\theta_y t - |t|\, d_y}$$

ove si è posto

$$\theta_y = \sum_{k=1}^{N} a_k \theta_k \qquad \text{e} \qquad d_y = \sum_{k=1}^{N} |a_k|\, d_k$$

Rapporto tra due variabili indipendenti

Consideriamo il rapporto tra due variabili casuali statisticamente indipendenti, supponendo che la seconda non sia nulla:

$$\begin{cases} x \sim f(x) \\ y \sim g(y) \qquad\qquad y \neq 0 \end{cases}$$

$$\begin{cases} u = \dfrac{x}{y} \\ v = y \end{cases} \qquad\qquad \begin{cases} x = uv \\ y = v \end{cases}$$

$$f(x,y) = f(x)\, g(y)$$

$$\varphi(u,v) = f(x)\, g(y) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

$$\frac{\partial(x,y)}{\partial(u,v)} = \left\| \begin{matrix} v & u \\ 0 & 1 \end{matrix} \right\|$$

$$\varphi(u,v) = f(uv)\, g(v)\, |v|$$

La distribuzione di Cauchy (o di Breit–Wigner) - IV

Il rapporto tra due variabili casuali normali statisticamente indipendenti è una variabile che segue la distribuzione di Cauchy:

$$f(x), g(y) \sim N(0,1) \qquad\qquad (y \neq 0)$$

$$\varphi(u,v) = \frac{1}{2\pi}\, e^{-\frac{1}{2}u^2v^2} e^{-\frac{1}{2}v^2} |v|$$

$$\frac{dP}{du} = 2\int_0^{+\infty} \varphi(u,v)\, dv$$

$$= \frac{1}{\pi}\int_0^{+\infty} e^{-\frac{1}{2}v^2(1+u^2)}\, v\, dv$$

$$t = \frac{1}{2}v^2(1+u^2) \qquad\qquad dt = (1+u^2)\, v\, dv$$

$$\frac{dP}{du} = \frac{1}{\pi(1+u^2)}\int_0^{+\infty} e^{-t}\, dt$$

$$= \frac{1}{\pi(1+u^2)}\left[-e^{-t}\right]_0^{+\infty}$$

$$= \frac{1}{\pi(1+u^2)}$$

La distribuzione log-normale - I

$$y \sim N(y; \mu, \sigma)$$

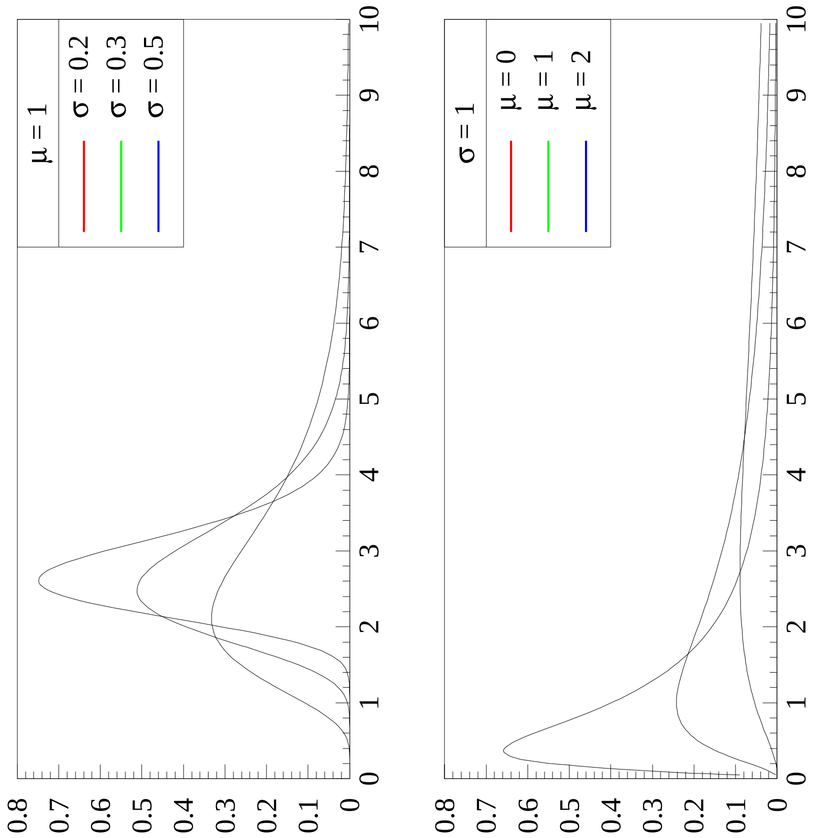
$$g(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

$$x = e^y \qquad \Longleftrightarrow \qquad y = \ln x$$

(corrispondenza biunivoca); $x > 0$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, \frac{1}{x}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$

- Il prodotto di due variabili log-normali ed indipendenti è ancora log-normale.



La distribuzione log-normale - II

$$\frac{\left(y-\mu-k\sigma^2\right)^2}{2\sigma^2}-k\mu-\frac{1}{2}k^2\sigma^2=\frac{(y-\mu)^2}{2\sigma^2}-ky$$

$$\begin{aligned}
\lambda_k &= \int_0^{+\infty} x^k\, f(x)\, dx \\
&= \int_{-\infty}^{+\infty} e^{ky}\, g(y)\, dy \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left[\frac{(y-\mu)^2}{2\sigma^2}-ky\right]}\, dy \\
&= e^{\left(k\mu+\frac{1}{2}k^2\sigma^2\right)} \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{\left(y-\mu-k\sigma^2\right)^2}{2\sigma^2}}\, dy \\
&= e^{\left(k\mu+\frac{1}{2}k^2\sigma^2\right)}
\end{aligned}$$

$$\begin{aligned}
E(x) \equiv \lambda_1 &= e^{\left(\mu+\frac{\sigma^2}{2}\right)} \\
E(x^2) \equiv \lambda_2 &= e^{\left(2\mu+2\sigma^2\right)} \\
Var(x) = \lambda_2-(\lambda_1)^2 &= e^{\left(2\mu+\sigma^2\right)}\left(e^{\sigma^2}-1\right)
\end{aligned}$$

La distribuzione binomiale (o di Bernoulli) - I

- Probabilità p costante su N prove (prove indipendenti):

$$\Pr(x;N) = \frac{N!}{x!\,(N-x)!}\,p^x\,q^{N-x}$$

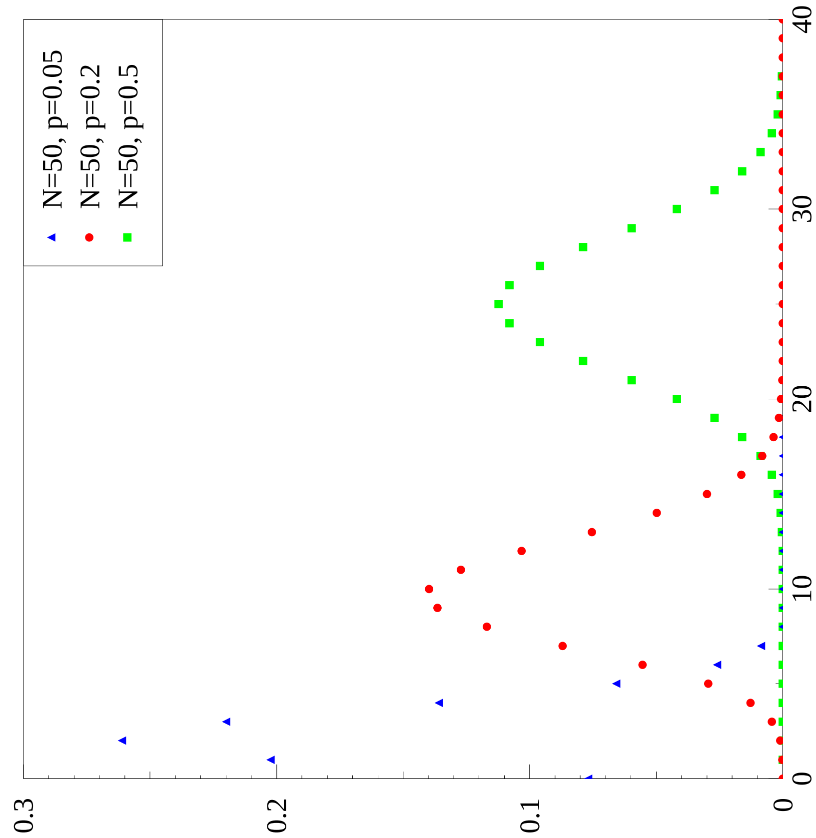
- Variabile di Bernoulli, y:

$$E(y) = p$$

$$E(y^2) = p$$

$$\mathrm{Var}(y) = pq$$

$$x = \sum_{i=1}^{N} y_i \qquad \Rightarrow \qquad \begin{cases} E(x) = Np \\ \mathrm{Var}(x) = Np\,q \end{cases}$$



La distribuzione binomiale (o di Bernoulli) - II

Funzione generatrice dei momenti:

$$\begin{aligned}
M_x(t) &= E\left(e^{tx}\right) \\
&= \sum_{x=0}^{N} e^{tx} \binom{N}{x} p^x q^{N-x} \\
&= \sum_{x=0}^{N} \binom{N}{x} \left(pe^t\right)^x q^{N-x} \\
&= \left(pe^t + q\right)^N
\end{aligned}$$

o anche, ricordando che $q = 1 - p$,

$$M_x(t) = \left[1 + p\left(e^t - 1\right)\right]^N$$

Per la variabile standardizzata

$$z = \frac{x - E(x)}{\sigma_x} = \frac{x - Np}{\sqrt{Np\,q}}$$

$$M_z(t) = e^{-\frac{Np}{\sqrt{Np\,q}}t} \left[1 + p\left(e^{\frac{t}{\sqrt{Np\,q}}} - 1\right)\right]^N$$

La distribuzione binomiale (o di Bernoulli) - III

$$\ln M_z(t) =$$

$$= -\frac{Np}{\sqrt{Np\,q}}\,t + N\,\ln\left[1 + p\left(e^{\frac{t}{\sqrt{Np\,q}}} - 1\right)\right]$$

Sviluppando in serie di MacLaurin prima

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots$$

e poi

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

e svolgendo le potenze, si ottiene:

$$\ln M_z(t) = \frac{1}{2}\,t^2 + O(t^3 N^{-\frac{1}{2}})$$

Per cui

$$\lim_{N\to\infty} M_z(t) = e^{\frac{t^2}{2}}$$

e la distribuzione binomiale tende asintoticamente ad una distribuzione normale con la stessa media e la stessa varianza.

La distribuzione binomiale (o di Bernoulli) - IV

- Per la frequenza relativa $f = \frac{x}{N}$ nelle N prove:

$$E(f) = \frac{E(x)}{N} = p$$

$$Var(f) = \frac{Var(x)}{N^2} = \frac{p\,q}{N}$$

- Applicazione: gas perfetto

$$E(N_1) = Np \qquad p = \frac{V_1}{V_1 + V_2}$$

$$\sigma_{N_1} = \sqrt{Np\,q}$$

$$\frac{\sigma_{N_1}}{N_1} \approx \sqrt{\frac{q}{Np}}$$

Scuola di Dottorato in Fisica

Esercizio

Ci sono 8 dottorandi, ognuno dei quali passa in Dipartimento il 50% del proprio tempo; quante scrivanie servono per attrezzare un'aula di studio riservata ai dottorandi, nell'ipotesi che ognuno dei presenti possa trovare un posto disponibile per almeno il 90% del tempo?

Soluzione

$p = 0.5$; $q = 1 - p = 0.5$. La probabilità che in un gruppo di 8 dottorandi N siano presenti contemporaneamente è data dalla distribuzione di Bernoulli: $P(N) = C^8_N p^N q^{8-N} = C^8_N 0.5^8$. Quello che si chiede è di calcolare il minimo intero N per cui risulti $\sum_{i=0}^N P(i) \geq 0.9$

N	$P(N)$	$\sum_{i=0}^{N} P(i)$
0	0.004	0.004
1	0.031	0.035
2	0.109	0.145
3	0.219	0.363
4	0.273	0.637
5	0.219	0.855
6	0.109	0.965
7	0.031	0.996
8	0.004	1.000

Servono almeno 6 scrivanie (nel qual caso i dottorandi presenti troveranno posto nel 96.5% del tempo).

Scuola di Dottorato in Fisica

La distribuzione binomiale (o di Bernoulli) - V

Esempio: rapporto di asimmetria, R

$$R = \frac{F-B}{F+B} = \frac{F-B}{N} = \frac{2F}{N} - 1$$

$$F = \frac{N(R+1)}{2}$$

$$\Pr(F) = \binom{N}{F} p^F (1-p)^{N-F}$$

$$E(F) = Np$$

$$\mathrm{Var}(F) = Np\,(1-p)$$

$$\Pr(R) = \binom{N}{\frac{N(R+1)}{2}} p^{\frac{N(R+1)}{2}} (1-p)^{N-\frac{N(R+1)}{2}}$$

$$E(R) = \frac{2}{N} E(F) - 1 = 2p - 1$$

$$\mathrm{Var}(R) = \frac{4}{N^2} \mathrm{Var}(F) = \frac{4p\,(1-p)}{N}$$

Se il numero di eventi nel campione è elevato,

$$p \approx \frac{F}{N} \qquad q \approx \frac{B}{N}$$

$$E(R) \approx 2\frac{F}{N} - 1 \qquad \mathrm{Var}(R) \approx 4\frac{FB}{N^3}$$

La distribuzione multinomiale

$$\Pr(x_1, x_2, \ldots, x_M) =$$
$$= \frac{N!}{x_1!\, x_2! \cdots x_M!}\, p_1^{x_1}\, p_2^{x_2} \cdots p_M^{x_M}$$

con

$$\sum_{i=1}^{M} x_i = N \qquad \sum_{i=1}^{M} p_i = 1$$

Risulta:

$$E(x_i) = Np_i \qquad i = 1, \ldots N$$
$$Var(x_i) = Np_i(1 - p_i) \qquad i = 1, \ldots N$$
$$Cov(x_i, x_j) = -Np_i p_j \qquad i \neq j$$

- Esempio: eventi nei bins di un istogramma (la frequenza assoluta in due bins è correlata negativamente)

Scuola di Dottorato in Fisica

La distribuzione di Poisson - I

- Ipotesi:
  - ⋆ $\Pr(1; dt) = \lambda\, dt$
  - ⋆ Indipendenza statistica
  - ⋆ La probabilità che più di un evento si verifichi in un tempo dt è infinitesima di ordine superiore a dt

Eventi rari su larga base statistica

$$P(x; t + dt) =$$
$$= P(x - 1; t)\, \lambda\, dt + P(x; t)\, (1 - \lambda\, dt)$$

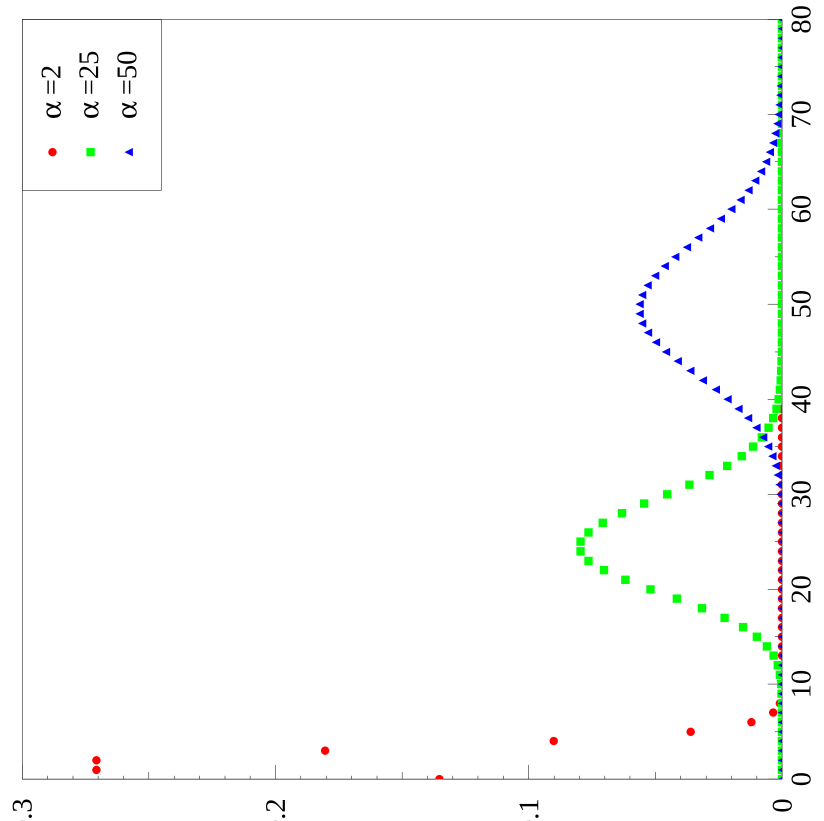
$$\frac{d}{dt} P(x; t) = -\lambda P(x; t) + \lambda P(x - 1; t)$$

$$\frac{d}{dt} P(0; t) = -\lambda P(0; t)$$

$$P(0; t) = e^{-\lambda t}$$

$$P(x; t) = \frac{(\lambda t)^x}{x!}\, e^{-\lambda t}$$

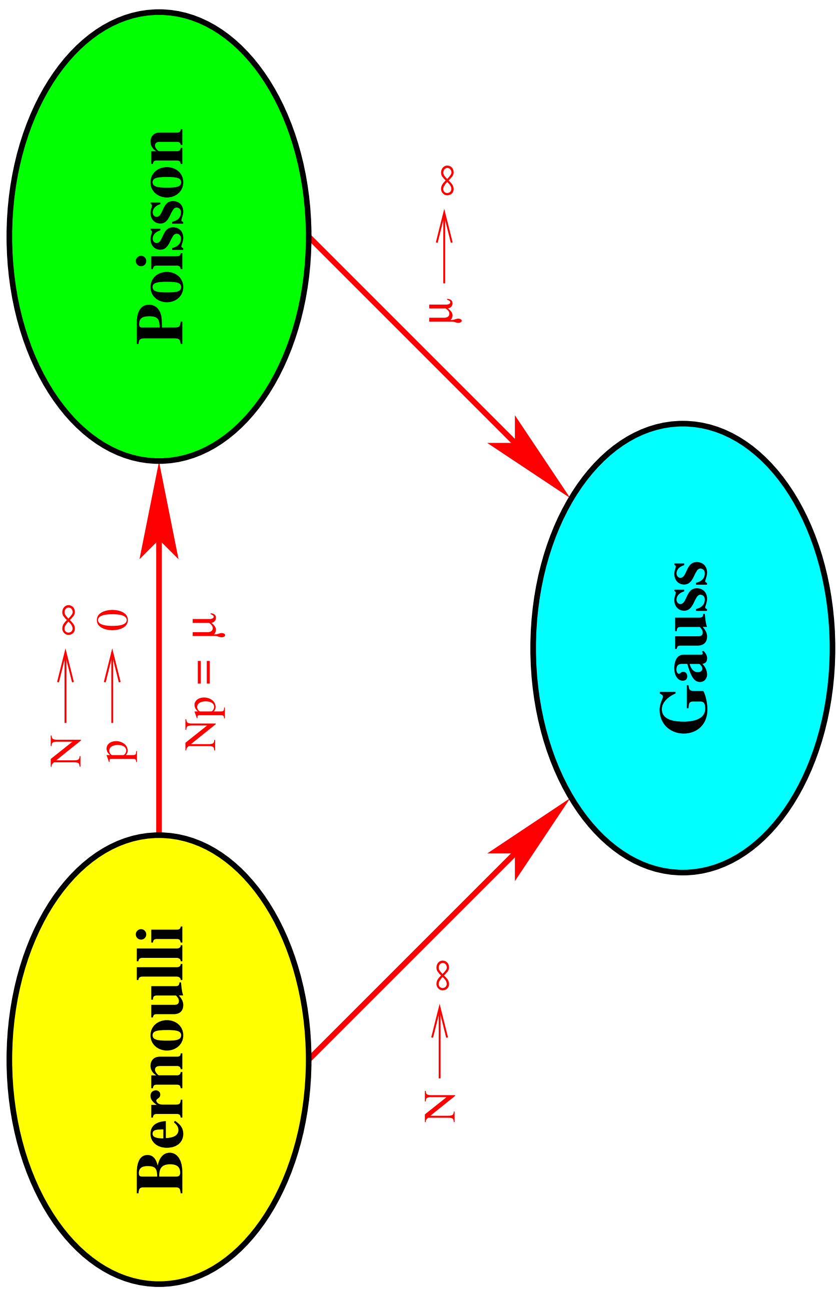
$$P(x; a) = \frac{a^x}{x!}\, e^{-a}$$



La distribuzione di Poisson - II

$$
\begin{aligned}
M_x(t) = E\left(e^{tx}\right) &= \sum_{x=0}^{\infty} e^{tx} \cdot \frac{a^x}{x!}\, e^{-a} \\
&= e^{-a} \sum_{x=0}^{\infty} \frac{\left(ae^t\right)^x}{x!} \\
&= e^{-a} e^{ae^t} \\
&= e^{a(e^t-1)}
\end{aligned}
$$

- $E(x) = a$
- $\mathrm{Var}(x) = a$
- Tende asintoticamente ad una distribuzione normale
- Esempi: decadimenti; urti nei gas; numero di particelle per burst; numero di interazioni per unità di lunghezza; ionizzazione, bremsstrahlung, etc. delle particelle; conteggio di cellule; gocce di pioggia; incidenti stradali; . . .



La distribuzione di Poisson - III

- La somma $z = x + y$ di due variabili di Poisson indipendenti è ancora distribuita secondo Poisson:

$$\begin{aligned} M_z(t) &= e^{\xi(e^t-1)} e^{\eta(e^t-1)} \\ &= e^{(\xi+\eta)(e^t-1)} \end{aligned}$$

Il valor medio di z è $\zeta = \xi + \eta$.

  - ⋆ Non vale per la differenza (che può essere negativa);
  - ⋆ Vale per la somma di un numero qualsiasi di variabili di Poisson tutte indipendenti.

La distribuzione di Poisson - IV

Applicazione: esperimenti negativi

$$a \;=\; N_0(1 - e^{-\frac{t}{\tau}}) \;\approx\; N_0\,\frac{t}{\tau}$$

$$\Pr(0) \;=\; \frac{a^0}{0!}\,e^{-a} \;=\; e^{-a} \;\approx\; e^{-N_0\frac{t}{\tau}}$$

$$\Pr(0) > CL \qquad \Rightarrow \qquad \tau > -\frac{N_0\,t}{\ln(CL)}$$

Scuola di Dottorato in Fisica

La distribuzione esponenziale - I

Sia la variabile casuale δ, intervallo di tempo tra due eventi successivi il cui numero x segua la distribuzione di Poisson:

$$\Pr(x;t) = \frac{(\lambda t)^x}{x!}\,e^{-\lambda t}$$

Visto che nel tempo δ non si verificano eventi,

$$\Pr(\delta > d) \;\equiv\; \Pr(0;d) \;=\; e^{-\lambda d}$$

quindi la funzione di distribuzione di δ è

$$F(d) \;=\; \Pr(\delta \leq d) \;=\; 1 - e^{-\lambda d}$$

e la funzione di frequenza è esponenziale:

$$f(\delta) = \frac{d\,F(\delta)}{d\delta} \qquad \Rightarrow \qquad f(\delta) = \lambda\,e^{-\lambda\delta}$$

Si può trovare, volendo, la funzione caratteristica; che vale

$$\phi_\delta(t) = \frac{\lambda}{\lambda - it}$$

La distribuzione esponenziale - II

I primi due momenti della distribuzione esponenziale si possono trovare o integrando direttamente $f(\delta)$, o derivando successivamente $\phi_\delta(t)$:

$$\frac{d\,\phi_\delta(t)}{dt} = \frac{d}{dt}\left(\frac{\lambda}{\lambda - it}\right) = \frac{i\lambda}{(\lambda - it)^2}$$

$$\left.\frac{d\,\phi_\delta(t)}{dt}\right|_{t=0} = \frac{i}{\lambda} = i \cdot E(\delta)$$

$$E(\delta) = \frac{1}{\lambda}$$

$$\frac{d^2\,\phi_\delta(t)}{dt^2} = \frac{-i\lambda \cdot 2\,(\lambda - it)(-i)}{(\lambda - it)^4} = -\frac{2\,\lambda}{(\lambda - it)^3}$$

$$\left.\frac{d^2\,\phi_\delta(t)}{dt^2}\right|_{t=0} = -\frac{2}{\lambda^2} = i^2 \cdot E(\delta^2)$$

$$E(\delta^2) = \frac{2}{\lambda^2}$$

$$\mathrm{Var}(\delta) = E(\delta^2) - \big[E(\delta)\big]^2 = \frac{1}{\lambda^2}$$

La distribuzione esponenziale - III

Se la variabile casuale t segue la distribuzione esponenziale, calcoliamo la probabilità che t sia maggiore di $t_0 + \Delta t$ condizionata dal sapere che t è sicuramente maggiore di t_0:

$$\begin{aligned}
\Pr\big(t > t_0 + \Delta t | t > t_0\big) &= \frac{\Pr(t > t_0 + \Delta t)}{\Pr(t > t_0)} \\
&= \frac{\int_{t_0+\Delta t}^{+\infty} e^{-\lambda t} dt}{\int_{t_0}^{+\infty} e^{-\lambda t} dt} \\
&= \frac{\big[-e^{-\lambda t}\big]_{t_0+\Delta t}^{+\infty}}{\big[-e^{-\lambda t}\big]_{t_0}^{+\infty}} \\
&= \frac{e^{-\lambda(t_0+\Delta t)}}{e^{-\lambda t_0}} \\
&= e^{-\lambda \Delta t} \\
&= \Pr(t > \Delta t)
\end{aligned}$$

La distribuzione esponenziale non ricorda la storia precedente: il presentarsi dell'evento nel tempo Δt non dipende da t_0.

Esercizio

Ad un incrocio stradale passa in media un'auto ogni 15 secondi; quale è la probabilità P che tra due passaggi successivi intercorra un tempo compreso tra 12 e 16 secondi?

Scuola di Dottorato in Fisica

Soluzione

Soluzione I

$$
\begin{aligned}
P &= \sum_{i=1}^{\infty} P(0;12) \cdot P(i;4) \\
&= P(0;12) \sum_{i=1}^{\infty} P(i;4) \\
&= P(0;12) \left[1 - P(0;4) \right] \\
&= e^{-\frac{12}{15}} \left[1 - e^{-\frac{4}{15}} \right] \\
&\approx 0.4493 \cdot [1 - 0.7659] \\
&\approx 10.52\%
\end{aligned}
$$

Soluzione II

$$
\begin{aligned}
f(\delta) &= \lambda e^{-\lambda\delta} \\
P &= \int_{t_1}^{t_2} f(\delta)\, d\delta = \left[-e^{-\lambda\delta} \right]_{t_1}^{t_2} = e^{-\lambda t_1} - e^{-\lambda t_2}
\end{aligned}
$$

$$
\left.\begin{aligned}
\lambda &= \tfrac{1}{15} \\
t_1 &= 12 \\
t_2 &= 16
\end{aligned}\right\} \quad \Rightarrow \quad P \approx 10.52\%
$$

La distribuzione di Poisson - V

Esempio: rapporto di asimmetria. Nell'esempio precedente abbiamo supposto N dato; invece N è una variabile casuale distribuita secondo Poisson con media ν. Indicando con p la probabilità $\Pr(F)$,

$$\Pr(F, N) = \\ = \frac{\nu^N}{N!} e^{-\nu} \frac{N!}{F!\,(N-F)!} p^F (1-p)^{N-F}$$

Ponendo $B = N - F$ e $q = 1 - p$,

$$\Pr(F, B) = e^{-\nu} \frac{(\nu p)^F (\nu q)^B}{F!\,B!} \\ = \frac{(\nu p)^F}{F!} e^{-\nu p} \frac{(\nu q)^B}{B!} e^{-\nu q}$$

che è il prodotto di due funzioni di frequenza di Poisson.

La composizione di binomiale e Poissoniana equivale al prodotto di due Poissoniane: il numero di decadimenti segue la statistica di Poisson; la scelta dello stato finale quella binomiale; e tutto avviene come se i decadimenti dei due tipi si verificassero separatamente ed indipendentemente secondo la statistica di Poisson.

La distribuzione di Poisson - VI

Accertato che F e B sono variabili indipendenti che seguono la statistica di Poisson, per il rapporto di asimmetria asintoticamente (quindi grandi N) si ricava:

$$\begin{aligned} \mathrm{Var}(F) &= E(F) \approx F \\ \mathrm{Var}(B) &= E(B) \approx B \end{aligned}$$

$$R = \frac{F-B}{F+B} \approx \frac{E(F)-E(B)}{E(F)+E(B)} + \\ + \frac{\partial R}{\partial F}\big[F - E(F)\big] + \frac{\partial R}{\partial B}\big[B - E(B)\big]$$

$$E(R) \approx \frac{E(F)-E(B)}{E(F)+E(B)} \approx \frac{2F}{N} - 1$$

$$\mathrm{Var}(R) \approx \left(\frac{\partial R}{\partial F}\right)^2 \mathrm{Var}(F) + \left(\frac{\partial R}{\partial B}\right)^2 \mathrm{Var}(B) \\ = \frac{4}{(F+B)^4}\big[B^2\,\mathrm{Var}(F) + F^2\,\mathrm{Var}(B)\big] \\ = \frac{4FB}{(F+B)^3} = 4\,\frac{FB}{N^3}$$

e le cose non cambiano rispetto alla precedente analisi.

Distribuzione composta di Poisson — I

La funzione caratteristica della distribuzione di Poisson può essere scritta nelle due forme

$$\phi_x(t) = e^{a\left(e^{it}-1\right)}$$

e

$$\phi_x(z) = e^{a(z-1)}$$

Riprendiamo la somma di un numero casuale di variabili casuali aventi tutte la stessa distribuzione,

$$S = \sum_{i=1}^{N} x_i$$

Se sia N che ognuna delle x seguono la distribuzione di Poisson, con medie ν e ξ rispettivamente,

$$\begin{cases} \phi_N(z) = e^{\nu(z-1)} \\ \phi_x(z) = e^{\xi(z-1)} \end{cases} \quad \Rightarrow \quad \phi_S(z) = e^{\nu\left[e^{\xi(z-1)}-1\right]}$$

(distribuzione composta di Poisson).

Scuola di Dottorato in Fisica

Distribuzione composta di Poisson — II

Per la distribuzione composta di Poisson si conosce dunque la funzione caratteristica di variabile reale $\phi_S(t)$, che vale

$$\phi_S(t) = e^{\nu\left[e^{\xi\left(e^{it}-1\right)}-1\right]}$$

Da essa si ricava:

$$\frac{d\,\phi_S(t)}{dt} = \phi_S(t) \cdot \nu\, e^{\xi\left(e^{it}-1\right)} \cdot \xi e^{it} \cdot i$$

$$\left.\frac{d\,\phi_S(t)}{dt}\right|_{t=0} = i\nu\,\xi$$

$$E(S) = \nu\,\xi$$

e, similmente, si potrebbe ottenere

$$Var(S) = \nu\,\xi\,(1+\xi)$$

e

$$Pr(S) = \sum_{N=0}^{\infty} \left[\frac{(N\xi)^S}{S!}\, e^{-N\xi} \cdot \frac{\nu^N}{N!}\, e^{-\nu} \right]$$

Esempio: l'osservazione di un quark isolato — I

1. McCusker and Cairns
   Evidence of quarks in air-shower cores
   Phys. Rev. Lett. **23** (1969), pagg. 658–659
2. Adair and Kasha
   Analysis of some results of quark searches
   Phys. Rev. Lett. **23** (1969), pagg. 1355–1358
3. Eadie, Drijard, James, Roos e Sadoulet
   Statistical methods in experimental physics
   pag. 53

In una camera a nebbia il numero medio di gocce generate per unità di lunghezza da tracce di un certo tipo è $a = 229$; è stata osservata una traccia con $n = 110$ gocce per unità di lunghezza sulle 55000 esaminate ($f = 1/55000 \approx 2 \times 10^{-5}$).

L'osservazione è significativa?

Scuola di Dottorato in Fisica

Esempio: l'osservazione di un quark isolato — II

Prima interpretazione

- Il processo di generazione delle gocce segue la statistica di Poisson.

La probabilità di osservare una traccia con 110 (o meno) gocce per unità di lunghezza è quindi

$$\Pr(n \leq 110) = \sum_{k=0}^{110} \frac{229^k}{k!} e^{-229} \approx 1.6 \times 10^{-18}$$

13 ordini di grandezza più piccola della frequenza sperimentale.

Esempio: l'osservazione di un quark isolato — III

Seconda interpretazione

- In ogni *scattering elementare* vengono prodotte in media $\nu = 4$ gocce.
- Il numero medio di scattering elementari per unità di lunghezza è $229/4 = 57.25$.
- L'evento in questione ne ha $110/4 = 27.5$.

La probabilità di causare meno di 28 scattering elementari per unità di lunghezza vale

$$\Pr(n \leq 110) \approx \sum_{k=0}^{27} \frac{57.25^k}{k!} e^{-57.25} \approx 6.7 \times 10^{-6}$$

più di 33 volte superiore alla frequenza osservata.

Esempio: l'osservazione di un quark isolato — IV

Terza (e ultima) interpretazione

- Il numero s di scattering elementari per unità di lunghezza è una variabile casuale che segue la statistica di Poisson con media $\lambda = 57.25$:

$$\Pr(s) \equiv P(s;\lambda) = \frac{\lambda^s}{s!} e^{-\lambda}$$

- Il numero g di gocce prodotte nel singolo scattering segue anch'esso la statistica di Poisson ed ha probabilità $P(g;\mu)$, con media $\mu = 4$.
- La distribuzione del numero k di gocce per unità di lunghezza quindi segue una distribuzione composta di Poisson.

$$\Pr(k) = \sum_{s=0}^{\infty} \left[\frac{(s\mu)^k}{k!} e^{-s\mu} \cdot \frac{\lambda^s}{s!} e^{-\lambda} \right]$$

$$\Pr(k \leq 110) = \sum_{k=0}^{110} \Pr(k) \approx 4.7 \times 10^{-5}$$

dello stesso ordine di grandezza della frequenza osservata.

La distribuzione di Poisson - VII

Ipotesi: un processo fisico segue la distribuzione di Poisson; il valor medio x della distribuzione è noto con un certo errore che supponiamo normale

$$x \sim N(x;\mu,\sigma)$$

Qual'è la probabilità di osservare 0 eventi?

Scuola di Dottorato in Fisica

La distribuzione di Poisson - VIII

$$\Pr\left(x \in [x, x+dx]\right) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$\Pr(0|x) = e^{-x}$$

$$\begin{aligned}\Pr(0;\mu,\sigma) &= \int_{-\infty}^{+\infty} e^{-x}\, N(x;\mu,\sigma)\, dx \\ &\equiv M_x(-1) \\ &= e^{\left(-\mu+\frac{\sigma^2}{2}\right)}\end{aligned}$$

Insomma:

$$\Pr(0;\mu,\sigma) = e^{-\mu}\cdot e^{\frac{\sigma^2}{2}}$$

Può essere maggiore di 1 !!!

$$\mu \gg \sigma$$

La distribuzione di Poisson - IX

- Eventi rari in assenza di fondo: la probabilità di un segnale s è:

$$\Pr(s) = \frac{S^s}{s!}\, e^{-S} = P(s; S)$$

- Se S è ignoto e sono stati osservati N eventi, si può dare un limite su S calcolando il valore S_u per il quale la probabilità di osservare un numero di eventi non superiore a N vale una quantità predeterminata: se

$$\sum_{x=0}^{N} P(x; S_u) = \epsilon$$

diremo di poter affermare che $S \leq S_u$ con un livello di confidenza

$$C_L = \epsilon$$

intendendo con questo che, *se risulta $S \leq S_u$, in una frazione almeno pari ad ϵ di esperimenti analoghi al nostro otterremmo al massimo N eventi.*

Scuola di Dottorato in Fisica

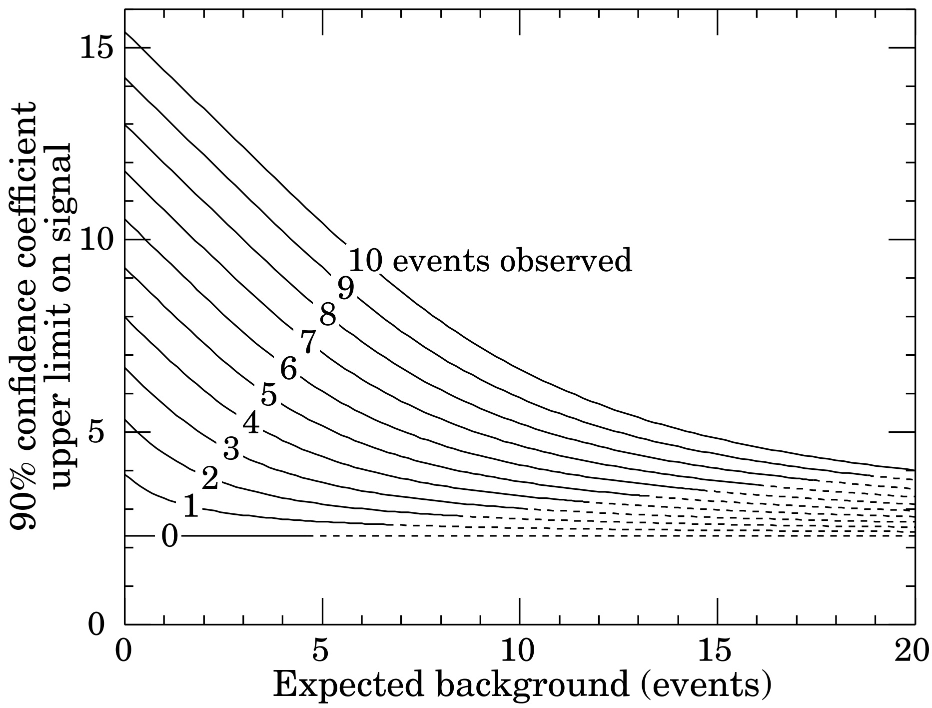
La distribuzione di Poisson - X

- Se il fondo è presente, se è indipendente dal segnale e se segue la distribuzione di Poisson con valore medio noto F, già sappiamo che la probabilità di osservare N eventi tra fondo e segnale segue la distribuzione di Poisson con valore medio $F + S$:

$$\Pr(N) \equiv P(N; F + S) = \frac{(F+S)^N}{N!}\, e^{-(F+S)}$$

- Se sono stati realmente osservati N eventi, si può determinare un limite superiore su S calcolando il valore S_u per il quale la probabilità di osservare un numero di eventi (tra fondo e segnale) non superiore a N *e condizionato all'avere ottenuto un numero di eventi di fondo che non può superare N* vale una quantità predeterminata (PDG):

$$\frac{\sum_{n=0}^{N} P(n; F + S_u)}{\sum_{f=0}^{N} P(f; F)} = \epsilon \quad \Rightarrow \quad S \leq S_u \quad (\mathrm{CL} = \epsilon)$$



Limite superiore (per un livello di confidenza del 90%) sugli eventi di segnale, in funzione del numero atteso di eventi di fondo.

La distribuzione di Poisson - XI

- In presenza di indeterminazione Gaussiana del fondo, $F \sim N(F; \mu_F, \sigma_F)$ con $\mu_F \gg \sigma_F$:

$$\frac{\sum_{n=0}^{N} \int dF\, P(n; S_u + F)\, e^{-\frac{1}{2}\left(\frac{F-\mu_F}{\sigma_F}\right)^2}}{\sum_{f=0}^{N} \int dF\, P(f; F)\, e^{-\frac{1}{2}\left(\frac{F-\mu_F}{\sigma_F}\right)^2}} = \epsilon \quad \Rightarrow$$

$$\Rightarrow \quad S \le S_u \quad (CL = \epsilon)$$

Teoria del campionamento casuale - I

Ipotesi: popolazione infinita (o reinserimento); indipendenza statistica delle estrazioni; funzione di frequenza **qualsiasi**; esistenza della media.

$$f(\underline{X}) = f(x_1) \cdot f(x_2) \cdots f(x_N)$$

$\bar{x}$ è essa stessa una variabile casuale:

$$\bar{x} = \sum_{k=1}^{N} \frac{x_k}{N}$$

$$\phi_{\bar{x}}(t) = \left[\phi_x\left(\frac{t}{N}\right)\right]^N = \left[\phi_x(t')\right]^N$$

$$\frac{d\,\phi_{\bar{x}}(t)}{dt} = N\left[\phi_x(t')\right]^{N-1} \cdot \frac{d\,\phi_x(t')}{dt'} \cdot \frac{1}{N}$$

$$\phi_x(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\,dx \qquad \Rightarrow \qquad \phi_x(0) \equiv 1$$

$$\left.\frac{d\,\phi_{\bar{x}}(t)}{dt}\right|_{t=0} = \left.\frac{d\,\phi_x(t')}{dt'}\right|_{t'=0}$$

$$\boxed{E(\bar{x}) = E(x)}$$

Scuola di Dottorato in Fisica

Teoria del campionamento casuale - II

Ipotesi aggiuntiva: esistenza della varianza.

$$\begin{aligned}\frac{d^2\,\phi_{\bar{x}}(t)}{dt^2} &= \\ &= (N-1)\left[\phi_x(t')\right]^{N-2} \cdot \left[\frac{d\,\phi_x(t')}{dt'}\right]^2 \cdot \frac{1}{N} + \\ &\qquad + \left[\phi_x(t')\right]^{N-1} \cdot \frac{d^2\,\phi_x(t')}{dt'^2} \cdot \frac{1}{N}\end{aligned}$$

$$E(\bar{x}^2) = \frac{N-1}{N}\left[E(x)\right]^2 + \frac{1}{N}E(x^2)$$

$$\begin{aligned}Var(\bar{x}) &= E(\bar{x}^2) - \left[E(\bar{x})\right]^2 \\ &= \frac{N-1}{N}\left[E(x)\right]^2 + \frac{1}{N}E(x^2) - \left[E(x)\right]^2 \\ &= \frac{1}{N}E(x^2) - \frac{1}{N}\left[E(x)\right]^2\end{aligned}$$

$$\boxed{Var(\bar{x}) = \frac{Var(x)}{N}}$$

Il teorema di Čebišef

Applicando la diseguaglianza di Bienaymé–Čebišef

$$\Pr\Big(|x - E(x)| \geq \epsilon\Big) \;\leq\; \frac{Var(x)}{\epsilon^2}$$

alla media del campione $\bar{x}$ si ottiene:

$$\Pr\Big(|\bar{x} - E(x)| \geq \epsilon\Big) \;\leq\; \frac{Var(x)}{N\epsilon^2}$$

che implica

$$\lim_{N\to\infty} \bar{x} = E(x)$$

(*teorema di Čebišef*); l'unica ipotesi è l'esistenza dei primi due momenti (media e varianza).

La legge dei grandi numeri

Sia un evento casuale A qualsiasi, e y la variabile di Bernoulli connessa ad una prova:

$$f = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Già sappiamo come risulti

$$E(y) = p$$

Applicando alla y il teorema di Čebišef, ne discende come corollario il Teorema di Bernoulli, o legge "dei grandi numeri": la frequenza di un evento casuale **qualsiasi** (per cui esista media e varianza) converge statisticamente alla sua probabilità.

Teoria del campionamento casuale - III

Per la distribuzione di Cauchy,

$$\phi_x(t) = e^{(i\theta t - |t|\,d)}$$

$$\begin{aligned}\phi_{\bar{x}}(t) &= \left[\phi_x\left(\frac{t}{N}\right)\right]^N \\ &= \left[e^{\left(i\theta \frac{t}{N} - \frac{|t|}{N}\,d\right)}\right]^N \\ &= e^{(i\theta t - |t|\,d)} \;\equiv\; \phi_x(t)\end{aligned}$$

La media del campione è distribuita **come la singola misura**; effettuando più di una misura non si guadagna alcuna informazione (dalla media del campione).

In casi come questo si può usare la mediana del campione, che, si può dimostrare, ha distribuzione asintoticamente normale con media θ e varianza

$$\mathrm{Var}(\tilde{x}) = \frac{1}{4N\,[f(\tilde{x};\theta,d)]^2}$$

e per la quale, ovviamente,

$$\lim_{N\to\infty} \mathrm{Var}(\tilde{x}) = 0$$

Scuola di Dottorato in Fisica

Teoria del campionamento casuale - IV

La varianza del campione è

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Si prova facilmente che risulta

$$\frac{1}{N}\sum_{i=1}^{N}\big[x_i - E(x)\big]^2 = s^2 + \big[\bar{x} - E(x)\big]^2$$

Prendendo i valori medi

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left\{E\big[x_i - E(x)\big]^2\right\} &= \\ &= E(s^2) + E\left\{\big[\bar{x} - E(x)\big]^2\right\}\end{aligned}$$

e quindi

$$\mathrm{Var}(x) = E(s^2) + \mathrm{Var}(\bar{x}) = E(s^2) + \frac{\mathrm{Var}(x)}{N}$$

$$E(s^2) = \frac{N-1}{N}\,\mathrm{Var}(x) \quad \xrightarrow{N\to\infty} \quad \mathrm{Var}(x)$$

Il teorema del limite centrale - I

Siano N variabili indipendenti provenienti dalla stessa popolazione, avente funzione di frequenza **qualsiasi** purchè dotata di media μ e di varianza σ^2; la funzione di frequenza della media aritmetica del campione $\bar{x}$ è sotto queste ipotesi asintoticamente **normale** con media μ e varianza σ^2/N.

$$S = \sum_{k=1}^{N} x_k \qquad \phi_S(t) = \left[\phi_x(t)\right]^N$$

$$E(S) = N\mu \qquad Var(S) = N\sigma^2$$

Passando agli scarti dalla media, $z = x - \mu$:

$$\phi_z(t) = e^{-i\mu t}\phi_x(t)$$

Se esistono **tutti** i momenti (restrittivo!)

$$\phi_z(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\lambda_k$$

$$= 1 + 0 - \frac{1}{2}t^2\sigma^2 + O(t^3)$$

Scuola di Dottorato in Fisica

Il teorema del limite centrale - II

La variabile standardizzata

$$y = \frac{S - E(S)}{\sigma_S} = \frac{1}{\sigma\sqrt{N}}S - \frac{N\mu}{\sigma\sqrt{N}}$$

ha funzione caratteristica

$$\phi_y(t) = e^{-it\frac{N\mu}{\sigma\sqrt{N}}} \cdot \phi_S\left(\frac{t}{\sigma\sqrt{N}}\right)$$

$$= e^{-it\frac{N\mu}{\sigma\sqrt{N}}}\left[\phi_x\left(\frac{t}{\sigma\sqrt{N}}\right)\right]^N$$

$$= \left[e^{-i\mu\frac{t}{\sigma\sqrt{N}}} \cdot \phi_x\left(\frac{t}{\sigma\sqrt{N}}\right)\right]^N$$

$$= \left[\phi_z\left(\frac{t}{\sigma\sqrt{N}}\right)\right]^N$$

$$= \left[1 - \frac{1}{2}\frac{t^2}{N\sigma^2}\sigma^2 + O\left(\frac{t^3}{N^{\frac{3}{2}}}\right)\right]^N$$

$$= \left[1 - \frac{t^2}{2N} + O\left(N^{-\frac{3}{2}}\right)\right]^N$$

Il teorema del limite centrale - III

Sfruttando il limite notevole

$$\lim_{x\to+\infty}\left(1+\frac{k}{x}\right)^x = e^k$$

si ottiene infine

$$\lim_{N\to+\infty}\phi_y(t) = e^{-\frac{t^2}{2}}$$

- La y è asintoticamente normale con media 0 e varianza 1;
- La S è asintoticamente normale con media $N\mu$ e varianza $N\sigma^2$;
- La $\bar{x}$ è asintoticamente normale con media μ e varianza $\dfrac{\sigma^2}{N}$.

★ Distribuzione asintotica delle medie dei campioni

★ Poor man's Gaussian random numbers generator

★ Il prodotto di molte variabili casuali indipendenti è asintoticamente log-normale.

Stime - I

Sia

$$x \sim f(x;\theta)$$

con θ parametro di valor vero (incognito) θ^*. Si definisce come funzione di stima (o, brevemente, *stima*) una funzione che permetta di attribuire un valore $\overline{\theta}$ al parametro θ a partire da N valori misurati $\underline{X}$:

$$\overline{\theta} = \overline{\theta}(\underline{X})$$

$\overline{\theta}$ è una variabile casuale, quindi si possono anche per essa definire media e varianza.

Stime - II

Proprietà delle stime

- Stime consistenti:

$$\lim_{N \to \infty} \overline{\theta} = \theta^*$$

- Stime indistorte:

$$E(\overline{\theta}) = \theta^*$$

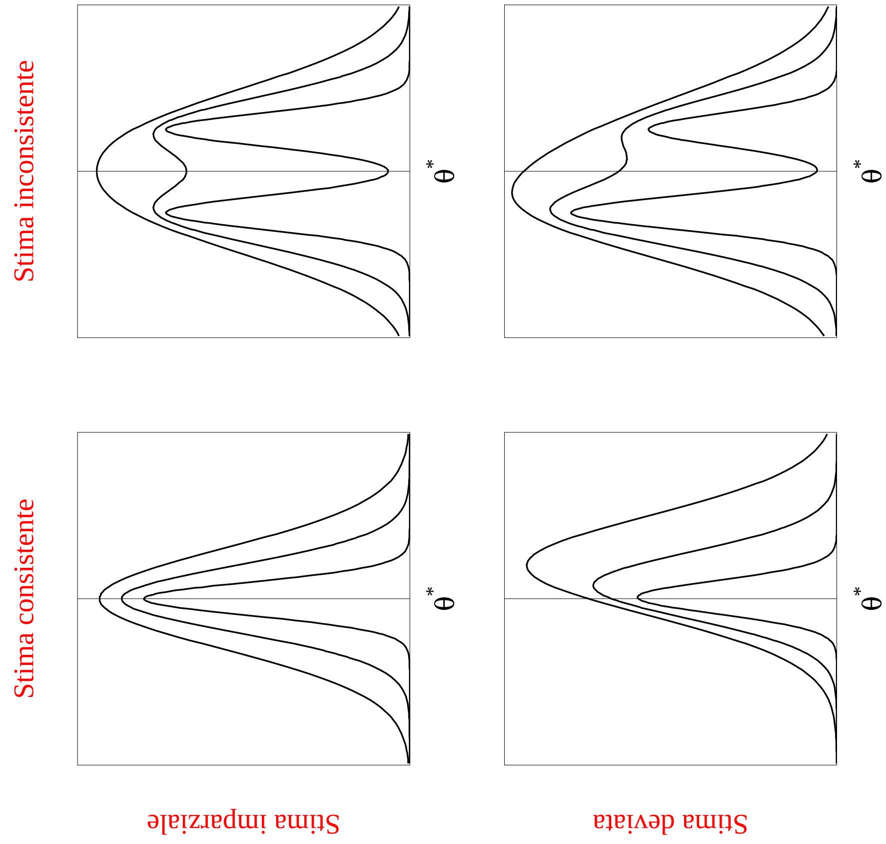
- Stime efficienti:
  - ⋆ Concetto relativo
  - ⋆ Il teorema di Cramér–Rao:
    - ▷ Limite superiore per l'efficienza delle stime
    - ▷ Condizioni per l'esistenza di una stima di efficienza massima

- Stime sufficienti: in generale

$$f(\theta_1, \theta_2, \ldots, \theta_M; \theta^*) =$$

$$= f^M(\theta_1; \theta^*) \cdot \varphi(\theta_2, \theta_3, \ldots, \theta_M; \theta^* | \theta_1)$$

La stima è sufficiente se φ non dipende da θ^*.



Stime - III

- Funzione di verosimiglianza (o di *likelihood*):

$$\mathcal{L}(\underline{X};\theta) = \prod_{i=1}^{N} f(x_i;\theta)$$

- Significato della funzione di verosimiglianza
- Stima di ***massima verosimiglianza***, $\hat{\theta}$:
  - ▷ E' sempre asintoticamente consistente.
  - ▷ E' asintoticamente normale.
  - ▷ E' sempre, asintoticamente, la stima più efficiente.
  - ▷ Se esiste una stima sufficiente, essa può essere espressa come funzione della sola $\hat{\theta}$.

Scuola di Dottorato in Fisica

Stime - IV

- E' spesso preferibile trovare il massimo del logaritmo della funzione di verosimiglianza, $\ln(\mathcal{L})$:

$$\ln(\mathcal{L}) = \sum_{i=1}^{N} \ln f(x_i;\theta)$$

- La normalità (asintotica) di $\hat{\theta}$ implica che la sua varianza sia approssimativamente data da

$$\mathrm{Var}(\hat{\theta}) = -\frac{1}{\left.\dfrac{d^2(\ln\mathcal{L})}{d\theta^2}\right|_{\theta=\hat{\theta}}}$$

La stima di massima verosimiglianza

E' asintoticamente

- consistente: se
  - ⋆ $f(x;\theta)$ è continua;
  - ⋆ $f(x;\theta)$ è dotata di derivata prima e seconda rispetto al parametro;
  - ⋆ se $\int dx_1 \cdot dx_2 \cdots dx_N$ commuta con $\frac{\partial}{\partial\theta}$ (ovvero se il dominio di definizione della x non dipende dal parametro).
- normale: se
  - ⋆ esistono i primi due momenti della f.
- di massima efficienza:
  - ⋆ sotto le stesse ipotesi necessarie per la consistenza.

Un esempio di stima sufficiente - I

Sia un campione di N determinazioni indipendenti x_i di una variabile distribuita secondo Poisson:

$$\Pr(x;a) = \frac{a^x}{x!}\,e^{-a}$$

$$\Pr(x_1,x_2,\ldots,x_N) =$$

$$= \frac{a^{x_1}}{x_1!}\,e^{-a}\cdot\frac{a^{x_2}}{x_2!}\,e^{-a}\cdots\frac{a^{x_N}}{x_N!}\,e^{-a}$$

$$= \frac{a^{\sum_i x_i}\cdot e^{-Na}}{x_1!\cdot x_2!\cdots x_N!}\cdot\frac{(N\bar{x})!}{(N\bar{x})!}$$

$$= \frac{a^{N\bar{x}}}{(N\bar{x})!}\,e^{-Na}\cdot\frac{(N\bar{x})!}{x_1!\cdot x_2!\cdots x_N!}\cdot\frac{N^{N\bar{x}}}{N^{N\bar{x}}}$$

$$= \left\{\frac{(Na)^{N\bar{x}}}{(N\bar{x})!}\,e^{-Na}\right\}\left\{\frac{(N\bar{x})!}{x_1!\cdot x_2!\cdots x_N!}\,\frac{1}{N^{N\bar{x}}}\right\}$$

- Il primo termine è la probabilità che la variabile

$$S = \sum_{i=1}^{N} x_i = N\bar{x}$$

abbia un certo valore; $\Pr(S)$ segue la statistica di Poisson (come noto). Si osservi che, essendo noto N, $\Pr(S)$ determina univocamente $\Pr(\bar{x})$.

Un esempio di stima sufficiente - II

- Il secondo termine **deve** allora essere la probabilità che i dati osservati valgano $x_1, x_2, \ldots, x_N$ condizionata all'avere la loro somma il valore S (o, il che è equivalente, condizionata all'avere la loro media il valore $\bar{x}$). **A posteriori** si vede che la statistica seguita è una multinomiale in cui i vari casi sono tutti **equiprobabili**: $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$.

$$\begin{aligned}
\Pr(x_1, x_2, \ldots, x_N; a) & \\
&\equiv \Pr(\bar{x}, x_1, x_2, \ldots, x_N; a) \\
&= \Pr(\bar{x}; a) \cdot \Pr(x_1, x_2, \ldots, x_N; a | \bar{x})
\end{aligned}$$

▷ Il secondo termine non dipende da a; qualunque sia $\bar{x}$, una volta noto il suo valore le x_i sono distribuite allo stesso modo (sono equiprobabili).

▷ Quindi $\bar{x}$ è una stima sufficiente di a.

Un esempio di stima sufficiente - III

$\bar{x}$ riassume la totalità delle informazioni contenute nei dati; non ha alcuna importanza quanto valgano le singole x_i: una volta che se ne conosca la somma (o il valor medio) questo è tutto quello che occorre per stimare a.

In effetti, se la densità di probabilità delle x_i condizionata dal valore di $\bar{x}$ non dipende dal parametro, questo implica che qualunque funzione dei dati ha densità (condizionata) che gode della stessa proprietà.

Teorema[a]: $\bar{\theta}$ è una stima sufficiente di θ se e solo se la funzione di verosimiglianza è fattorizzabile nella forma

$$\mathcal{L}(x_1, \ldots, x_N; \theta) = f(\bar{\theta}, \theta) \cdot \phi(x_1, \ldots, x_N)$$

[a] Eadie, Drijard, James, Roos, Sadoulet: Statistical methods in Experimental Physics, Pag. 96.

Il teorema di Cramér-Rao — I

Se

- $f(x;\theta)$ è definita in una regione dell'asse x indipendente da θ;
- esiste ovunque $\dfrac{\partial\,(\ln f)}{\partial\theta}$ ed esiste finito il valor medio del suo quadrato, $E\left\{\left[\dfrac{\partial\,(\ln f)}{\partial\theta}\right]^2\right\}$

allora una qualsiasi stima imparziale $\bar{\theta}$ di θ ha varianza che non può essere inferiore al limite di Cramér—Rao:

$$\mathrm{Var}(\bar{\theta}) \geq \frac{1}{N \cdot E\left\{\left[\dfrac{\partial\,(\ln f)}{\partial\theta}\right]^2\right\}}$$

Scuola di Dottorato in Fisica

Il teorema di Cramér-Rao — II

Il segno di uguaglianza vale se e solo se esiste una funzione $R(\theta)$ del solo parametro θ per la quale risulti

$$\frac{\partial\,(\ln \mathcal{L})}{\partial\theta} = \sum_{i=1}^{N} \frac{\partial\,(\ln f)}{\partial\theta} = \frac{\bar{\theta}(x_1, x_2, \ldots, x_N) - \theta}{R(\theta)}$$

e, in tal caso

- la stima di minima varianza **rende massima la funzione di verosimiglianza**;
- questa varianza minima vale $R(\theta)$.

La condizione è restrittiva, e si può dimostrare come implichi che la $f(x;\theta)$ sia di tipo *esponenziale*.

Esempio: stima del valor medio di una popolazione normale di varianza nota

$$f(x_i;\mu) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$\ln f = -\ln\sigma - \frac{1}{2}\ln(2\pi) - \frac{1}{2\sigma^2}(x_i-\mu)^2$$

$$\ln\mathcal{L} = -N\ln\sigma - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\frac{d\,(\ln\mathcal{L})}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) = \frac{N}{\sigma^2}\,(\bar{x}-\mu)$$

$$\frac{d^2\,(\ln\mathcal{L})}{d\mu^2} = -\frac{N}{\sigma^2} < 0$$

- La stima di massima verosimiglianza è la media del campione, $\hat{\mu} = \bar{x}$;
- *essa è (già lo sapevamo) consistente ed imparziale;*
- *essa è (per il teorema di Cramér–Rao) la stima avente la minima varianza, che vale*

  $$\mathrm{Var}(\hat{\mu}) = \frac{\mathrm{Var}(x)}{N} = \frac{\sigma^2}{N} = R(\theta)$$

  e quindi la stima di massima efficienza possibile;
- Inoltre $\bar{x}$ è una stima sufficiente di μ.

Scuola di Dottorato in Fisica

Esempio: stima della vita media - I

Nel processo di decadimento di una particella instabile, indichiamo con τ l'incognita vita media e con t i tempi (propri) di decadimento osservati, che seguono la distribuzione esponenziale:

$$f(t;\tau) = \frac{1}{\tau}\, e^{-\frac{t}{\tau}} \qquad \Rightarrow \qquad \begin{cases} E(t) = \tau \\ \mathrm{Var}(t) = \tau^2 \end{cases}$$

Ammettendo per l'osservazione una efficienza unitaria,

$$\mathcal{L} = \prod_{k=1}^{N} f(t_k;\tau) = \frac{1}{\tau^N}\, e^{-\frac{1}{\tau}\sum_k t_k}$$

$$\ln(\mathcal{L}) = -N\ln\tau - \frac{1}{\tau}\sum_{k=1}^{N} t_k = -N\left(\frac{\bar{t}}{\tau} + \ln\tau\right)$$

$$\frac{d\,\ln(\mathcal{L})}{d\tau} = \frac{N}{\tau^2}(\bar{t}-\tau) = 0 \qquad \Rightarrow \qquad \hat{\tau} = \bar{t}$$

$$\frac{d^2\,\ln(\mathcal{L})}{d\tau^2} = -\frac{N}{\tau^3}\,(2\bar{t}-\tau)$$

$$\left.\frac{d^2\,\ln(\mathcal{L})}{d\tau^2}\right|_{\tau=\bar{t}} < 0$$

Esempio: stima della vita media - II

La soluzione di massima verosimiglianza $\hat{\tau} = \bar{t}$ è

- consistente ed imparziale (essendo il valor medio del campione);
- di varianza minima (per il teorema di Cramér—Rao);
- questa varianza minima è inoltre data dalla

$$\mathrm{Var}(\hat{\theta}) = \frac{\mathrm{Var}(t)}{N} = \frac{\tau^2}{N} = R(\theta)$$

- sufficiente (riassume tutta l'informazione del campione).

Esempio: stima della vita media - III

Normalmente l'efficienza non è unitaria; ad esempio il nostro rivelatore può avere dimensioni confrontabili col cammino medio. In questo caso bisogna usare le probabilità condizionate:

$$\mathcal{L} = \prod_{i=1}^{N} \left[\frac{\frac{1}{\tau} e^{-\frac{t_i}{\tau}}}{e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}}} \right]$$

$$\ln(\mathcal{L}) = -N \ln \tau + \\ + \sum_{i=1}^{N} \left\{ -\frac{t_i}{\tau} - \ln \left[e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}} \right] \right\}$$

e, posto per brevità

$$\varphi_i(\tau) = \frac{(t_{\min})_i^2 \cdot e^{-\frac{(t_{\min})_i}{\tau}} - (t_{\max})_i^2 \cdot e^{-\frac{(t_{\max})_i}{\tau}}}{e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}}}$$

si arriva alla

$$\frac{d \ln(\mathcal{L})}{d\tau} = \frac{1}{\tau^2} \left\{ \sum_{i=1}^{N} \left[t_i - \varphi_i(\tau) \right] - N\tau \right\} = 0$$

che bisogna risolvere in modo numerico.

Esempio: stima della vita media - IV

- Non si può garantire che le proprietà precedentemente delineate per $\hat{\tau}$ (consistenza, normalità, efficienza, ...) siano ancora valide, almeno per N finito.
- Può darsi che la funzione di verosimiglianza ammetta più di un massimo, e non si sa a priori quale convergerà verso τ^*.
- L'errore della stima deve essere ricavato dalla concavità della funzione di verosimiglianza, supposta approssimativamente normale.

Stima del valor medio di una popolazione di Poisson

$$f(x_i; a) = \frac{a^{x_i}}{x_i!}\, e^{-a}$$

$$\ln\left[f(x_i; a)\right] = x_i \cdot \ln a - \ln(x_i!) - a$$

$$\ln(\mathcal{L}) = N\bar{x} \cdot \ln a - \sum_{i=1}^{N} \ln(x_i!) - Na$$

$$\frac{d\,\ln(\mathcal{L})}{da} = \frac{N\,\bar{x}}{a} - N = \frac{N}{a}(\bar{x} - a)$$

$$\frac{d^2\,\ln(\mathcal{L})}{da^2} = -\frac{N\,\bar{x}}{a^2} < 0$$

$$\hat{a} = \bar{x}$$

$$\mathrm{Var}(\hat{a}) = \frac{a}{N}$$

Stima di più parametri — I

Il metodo della massima verosimiglianza si generalizza facilmente alla stima contemporanea di più parametri:

$$x \sim f(x;\theta_1,\theta_2,\ldots,\theta_M) = f(x;\underline{\theta})$$

$$\mathcal{L}(\underline{X};\underline{\theta}) = \prod_{i=1}^{N} f(x_i;\underline{\theta})$$

$$\ln(\mathcal{L}) = \sum_{i=1}^{N} \ln\left[f(x_i;\underline{\theta})\right]$$

$$\theta_k \quad \longleftarrow \quad \begin{cases} \dfrac{\partial\mathcal{L}}{\partial\theta_k} = 0 & , \quad \dfrac{\partial^2\mathcal{L}}{\partial\theta_k^2} < 0 \\[2ex] \dfrac{\partial \ln(\mathcal{L})}{\partial\theta_k} = 0 & , \quad \dfrac{\partial^2 \ln(\mathcal{L})}{\partial\theta_k^2} < 0 \end{cases}$$

Scuola di Dottorato in Fisica

Stima di più parametri — II

- Le proprietà della stima di massima verosimiglianza rimangono valide:
  - ▷ Le stime $\hat{\underline{\theta}}$ sono asintoticamente consistenti;
  - ▷ La loro distribuzione è asintoticamente normale (in M dimensioni);
  - ▷ Le stime $\hat{\underline{\theta}}$ tendono asintoticamente alla massima efficienza;
  - ▷ Gli errori sono legati all'andamento di $\mathcal{L}$ nell'intorno del massimo:

$$\left(\underset{\sim}{V}^{-1}\right)_{ij} = -N \cdot E\left\{\frac{\partial^2 \ln f(x;\theta_1,\theta_2,\ldots,\theta_M)}{\partial\theta_i\,\partial\theta_j}\right\}$$

Stima di probabilità ignote

Siano M eventualità esaurienti e mutuamente esclusive; in N prove effettuate le frequenze assolute di ognuna di esse valgano n_i, con $i = 1, 2, \ldots, M$.

$$\mathcal{L} \propto \prod_{i=1}^{M} p_i^{n_i}$$

$$\ln(\mathcal{L}) = K + \sum_{i=1}^{M} n_i \cdot \ln(p_i)$$

con le p_i vincolate dalla $\sum_i p_i = 1$.

$$f(\underline{p}; \underline{n}) = \sum_{i=1}^{M} n_i \cdot \ln(p_i) - \lambda \left(\sum_{i=1}^{M} p_i - 1 \right)$$

$$\frac{\partial f}{\partial p_k} = \frac{n_k}{p_k} - \lambda = 0$$

$$\hat{p}_k = \frac{n_k}{\lambda}$$

$$\sum_{k=1}^{N} p_k = \frac{N}{\lambda} = 1 \quad \longrightarrow \quad \lambda = N$$

$$\hat{p}_k = \frac{n_k}{N}$$

Stima dei parametri della popolazione normale - I

Già sappiamo che

$$\ln \mathcal{L} = -N \ln \sigma - \frac{N}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2$$

Se σ è supposta ignota, le equazioni di massima verosimiglianza sono:

$$\begin{cases} \dfrac{\partial \ln(\mathcal{L})}{\partial \mu} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{N} (x_i - \mu) = 0 \\[2ex] \dfrac{\partial \ln(\mathcal{L})}{\partial \sigma} = \dfrac{1}{\sigma^3} \left[\displaystyle\sum_{i=1}^{N} (x_i - \mu)^2 - N\sigma^2 \right] = 0 \end{cases}$$

ed hanno l'unica soluzione (un massimo)

$$\begin{cases} \hat{\mu} = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} x_i \\[2ex] \hat{\sigma}^2 = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} (x_i - \hat{\mu})^2 \end{cases}$$

Stima dei parametri della popolazione normale - II

Entrambe le stime sono consistenti; la seconda non è imparziale, ma può essere resa tale:

$$\hat{\sigma}^2 \qquad \longrightarrow \qquad s^2 = \frac{N}{N-1}\,\hat{\sigma}^2$$

che è appunto imparziale ma non è di minima varianza:

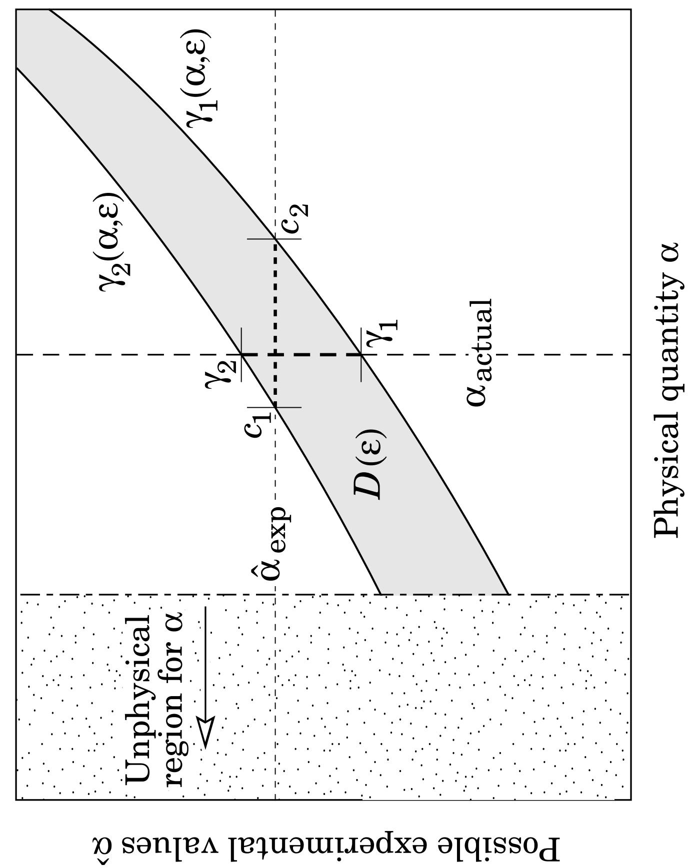
$$\mathrm{Var}(s^2) = \frac{N^2}{(N-1)^2}\,\mathrm{Var}(\hat{\sigma}^2) > \mathrm{Var}(\hat{\sigma}^2)$$

Il fatto che la varianza della popolazione abbia un determinato valore non cambia il fatto che la nostra migliore stima del valor medio della popolazione sia comunque data dalla media aritmetica del campione: valor medio e varianza del campione sono statisticamente indipendenti.

La matrice delle covarianze vale

$$\underset{\sim}{V} = \cdots = \left\| \begin{array}{cc} \dfrac{\hat{\sigma}^2}{N} & 0 \\ 0 & \dfrac{\hat{\sigma}^2}{2N} \end{array} \right\|$$

note espressioni della varianza di $\hat{\mu}$ e di $\hat{\sigma}$; la loro covarianza è ovviamente nulla.



Determinazione degli intervalli di confidenza per un parametro α da cui dipende una funzione di distribuzione nota di un esperimento.

La distribuzione del χ^2 - I

Come è distribuito il quadrato di una variabile casuale normale standardizzata?

$$\begin{cases} x \sim N(x;0,1) \\ y = x^2 \end{cases}$$

$$\phi_y(t) = E\left(e^{itx^2}\right)$$

$$= \int_{-\infty}^{+\infty} e^{itx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}(1-2it)}\, dx$$

e, posto $u^2 = x^2(1-2it)$,

$$\phi_y(t) = \frac{1}{\sqrt{1-2it}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$$

$$= (1-2it)^{-\frac{1}{2}}$$

Scuola di Dottorato in Fisica

La distribuzione del χ^2 - II

Generalizzando,

$$\begin{cases} x_i \sim N(0,1) \qquad (i = 1, 2, \ldots, N) \\ X = \sum_{i=1}^{N} x_i^2 \end{cases}$$

$$\phi_X(t) = (1-2it)^{-\frac{N}{2}}$$

$$f(X;N) = \frac{1}{2\Gamma\left(\frac{N}{2}\right)} \left(\frac{X}{2}\right)^{\frac{N}{2}-1} e^{-\frac{X}{2}}$$

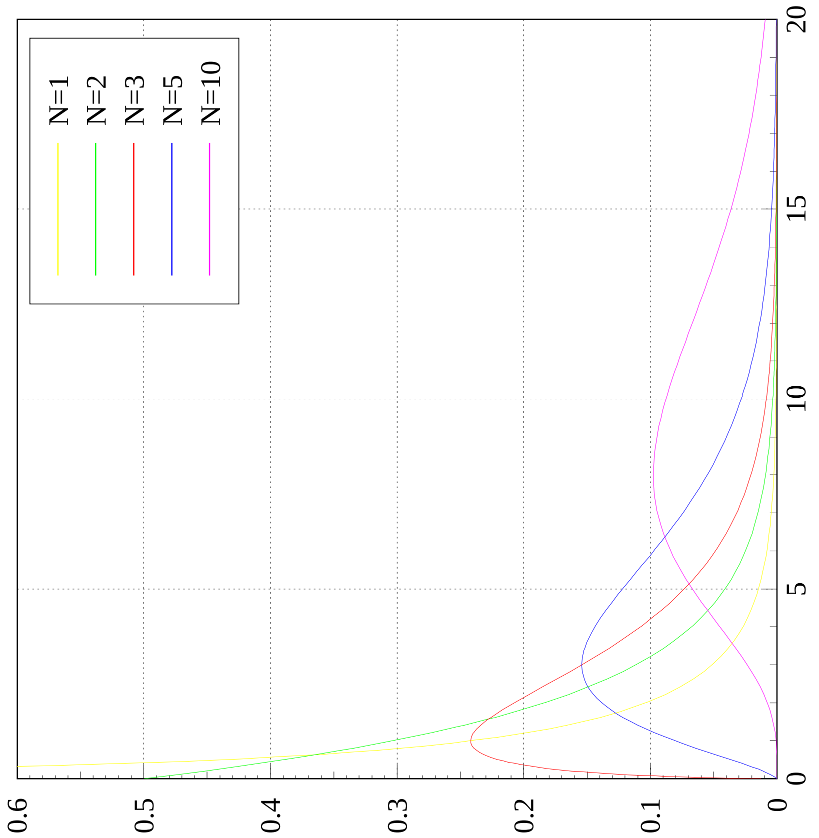
$$M_X(t) = (1-2t)^{-\frac{N}{2}}$$

$$E(X) = N$$

$$Var(X) = 2N$$

$$\frac{X-N}{\sqrt{2N}} \xrightarrow{N\to\infty} N(0,1)$$

$$\sqrt{2X} - \sqrt{2N-1} \xrightarrow{N\to\infty} N(0,1)$$



La distribuzione del χ^2 - III

- Regola di somma del χ^2.
- Variabili importanti distribuite come il χ^2:

★ $$X = \sum_{i=1}^{N} \left(\frac{x_i - \mu}{\sigma} \right)^2 \qquad \sim \chi^2(N)$$

★ $$X = \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \qquad \sim \chi^2(N-1)$$

★ $$X = (N-1)\,\frac{s^2}{\sigma^2} \qquad \sim \chi^2(N-1)$$

La distribuzione del χ^2 - IV

Consideriamo quella particolare trasformazione lineare $x_j \to y_i = \sum_j A_{ij} x_j$ in cui

$$\begin{cases} y_1 = \dfrac{1}{\sqrt{N}}\,(x_1 + x_2 + \cdots + x_N) \\ y_2 = \dfrac{1}{\sqrt{2}}\,(x_1 - x_2) \\ y_3 = \dfrac{1}{\sqrt{6}}\,(x_1 + x_2 - 2x_3) \\ \cdots \\ y_N = \dfrac{x_1 + x_2 + \cdots + x_{N-1} - (N-1)x_N}{\sqrt{N(N-1)}} \end{cases}$$

ossia

$$A_{ij} \equiv \begin{cases} i = 1: & \dfrac{1}{\sqrt{N}} \\ i > 1: & \begin{cases} j < i: & \dfrac{1}{\sqrt{i(i-1)}} \\ j = i: & -\dfrac{i-1}{\sqrt{i(i-1)}} \\ j > i: & 0 \end{cases} \end{cases}$$

Scuola di Dottorato in Fisica

La distribuzione del χ^2 - V

La matrice $\underset{\sim}{A}$ è ortogonale:

$$\sum_{i=1}^{N} A_{ij} A_{ik} = \delta_{jk} \qquad \Rightarrow \qquad \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2$$

ed è stata scelta in modo che

$$y_1 = \sqrt{N}\,\bar{x}$$

$$\sum_{j=1}^{N} A_{ij}^2 = 1 \qquad \forall i$$

$$\sum_{j=1}^{N} A_{ij} = 0 \qquad i > 1$$

per cui

$$\begin{cases} Var(y_i) = Var(x) = \sigma^2 & \forall i \\ E(y_i) = 0 & i > 1 \end{cases}$$

Insomma le y_i, per $i > 1$, sono variabili casuali *normali* con media 0 e varianza σ^2.

La distribuzione del χ^2 - VI

$$
\begin{aligned}
X &= \frac{1}{\sigma^2} \sum_{i=1}^{N} (x_i - \bar{x})^2 \\
&= \frac{1}{\sigma^2} \left(\sum_{i=1}^{N} {x_i}^2 - N\bar{x}^2 \right) \\
&= \frac{1}{\sigma^2} \left(N\bar{x}^2 + \sum_{i=2}^{N} {y_i}^2 - N\bar{x}^2 \right) \\
&= \sum_{i=2}^{N} \frac{{y_i}^2}{\sigma^2}
\end{aligned}
$$

X è distribuita come il χ^2 *a* $(N-1)$ *gradi di libertà.*

La regola generale è che il numero di gradi di libertà da associare a variabili che seguono la distribuzione del χ^2 è dato dal numero di contributi **indipendenti**: ovvero il numero di termini con distribuzione normale sommati in quadratura (qui N, uno per ogni determinazione x_i) diminuito del numero di parametri che compaiono nella formula e che sono stati **stimati dai dati** stessi (qui uno, la media della popolazione, stimata usando la media aritmetica delle misure).

Il test del χ^2

Dati raccolti in M classi di frequenza.

Ipotesi (ammessa vera la conoscenza della distribuzione):

- Classi di frequenza ristrette: $p_i \ll 1$, $p_i^2 \ll p_i$, in modo che la distribuzione (binomiale) nel singolo bin si possa approssimare con la statistica di Poisson:

$$\mathrm{Var}(A_i) = A_i$$

- Alta frequenza attesa: $A_i \gtrsim 5$, in modo che in ogni bin si possa ritenere valida l'approssimazione normale alla distribuzione di Poisson: pertanto

$$X = \sum_{i=1}^{M} \frac{(O_i - A_i)^2}{A_i} \qquad \sim \begin{cases} \chi^2(M-1) \\ \chi^2(M-R-1) \end{cases}$$

Se vi sono classi con frequenze troppo basse, si usano classi di ampiezza variabile in modo che $A_i \approx 5 \div 10$

Esercizio

In uno dei suoi esperimenti, l'abate Mendel osservò forma e colore dei frutti di molte piante di piselli, classificandole in quattro categorie come segue (O_i è qui il numero di piante osservate in ogni categoria):

i	Tipo	O_i
1	Rotondi e gialli	315
2	Rotondi e verdi	108
3	Oblunghi e gialli	101
4	Oblunghi e verdi	32

Mendel si aspettava un rapporto tra le popolazioni delle quattro categorie di $9:3:3:1$; i risultati sono in accordo con queste previsioni?

Soluzione

Il numero totale di osservazioni è $N = 556$.

i	p_i	$A_i = Np_i$	O_i	q_i	$V_i = Np_iq_i$
1	$\frac{9}{16}$	312.75	315	$\frac{7}{16}$	136.83
2	$\frac{3}{16}$	104.25	108	$\frac{13}{16}$	84.70
3	$\frac{3}{16}$	104.25	101	$\frac{13}{16}$	84.70
4	$\frac{1}{16}$	34.75	32	$\frac{15}{16}$	32.58

$$X = \sum_{i=1}^{4} \frac{(A_i - O_i)^2}{A_i} \approx 0.47$$

è distribuita come il χ^2 a 3 gradi di libertà: un valore superiore a quello osservato si presenta casualmente nel 92.54% dei casi.

$$X' = \sum_{i=1}^{4} \frac{(A_i - O_i)^2}{V_i} \approx 0.56$$

$$X' \sim \chi^2(X';3) \qquad \Rightarrow \qquad P \approx 90.56\%$$

Esercizio

Il Bortkewitch studiò il numero di morti per calci di cavallo nell'esercito prussiano, registrando i decessi verificatisi in 10 corpi d'armata nel corso di 20 anni (per un totale quindi di $N = 200$ casi). Le frequenze assolute n_i del numero di morti per corpo d'armata e per anno i sono riassunte nella tabella seguente:

i	n_i
0	109
1	65
2	22
3	3
4	1
Totale	200

Si chiede di verificare se i dati sono in accordo con la distribuzione di Poisson.

Soluzione

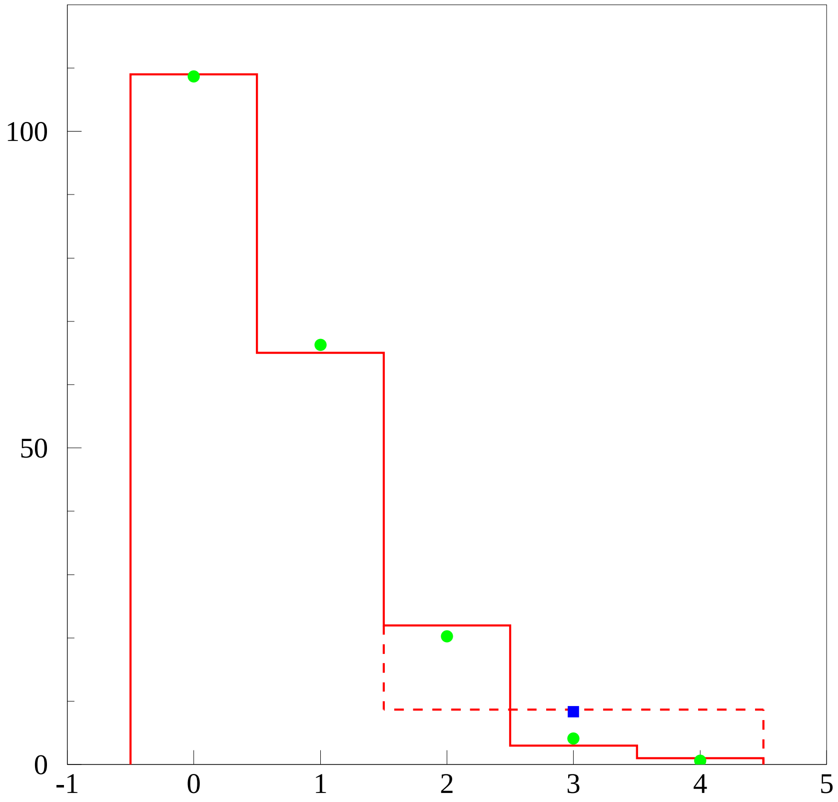
$$a = \frac{1}{N}\sum_{i=0}^{4} i \cdot n_i \approx 0.61$$

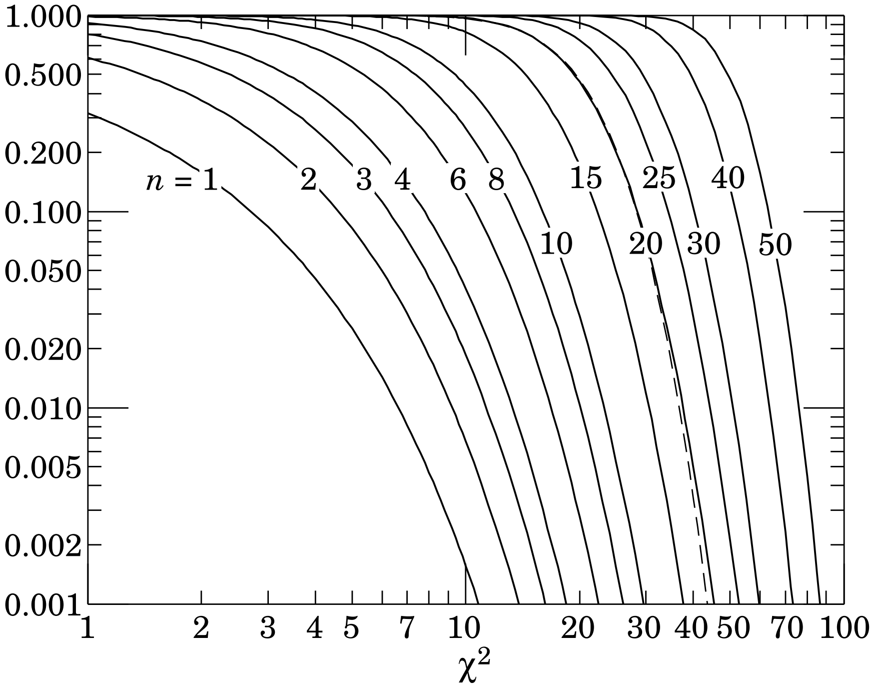
i	p_i	A_i	n_i	V_i
0	0.5434	108.67	109	49.62
1	0.3314	66.29	65	44.31
2	0.1011	20.22	22	
3	0.0206	4.11	3	
4	0.0031	0.63	1	
¿1	0.1252	25.04	26	21.90

$$X = \frac{(109-108.67)^2}{108.67} + \frac{(65-66.29)^2}{66.29} + \\ + \frac{(26-25.04)^2}{25.04} \approx 0.06291$$

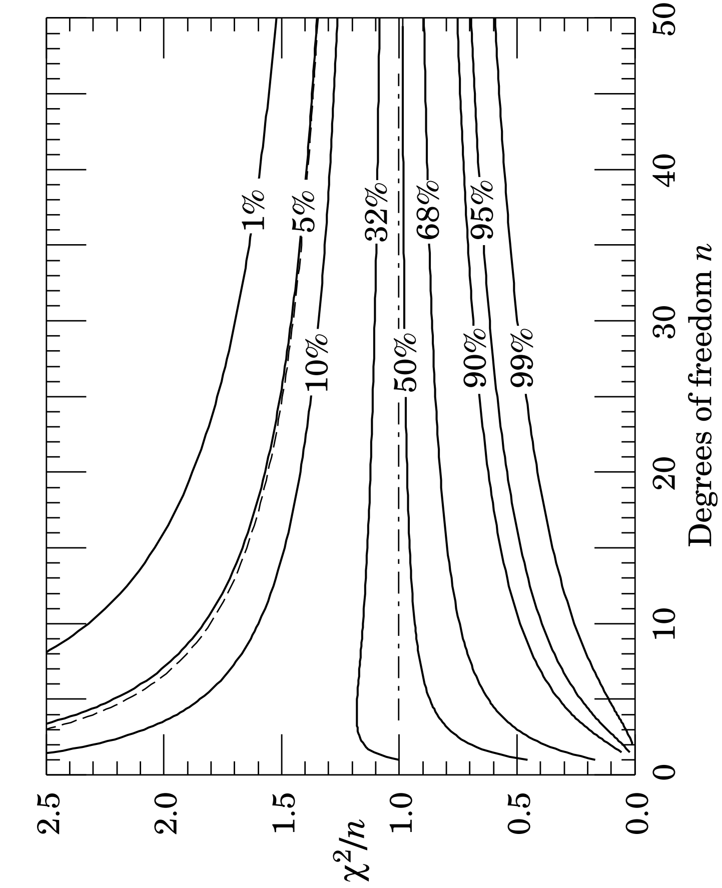
è distribuita come il χ^2 a 1 grado di libertà: la probabilità di ottenere per motivi puramente casuali un valore almeno pari a quello osservato è dell'80.35%.

$$X' \approx 0.07997 \qquad \Rightarrow \qquad P \sim 77.73\%$$





Il livello di confidenza come funzione del valore del χ^2 e del numero n di gradi di libertà.



Il livello di confidenza come funzione del χ^2 ridotto (χ^2/n), in corrispondenza di vari gradi di libertà n.

Il metodo del minimo χ^2 - I

$$\begin{cases} X = \sum_{i=1}^{M} \frac{(O_i - A_i)^2}{A_i} = \sum_{i=1}^{M} \frac{\left(n_i - Np_i\right)^2}{Np_i} \\ p_i = p_i(a_1, a_2, \ldots, a_R) \end{cases}$$

$$\frac{\partial X}{\partial a_k} =$$

$$- \sum_{i=1}^{M} \frac{2\left(n_i - Np_i\right) N^2 p_i + N\left(n_i - Np_i\right)^2}{N^2 {p_i}^2} \frac{\partial p_i}{\partial a_k}$$

$$-\frac{1}{2} \frac{\partial X}{\partial a_k} =$$

$$\sum_{i=1}^{M} \left[\frac{n_i - Np_i}{p_i} + \frac{\left(n_i - Np_i\right)^2}{2\,N{p_i}^2} \right] \frac{\partial p_i}{\partial a_k}$$

Il rapporto tra il secondo termine ed il primo vale

$$\frac{n_i - Np_i}{2\,Np_i} = \frac{1}{2p_i}\left(\frac{n_i}{N} - p_i\right) \quad \xrightarrow{N\to\infty} \quad 0$$

Il metodo del minimo χ^2 - II

Trascurando il secondo termine (lecito asintoticamente; metodo **semplificato** del minimo χ^2):

$$\sum_{i=1}^{M} \left(\frac{n_i - Np_i}{p_i} \right) \frac{\partial p_i}{\partial a_k} = 0$$

Se i bins coprono **tutti** i valori permessi,

$$\sum_{i=1}^{M} \left(\frac{n_i - Np_i}{p_i} \right) \frac{\partial p_i}{\partial a_k} =$$

$$= \sum_{i=1}^{M} \frac{n_i}{p_i} \frac{\partial p_i}{\partial a_k} - N \sum_{i=1}^{M} \frac{\partial p_i}{\partial a_k}$$

e l'ultimo termine si annulla, essendo

$$\sum_{i=1}^{M} \frac{\partial p_i}{\partial a_k} = \frac{\partial}{\partial a_k} \sum_{i=1}^{M} p_i = \frac{\partial}{\partial a_k} 1 \equiv 0$$

quindi

$$\sum_{i=1}^{M} \frac{n_i}{p_i} \frac{\partial p_i}{\partial a_k} = 0 \qquad (k = 1, 2, \ldots, R)$$

Scuola di Dottorato in Fisica

Il metodo del minimo χ^2 - III

Applicando invece il metodo della massima verosimiglianza:

$$\mathcal{L}(a_1, a_2, \ldots, a_R) \propto \prod_{i=1}^{M} p_i^{n_i}$$

$$\ln(\mathcal{L}) = K + \sum_{i=1}^{M} \left(n_i \cdot \ln p_i \right)$$

$$\frac{\partial \ln(\mathcal{L})}{\partial a_k} = \sum_{i=1}^{M} \frac{n_i}{p_i} \frac{\partial p_i}{\partial a_k} = 0 \qquad (k = 1, 2, \ldots, R)$$

I due metodi (della massima verosimiglianza e del χ^2 semplificato) conducono **asintoticamente** allo stesso risultato.

Test di omogeneità — I

Siano Q campioni indipendenti di dati, composti da $n_1, n_2, \ldots, n_Q$ valori ciascuno; e, all'interno di ogni campione, i dati siano divisi tra P classi esaurienti e mutuamente esclusive; consideriamo il problema di verificare l'ipotesi che i dati provengano dalla stassa distribuzione.

	Campioni					
Gruppi	ν_{11}	ν_{12}	ν_{13}	$\cdots$	ν_{1Q}	m_1
	ν_{21}	ν_{22}	$\cdots$	$\cdots$	ν_{2Q}	m_2
	ν_{31}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	m_3
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	ν_{P1}	ν_{P2}	$\cdots$	$\cdots$	ν_{PQ}	m_P
	n_1	n_2	n_3	$\cdots$	n_Q	N

Ammessa vera l'ipotesi, si dimostra che

$$X = N \left[\sum_{i,j} \frac{\nu_{ij}^2}{m_i \cdot n_j} - 1 \right] \qquad \sim \chi^2(X;K)$$

$$K = (P-1)(Q-1)$$

Test di omogeneità — II

Sia p_i, con $i = 1, 2, \ldots, P$, la probabilità che un dato scelto a caso appartenga alla i-esima classe; e sia $q_j (j = 1, 2, \ldots, Q)$ la probabilità di appartenenza al campione j-esimo. Dall'insieme dei dati possiamo ricavare $(P-1)$ stime indipendenti delle p_i e $(Q-1)$ stime indipendenti delle q_j:

$$p_i = \frac{m_i}{N} \qquad e \qquad q_j = \frac{n_j}{N}$$

La variabile casuale

$$\begin{aligned} X &= \sum_{i,j} \frac{\left(\nu_{ij} - N p_i q_j\right)^2}{N p_i q_j} \\ &= \sum_{i,j} \left[\frac{\left(\nu_{ij}\right)^2}{N p_i q_j} - 2\nu_{ij} + N p_i q_j \right] \\ &= \sum_{i,j} \frac{\left(\nu_{ij}\right)^2}{N p_i q_j} - 2N + N \\ &= \sum_{i,j} \frac{\left(\nu_{ij}\right)^2}{N p_i q_j} - N \end{aligned}$$

deve essere distribuita come il χ^2.

Test di omogeneità — III

Sostituendo le espressioni per i valori stimati delle p_i e delle q_j, si ottiene

$$X = N \left[\sum_{i,j} \frac{\nu_{ij}^2}{m_i \cdot n_j} - 1 \right]$$

Il numero di gradi di libertà è dato dal numero di contributi (PQ), diminuito di 1 per tener conto del vincolo sul numero totale di dati, e di $(P-1)+(Q-1)$ che è il numero dei parametri indipendenti stimati dai dati stessi; e vale quindi

$$\begin{aligned} K &= PQ - 1 - (P-1) - (Q-1) \\ &= PQ - P - Q + 1 \\ K &= (P-1)(Q-1) \end{aligned}$$

Il test di Kolmogorov e Smirnov

Confronto tra campione e distribuzione:

$$\delta = \max\left\{ \left| F(x) - \Phi(x) \right| \right\}$$

$$\Pr(\delta \geq \delta_0) = F_{KS}(\delta_0')$$

$$F_{KS}(x) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2}$$

$$\delta_0' = \left(\sqrt{N} + 0.12 + \frac{0.11}{\sqrt{N}} \right) \delta_0$$

Confronto tra due campioni:

$$\delta = \max\left\{ \left| F_1(x) - F_2(x) \right| \right\}$$

$$N = \frac{1}{\frac{1}{N_1} + \frac{1}{N_2}} = \frac{N_1 N_2}{N_1 + N_2}$$

- Non richiede la definizione (in genere arbitraria) delle classi di frequenza;
- E' (relativamente) indipendente dalla variabile casuale usata;
- E' poco sensibile nelle code e molto nella zona mediana;
- *Nessun* parametro deve essere stato stimato dai dati.

Esercizio

Per una ricerca statistica si è misurato il peso di 345 bambini e 326 bambine di 11 anni; le frequenze osservate, per classi ampie 2.5 Kg, sono riportate nella tabella seguente:

Peso	M	F	Peso	M	F	Peso	M	F
40.0-42.5	1	0	52.5-55.0	80	59	65.0-67.5	4	16
42.5-45.0	3	1	55.0-57.5	72	58	67.5-70.0	2	5
45.0-47.5	9	7	57.5-60.0	41	48	70.0-72.5	1	3
47.5-50.0	33	37	60.0-62.5	27	23	72.5-75.0	0	2
50.0-52.5	65	41	62.5-65.0	7	26			

Le osservazioni sono compatibili con popolazioni normali? E sono compatibili tra di loro?

Soluzione - I

$\mu_M = 54.67$ $\qquad$ $\mu_F = 56.31$

$\sigma_M = 4.497$ $\qquad$ $\sigma_F = 5.631$

Bambini: raggruppiamo i dati $P < 47.5$ e $P > 62.5$; $A_i(> 65) = 3.73$.

Peso	M: P_i^M	M: A_i^M	M: O_i^M
¡47.5	0.055	19.121	13
47.5-50.0	0.094	32.465	33
50.0-52.5	0.165	56.989	65
52.5-55.0	0.214	74.016	80
55.0-57.5	0.206	71.131	72
57.5-60.0	0.147	50.581	41
60.0-62.5	0.077	26.611	27
¿62.5	0.041	14.086	14

$$X_M = \sum_{i=1}^{8} \frac{(A_i^M - O_i^M)^2}{A_i^M} \approx 5.41 \sim \chi^2(5)$$

$$\Pr(X > X_M) \approx 36.8\%$$

Soluzione - II

Bambine: raggruppando $P > 65$

Peso	F		
	P_i^F	A_i^F	O_i^F
¡47.5	0.059	19.183	8
47.5-50.0	0.072	23.599	37
50.0-52.5	0.118	38.498	41
52.5-55.0	0.159	51.734	59
55.0-57.5	0.176	57.267	58
57.5-60.0	0.160	52.218	48
60.0-62.5	0.120	39.222	23
62.5-65.0	0.074	24.267	26
¿65.0	0.061	20.012	26

$$X_F = \sum_{i=1}^{9} \frac{(A_i^F - O_i^F)^2}{A_i^F} \approx 24.287 \sim \chi^2(6)$$

$\Pr(X > X_F) \approx 0.05\%$

Soluzione - III

Tabella delle contingenze: raggruppiamo $P < 47.5$ e $P > 65$

13	33	65	80	72	41	27	7	7	345
8	37	41	59	58	48	23	26	26	326
21	70	106	139	130	89	50	33	33	671

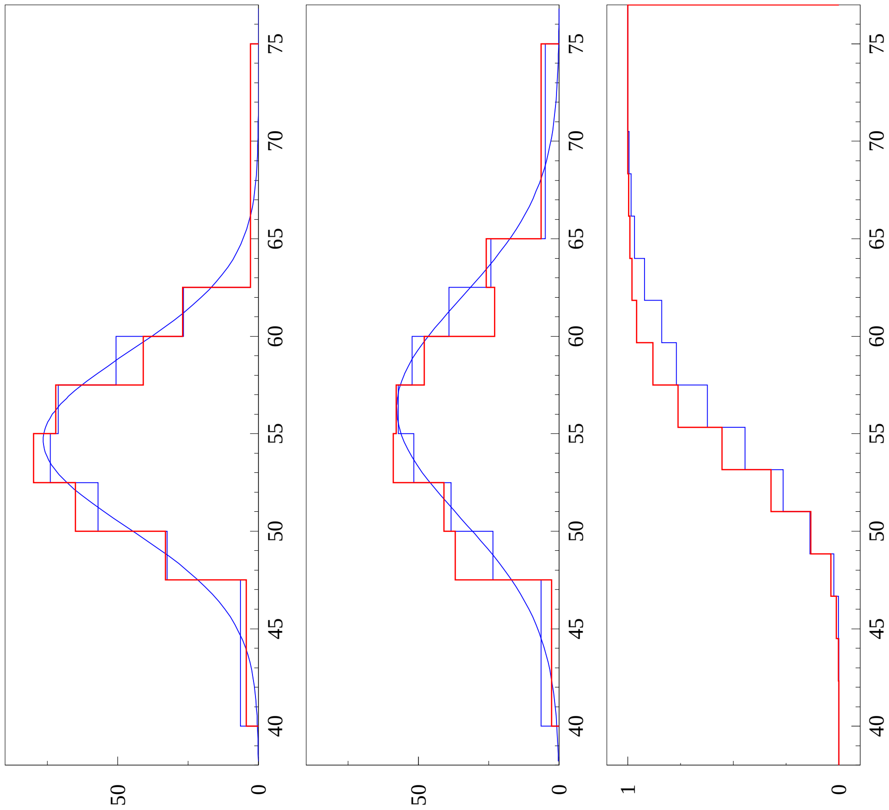
$X \approx 33.77 \sim \chi^2(8)$

$\Pr(\chi^2 > X) \approx 0$

Kolmogorov & Smirnov:

$\delta_{obs} = 0.1396 \qquad (P \in [55, 57.5])$

$\Pr(\delta > \delta_{obs}) \approx 0.25\%$



Test basati sulla curva normale

- Compatibilità del valore misurato con un numero prefissato
- Tests two-tailed e tests one-tailed:

$$u = \frac{\bar{x} - \xi}{\sqrt{\frac{\mathrm{Var}(x)}{N}}} \qquad \sim N(u; 0, 1)$$

  se l'ipotesi è vera.
- Compatibilità di due valori misurati:

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\mathrm{Var}(x)}{N_x} + \frac{\mathrm{Var}(y)}{N_y}}} \qquad \sim N(u; 0, 1)$$

  se l'ipotesi è vera.
- Limiti della validità del test; i piccoli campioni.

Esercizio

Su N campioni scelti a caso dalla lavorazione di un conduttore di rame si misura la forza F_i (in Kg-peso; $i = 1, \ldots, N$) necessaria per spezzare il cavo stesso. Valor medio ed errore quadratico medio dei valori F_i valgono rispettivamente (sempre in Kg-peso)

$$\mu = 261.1 \qquad e \qquad \sigma = 3.8$$

Supposto che la popolazione delle F segua la legge normale di distribuzione, e che la dimensione N del campione ci consenta di ottenere una buona stima della varianza della popolazione, a quale forza i cavi resisteranno con un livello di confidenza $P = 99.99\%$?

Soluzione — I

Ammettendo di conoscere valor medio F^* e scarto quadratico medio σ^* della popolazione delle F, per ipotesi

$$\frac{dp}{dF} = N(F; F^*, \sigma^*)$$

e si chiede di stimare un valore F_0 tale che

$$\int_{F_0}^{+\infty} dF \, \frac{1}{\sigma^* \sqrt{2\pi}} \, e^{-\frac{1}{2}\left(\frac{F-F^*}{\sigma^*}\right)^2} \; = \; P$$

Sempre per ipotesi, $\sigma^* = \sigma$; se fosse anche $F^* = \mu$, la risposta (one-tailed test all'uno per diecimila) sarebbe

$$F \; = \; \mu - 3.71\,\sigma \; \approx \; 246.9 \text{ Kg}$$

F^* non è nota; però si sa che è distribuita normalmente con densità di probabilità

$$\frac{dp}{dF^*} = N\left(F^*; \mu, \sigma_\mu\right) \qquad \text{ove} \qquad \sigma_\mu = \frac{\sigma}{\sqrt{N}}$$

Soluzione — II

$$\frac{dp}{dF} =$$

$$= \int_{-\infty}^{+\infty} dF^* \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(F-F^*)^2}{2\sigma^2}} \frac{1}{\sigma_\mu\sqrt{2\pi}} e^{-\frac{(F^*-\mu)^2}{2\sigma_\mu^2}}$$

che è la convoluzione delle due funzioni $N(F; 0, \sigma)$ e $N(F^*; \mu, \sigma_\mu)$. La trasformata di Fourier della convoluzione di due funzioni è il prodotto delle trasformate, per cui

$$\phi_F(t) = e^{-\frac{1}{2}\sigma^2 t^2} \cdot e^{\left(it\mu - \frac{1}{2}\sigma_\mu^2 t^2\right)}$$

$$= e^{\left[it\mu - \frac{t^2}{2}\left(\sigma^2 + \sigma_\mu^2\right)\right]}$$

che è la funzione caratteristica di una distribuzione ancora normale, con valor medio μ e varianza

$$\sigma^2 + \sigma_\mu^2 = \frac{N+1}{N}\,\sigma^2$$

per cui la risposta è

$$F_0 = \mu - 3.71\,\sigma\sqrt{\frac{N+1}{N}}$$

La distribuzione di Student - I

$$t = \frac{u}{\sqrt{\frac{X}{N}}} \quad \longleftarrow \quad \begin{cases} u \sim N(u; 0, 1) \\ X \sim \chi^2(X; N) \end{cases}$$

(statisticamente indipendenti)

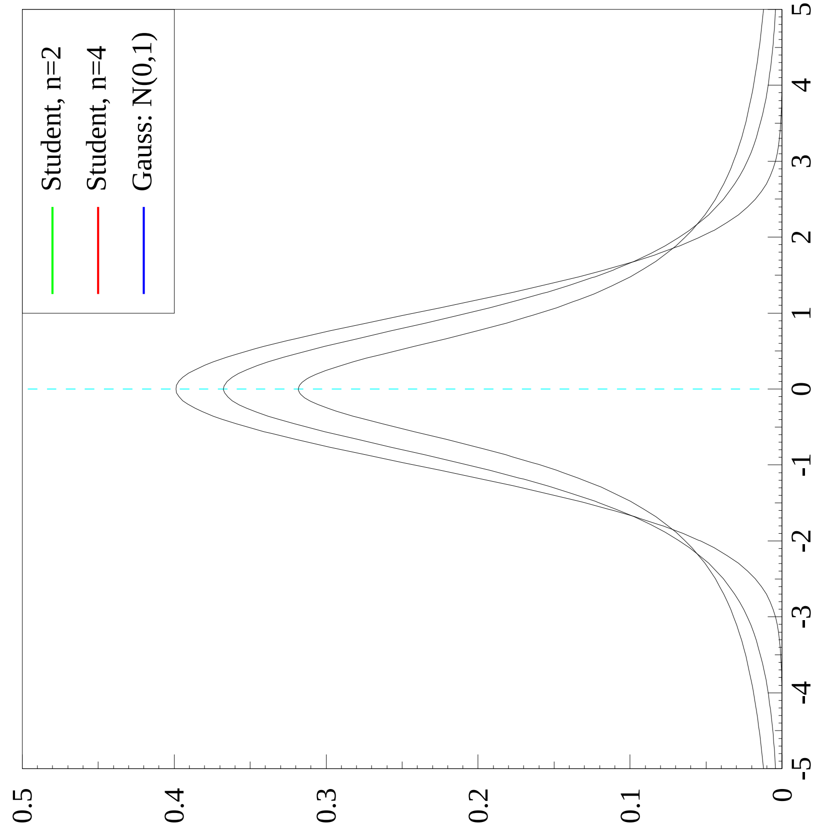
$$f(t) = \frac{dp}{dt} = \frac{\Gamma\left(\frac{N+1}{2}\right)}{\sqrt{N\pi} \cdot \Gamma\left(\frac{N}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{N}\right)^{\frac{N+1}{2}}}$$

$$E(t) = 0$$

$$Var(t) = \frac{N}{N-2} \qquad (N > 2)$$

$$f(t) \xrightarrow{N \to \infty} N(t; 0, 1)$$

Per $N = 1$ la distribuzione di Student si riduce a quella di Cauchy.



La distribuzione di Student - II

$$u = \frac{\bar{x} - E(x)}{\frac{\sigma}{\sqrt{N}}} \qquad \sim N(u; 0, 1)$$

$$X = (N-1)\frac{s^2}{\sigma^2} \qquad \sim \chi^2(X; N-1)$$

$$x = \frac{u}{\sqrt{\frac{X}{N-1}}}$$

$$= \frac{\bar{x} - E(x)}{\frac{s}{\sqrt{N}}} \qquad \sim t(x; N-1)$$

Compatibilità tra valore misurato e numero predeterminato per piccoli campioni: ammessa vera l'ipotesi,

$$x = \frac{\bar{x} - \xi}{\frac{s}{\sqrt{N}}} \qquad \sim t(x; N-1)$$

La distribuzione di Student - III

Compatibilità tra due valori misurati per piccoli campioni, ammesso che abbiano la stessa varianza σ^2:

$$S^2 = \frac{\sum_i \left[x_i - \bar{x}\right]^2 + \sum_k \left[y_k - \bar{y}\right]^2}{N+M-2} = \frac{(N-1)\, s_x^2 + (M-1)\, s_y^2}{N+M-2}$$

$$X = \frac{(N-1)\, s_x^2 + (M-1)\, s_y^2}{\sigma^2} = (N+M-2)\,\frac{S^2}{\sigma^2} \qquad \sim \chi^2(N+M-2)$$

$$u = \frac{(\bar{x}-\bar{y}) - [E(x)-E(y)]}{\sqrt{\sigma^2\left(\frac{1}{N}+\frac{1}{M}\right)}} \qquad \sim N(0,1)$$

$$t = \frac{u}{\sqrt{\frac{X}{N+M-2}}} \qquad \sim t(N+M-2)$$

e, ammessa vera l'ipotesi

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{N}+\frac{1}{M}\right)}} \qquad \sim t(N+M-2)$$

Scuola di Dottorato in Fisica

La distribuzione di Fisher - I

$$X \sim \chi^2(M) \qquad\qquad Y \sim \chi^2(N)$$

(statisticamente indipendenti)

$$w = \frac{\frac{X}{M}}{\frac{Y}{N}}$$

$$f(w;M,N) = \frac{\Gamma\left(\frac{M+N}{2}\right)\left(\frac{M}{N}\right)^{\frac{M}{2}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{N}{2}\right)} \cdot \frac{w^{\frac{M}{2}-1}}{\left(1+\frac{M}{N}w\right)^{\frac{M+N}{2}}}$$

$$E(w) = \frac{N}{N-2} \qquad N > 2$$

$$Var(w) = \frac{2\,N^2(M+N-2)}{M(N-2)^2(N-4)} \qquad N > 4$$

La distribuzione di Fisher - II

$$f(w; M, \infty) = \frac{1}{M}\,\chi^2(w; M)$$

$$f(w; \infty, N) = N\,\frac{1}{\chi^2(w; N)}$$

$$u \sim t(u; N) \qquad \Rightarrow \qquad u^2 \sim f(u^2; 1, N)$$

Per due campioni indipendenti

$$X = (M-1)\frac{s_1^2}{\sigma_1^2} \qquad\qquad Y = (N-1)\frac{s_2^2}{\sigma_2^2}$$

$$w = \frac{s_1^2}{s_2^2}\,\frac{\sigma_2^2}{\sigma_1^2} \qquad \sim f(w; M-1, N-1)$$

Test dell'ipotesi $\sigma_1 = \sigma_2$: ammessa vera l'ipotesi stessa,

$$w = \frac{s_1^2}{s_2^2} \qquad\qquad \sim f(w; M-1, N-1)$$

Scuola di Dottorato in Fisica

La verifica delle ipotesi

- Ipotesi nulla ed ipotesi alternativa
  - ⋆ Non necessariamente complementari
- La regione di rigetto, $\mathcal{R}$
- La regione di accettanza, $\mathcal{A} = S - \mathcal{R}$
- I possibili errori:
  - ⋆ Rigettare un'ipotesi vera (errori di prima specie):

    $$P_I = \Pr\big(E \in \mathcal{R} | H_0\big)$$

  - ⋆ Accettare un'ipotesi falsa (errori di seconda specie):

    $$P_{II} = \Pr\big(E \in \mathcal{A} | H_a\big)$$

- Significanza del criterio, $\quad a = 1 - P_I$
- Potenza del criterio, $\quad \beta = 1 - P_{II}$

Il criterio del **rapporto delle verosimiglianze**:

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\underline{x}|H_0)}{\mathcal{L}(\underline{x}|H_a)} < k \right\}$$

Primo esempio — I

$$\begin{cases} x \ \sim N(x; \mu, \sigma) & \text{(con } \sigma > 0 \text{ noto)} \\ H_0 \equiv \{\mu = 0\} \\ H_a \equiv \{\mu = 1\} \end{cases}$$

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$\ln f(x_i; \mu, \sigma) = -\ln\sigma - \ln\left(\sqrt{2\pi}\right) - \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\ln \mathcal{L} = -N\ln\sigma - N\ln\left(\sqrt{2\pi}\right) - \frac{\sum_i (x_i - \mu)^2}{2\sigma^2}$$

$$= -N\ln\sigma - N\ln\left(\sqrt{2\pi}\right) - \frac{\sum_i \left(x_i^2 - 2\mu x_i + \mu^2\right)}{2\sigma^2}$$

$$= -N\ln\sigma - N\ln\left(\sqrt{2\pi}\right) - \frac{\sum_i x_i^2 - 2N\bar{x}\mu + N\mu^2}{2\sigma^2}$$

Scuola di Dottorato in Fisica

Primo esempio — II

$$\ln\lambda = \ln\mathcal{L}(\underline{x}; 0, \sigma) - \ln\mathcal{L}(\underline{x}; 1, \sigma)$$

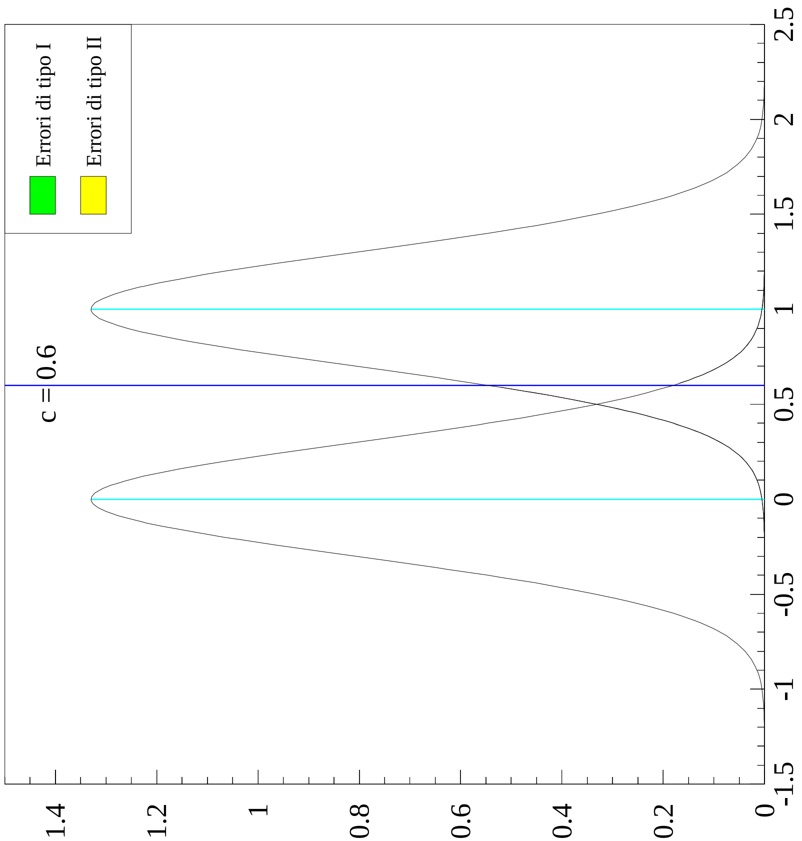
$$= \frac{N}{2\sigma^2}\,(1 - 2\bar{x})$$

$$\mathcal{R}_k \equiv \left\{ \ln\lambda = \frac{N}{2\sigma^2}\,(1 - 2\bar{x}) < \ln k \right\}$$

$$\equiv \left\{ \bar{x} > \frac{N - 2\sigma^2 \ln k}{2N} = c \right\}$$

$$P_I = \int_c^{+\infty} N\left(x; 0, \frac{\sigma}{\sqrt{N}}\right) dx$$

$$P_{II} = \int_{-\infty}^{c} N\left(x; 1, \frac{\sigma}{\sqrt{N}}\right) dx$$



Il lemma di Neyman-Pearson — I

Lemma: Se si ha a disposizione un campione di N valori indipendenti x_i da utilizzare per discriminare tra un'ipotesi nulla ed un'ipotesi alternativa entrambe semplici (ovvero del tipo $\theta = \theta_0$ e $\theta = \theta_a$), a parità di significanza α, la massima potenza del test (ovvero la minima probabilità di commettere errori di seconda specie) si ottiene definendo la regione di rigetto $\mathcal{R}_k$ attraverso una relazione del tipo:

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\underline{x}|H_0)}{\mathcal{L}(\underline{x}|H_a)} < k \right\}$$

Il lemma di Neyman-Pearson — II

Dobbiamo trovare tra tutte le regioni di rigetto $\mathcal{R}$ che soddisfano alla

$$\int_{\mathcal{R}} \mathcal{L}(\underline{x};\theta_0)\, d\underline{x} \;=\; P_I \;=\; 1-a$$

quella che rende massima

$$\beta = \int_{\mathcal{R}} \mathcal{L}(\underline{x};\theta_a)\, d\underline{x}$$

e dimostrare che è

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\underline{x}|\theta_0)}{\mathcal{L}(\underline{x}|\theta_a)} < k \right\}$$

Ora, per qualunque $\mathcal{R}$ e qualunque funzione $\phi(\underline{x})$,

$$\mathcal{R}_k = (\mathcal{R}_k \cap \mathcal{R}) \cup (\mathcal{R}_k \cap \overline{\mathcal{R}})$$

$$\mathcal{R} = (\mathcal{R} \cap \mathcal{R}_k) \cup (\mathcal{R} \cap \overline{\mathcal{R}}_k)$$

$$\int_{\mathcal{R}_k} \phi(\underline{x})\, d\underline{x} - \int_{\mathcal{R}} \phi(\underline{x})\, d\underline{x} =$$

$$= \int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \phi(\underline{x})\, d\underline{x} - \int_{\mathcal{R} \cap \overline{\mathcal{R}}_k} \phi(\underline{x})\, d\underline{x}$$

Il lemma di Neyman-Pearson — III

$$\int_{\mathcal{R}_k} \mathcal{L}(\underline{x}|\theta_a)\, d\underline{x} - \int_{\mathcal{R}} \mathcal{L}(\underline{x}|\theta_a)\, d\underline{x} =$$

$$= \int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \mathcal{L}(\underline{x}|\theta_a)\, d\underline{x} - \int_{\mathcal{R} \cap \overline{\mathcal{R}}_k} \mathcal{L}(\underline{x}|\theta_a)\, d\underline{x}$$

$$> \frac{1}{k} \left[\int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \mathcal{L}(\underline{x}|\theta_0)\, d\underline{x} - \int_{\mathcal{R} \cap \overline{\mathcal{R}}_k} \mathcal{L}(\underline{x}|\theta_0)\, d\underline{x} \right]$$

$$= \frac{1}{k} \left[\int_{\mathcal{R}_k} \mathcal{L}(\underline{x}|\theta_0)\, d\underline{x} - \int_{\mathcal{R}} \mathcal{L}(\underline{x}|\theta_0)\, d\underline{x} \right]$$

$$= \;\frac{1}{k} \cdot \left[P_I - P_I \right] \;=\; 0$$

Secondo esempio

$$\begin{cases} x \sim N(x;\mu,\sigma) & \text{(con } \sigma > 0 \text{ noto)} \\ H_0 \equiv \{\mu = 0\} \\ H_a \equiv \{\mu > 0\} \end{cases}$$

$$\ln \mathcal{L}(\underline{x};0,\sigma) = -N \ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N} {x_i}^2$$

$$\ln \mathcal{L}(\underline{x};\mu_a,\sigma) = -N \ln\left(\sigma\sqrt{2\pi}\right) - \\ - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{N} {x_i}^2 - 2N\bar{x}\mu_a + N{\mu_a}^2\right)$$

$$\ln\lambda = \ln \mathcal{L}(\underline{x};0,\sigma) - \mathcal{L}(\underline{x};\mu_a,\sigma)$$

$$= \frac{N\mu_a}{2\sigma^2}\left(\mu_a - 2\bar{x}\right) < \ln k$$

$$\mathcal{R}_k \equiv \left\{\bar{x} > \frac{N{\mu_a}^2 - 2\sigma^2 \ln k}{2N\mu_a} = c\right\}$$

$$P_I = \int_c^{+\infty} N\left(x;0,\frac{\sigma}{\sqrt{N}}\right) dx$$

Il rapporto delle massime verosimiglianze

Sia $\mathcal{L} = \mathcal{L}(\underline{x};\theta_1,\theta_2,\ldots,\theta_M)$.

- Se H_0 consiste nell'appartenere $P \leq M$ dei parametri θ_i ad una regione $\mathcal{Q}$ dello spazio dei parametri, e H_a nel non appartenervi;
- Se $\mathcal{L}(\hat{S})$ è il massimo valore della funzione di verosimiglianza nell'intero spazio dei parametri;
- Se $\mathcal{L}(\hat{R})$ è il massimo valore della funzione di verosimiglianza in $\mathcal{Q}$;

si può usare come criterio per definire $\mathcal{R}_k$

$$\mathcal{R}_k \equiv \left\{\lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} < k\right\}$$

Se si conosce la distribuzione di λ ammessa vera H_0 si può calcolare la significanza del test:

$$P_I = \Pr\left(\lambda \in [0,k] | H_0\right) = \int_0^k g(\lambda|H_0)\, d\lambda$$

e si può dimostrare che comunque

$$-2\ln\lambda \xrightarrow{N\to\infty} \chi^2(P)$$

Terzo esempio

$$\begin{cases} x \sim N(x;\mu,\sigma) & \text{(con } \sigma > 0 \text{ noto)} \\ H_0 \equiv \{\mu = 0\} & \\ H_a \equiv \{\mu \neq 0\} & \end{cases}$$

$$\ln \mathcal{L} = -N \ln(\sigma\sqrt{2\pi}) - \frac{\sum_i x_i^2 - 2N\bar{x}\mu + N\mu^2}{2\sigma^2}$$

$$\ln \mathcal{L}(\hat{S}) = -N \ln\left(\sigma\sqrt{2\pi}\right) - \frac{\sum_i x_i^2 - N\bar{x}^2}{2\sigma^2}$$

$$\ln \mathcal{L}(\hat{R}) = -N \ln\left(\sigma\sqrt{2\pi}\right) - \frac{\sum_i x_i^2}{2\sigma^2}$$

$$\ln \lambda = \ln \mathcal{L}(\hat{R}) - \ln \mathcal{L}(\hat{S}) = -\frac{1}{2\sigma^2} N\bar{x}^2$$

$$\mathcal{R}_k \equiv \left\{\bar{x}^2 > -\frac{2\sigma^2 \ln k}{N}\right\} \equiv \left\{|\bar{x}| > c\right\}$$

Il criterio di verifica è sostanzialmente un test sul χ^2:

$$\frac{\bar{x}^2}{\frac{\sigma^2}{N}} > \text{cost.}$$

Scuola di Dottorato in Fisica

Quarto esempio

$$\begin{cases} H_0 \equiv \{p_i = \pi_i\} & \left(\forall i \in \{1,2,\ldots,M\}\right) \\ H_a \equiv \{p_i \neq \pi_i\} & \left(\exists i \in \{1,2,\ldots,M\}\right) \end{cases}$$

$$\mathcal{L} \propto \prod_{i=1}^{M} p_i^{n_i}$$

$$\hat{p}_i = \frac{n_i}{N}$$

$$\mathcal{L}(\hat{S}) = \prod_{i=1}^{M} \left(\frac{n_i}{N}\right)^{n_i} = \frac{1}{N^N} \prod_{i=1}^{M} n_i^{n_i}$$

$$\mathcal{L}(\hat{R}) = \prod_{i=1}^{M} \pi_i^{n_i}$$

$$\lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} = N^N \prod_{i=1}^{M} \left(\frac{\pi_i}{n_i}\right)^{n_i}$$

$$\mathcal{R}_k \equiv \{\lambda < k\}$$

Il criterio di verifica dell'ipotesi è differente da quello dato in precedenza (χ^2).

Quinto esempio - I

$$\begin{cases} x \sim N(x;\mu,\sigma) & \text{(con } \sigma > 0 \text{ ignoto)} \\ H_0 \equiv \{\mu = \mu_0\} \\ H_a \equiv \{\mu \neq \mu_0\} \end{cases}$$

$$\ln \mathcal{L} = -N \ln\sigma - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2$$

Sappiamo già che il massimo assoluto si ha per

$$\hat{\mu} = \bar{x} \qquad e \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu})^2$$

e vale quindi

$$\ln \mathcal{L}(\hat{S}) = -\frac{N}{2}\ln\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right] + \\ + \frac{N}{2}\ln N - \frac{N}{2}\ln(2\pi) - \frac{N}{2}$$

Scuola di Dottorato in Fisica

Quinto esempio - II

Ammessa vera H_0

$$\ln \mathcal{L} = \\ = -N \ln\sigma - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu_0)^2$$

che ha un massimo per

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_0)^2$$

che vale

$$\ln \mathcal{L}(\hat{R}) = -\frac{N}{2}\ln\left[\sum_{i=1}^{N}(x_i - \mu_0)^2\right] + \\ + \frac{N}{2}\ln N - \frac{N}{2}\ln(2\pi) - \frac{N}{2}$$

Quinto esempio - III

$$
\begin{aligned}
\ln\lambda &= \ln\mathcal{L}(\hat{R}) - \ln\mathcal{L}(\hat{S}) \\
&= -\frac{N}{2}\left\{\ln\left[\sum_{i=1}^{N}(x_i-\mu_0)^2\right] - \right. \\
&\qquad \left. - \ln\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]\right\} \\
&= -\frac{N}{2}\,\ln\left[\frac{\sum_i (x_i-\mu_0)^2}{\sum_i (x_i-\bar{x})^2}\right] \\
&= -\frac{N}{2}\,\ln\left[1+\frac{N\,(\bar{x}-\mu_0)^2}{\sum_i (x_i-\bar{x})^2}\right] \\
&= -\frac{N}{2}\,\ln\left(1+\frac{t^2}{N-1}\right)
\end{aligned}
$$

definendo una nuova variabile casuale

$$
t = (\bar{x}-\mu_0)\sqrt{\frac{N(N-1)}{\sum_i (x_i-\bar{x})^2}} \;=\; \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{N}}}
$$

Il criterio di verifica dell'ipotesi è un test su una variabile di Student.

Dottorato di Ricerca in Fisica

XIV Ciclo (Anno Accademico 1998–99)

Statistica Applicata alla Fisica

The End

In realtà un lavoro simile non termina mai.
Lo si deve dichiarare concluso quando,
a seconda del tempo e delle circostanze,
si è fatto il possibile.

Johann Wolfgang von Goethe — *Italienische Reise*