Def. di probabilità

Combinatoria (Laplace): $P = \dfrac{n \text{ favorevoli}}{N \text{ tot. eventi}}$

Teoria frequentistica: $P = \lim\limits_{N \to \infty} \dfrac{n \text{ favorevoli}}{N \text{ tot.}}$

Def. assiomatica (teoria matematica)

Dato l'insieme S degli eventi elementari, si definisce "classe degli eventi" o "sigma algebra" una collezione Σ di sottoinsiemi di S tali che:

- $S \in \Sigma$
- $A \in \Sigma \Rightarrow \bar{A} \in \Sigma$
- $\{A_n\} \in \Sigma \Rightarrow \bigcup_n A_n \in \Sigma \wedge \bigcap_n A_n \in \Sigma$

Definiamo probabilità una misura tale che:

- $\forall A \in \Sigma \quad 0 \leq P(A) \leq 1$
- $P(S) = 1$
- $\bigcap_n A_n = \emptyset \Rightarrow P(\bigcup_n A_n) = \sum_n P(A_n)$

È detta variabile casuale una funzione del set di sample S che associa ad ogni evento casuale $X_i \in \Sigma$ un numero reale x_i.

Densità di probabilità $P(X \in (x, x+dx)) = f(x)\,dx$

tale che $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

Valore d'aspettazione di una variabile casuale: $E(x) = \int_{-\infty}^{+\infty} t\, f(t)\,dt$

Varianza $Var(x) = \int_{-\infty}^{+\infty} [t - E(x)]^2 f(t)\,dt$

Per variabili discrete:

$E(x) = \sum_i x_i P(x_i) \qquad Var(x) = \sum_i [x_i - E(x)]^2 P(x_i)$

Probabilità condizionata

$P(A \cap B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$

Se A e B sono eventi indipendenti:

$P(A|B) = P(A) \quad \wedge \quad P(A \cap B) = P(A) \cdot P(B)$

Teorema di Bayes

$\{H_i\} : \ S = \bigcup_i H_i \ \wedge \ \forall i,j \ H_i \cap H_j = \emptyset \quad \Rightarrow \sum_i P(H_i) = 1$

Se $A \in \Sigma$ allora $A = A \cap (\bigcup_i H_i) = \bigcup_i (A \cap H_i)$

$\Rightarrow P(A) = \sum_i P(A \cap H_i) = \sum_i P(A|H_i) P(H_i)$

Inoltre, dalla def. di probabilità condizionata,

$$P(H_i|A) = \frac{P(A|H_i) \cdot P(H_i)}{P(A)} \Rightarrow$$

$$P(H_i|A) = \frac{P(A|H_i) \cdot P(H_i)}{\sum_i P(A|H_i) \cdot P(H_i)}$$

Com: $P(H_i) \equiv$ probabilità a priori

$P(A|H_i) \equiv$ verosimiglianza

$P(H_i|A) \equiv$ probabilità a posteriori

Per applicare il teorema di Bayes dobbiamo conoscere le probabilità a priori.

Postulato di Bayes Se non c'è evidenza del contrario, allora la distribuzione di probabilità $P(H_i)$ si assume costante.

Principio di massima entropia Dati dei vincoli conosciuti, si ricava la distribuzione di probabilità che fa minore utilizzo di informazioni esterne massimizzando l'entropia di Shannon: $S = -\int_\Omega f(x) \log[f(x)]\, dx$

DEFINIZIONI

Il risultato di un esperimento è detto evento. L'insieme di tutti i possibili risultati dell'esperimento è detto popolazione. Una collezione di risultati ottenuta ripetendo l'esperimento è detto sample: un sample si dice fiduciario se a partire da esso possiamo inferire le caratteristiche della popolazione. Se un esperimento è privo di bias, qualsiasi sample ha lo stesso probabilità di essere misurato. Un evento casuale è un membro della popolazione, il suo valore numerico è invece una variabile casuale.

Covarianza $\mathrm{Cov}(x,y) = E[(x-\mu_x)(y-\mu_y)] = E(x\cdot y) - \mu_x\mu_y$

Momento m-esimo $\mu'_m = E(x^m)$ $\mu_m = E[(x-\mu_x)^m]$

Teorema Se due densità di probabilità f e g hanno gli stessi momenti e se $\forall x$ $f(x) - g(x) = \sum_m c_m x^m$ $\Rightarrow f = g$.

Funzione generatrice dei moment: $\Phi_x(t) = E(e^{xt}) = \int_{-\infty}^{\infty} e^{xt} f(x)dx$

tale che $\Phi_x(t) \simeq \sum_m \frac{E(x^m)}{m!} t^m = \sum_m \mu'_m \frac{t^m}{m!}$

E quindi $\mu'_m \equiv \frac{\partial^{(m)}}{\partial t^m} E(e^{xt})\Big|_{t=0}$.

- $\Phi_{ax}(t) = \Phi_x(at)$
- $\Phi_{x+c}(t) = e^{ct}\,\Phi_x(t)$
- $\Phi_{\sum_i x_i}(t) = \prod_i \Phi_{x_i}(t)$ (x_i variabili casuali indipendenti)

Funzione caratteristica $\phi(t) = E(e^{itx}) = \int_{-\infty}^{+\infty} e^{itx} f(x)\,dx$

E' la trasformata di Fourier della densità di probabilità.

- $\phi(0) = 1$
- $\mu'_m \equiv \frac{1}{i^m} \frac{\partial^{(m)}\phi(t)}{\partial t^m}$
- $f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \phi(t)\, e^{-itx}\, dt$

Funzione generatrice delle probabilità $\Pi(s) \equiv E(s^x) = \sum_i s^{x_i} p(x_i)$

Nel caso $\{x_i\} \in \mathbb{N}$: $\Pi(s) = p_0 + p_1 s + p_2 s^2 + \dots$ in cui il coefficiente di s^m è la probabilità del risultato m-esimo.

- $\Pi(1) = 1$
- $\Pi'(1) = E(x)$
- $\Pi''(1) = E(x^2) - E(x) \Rightarrow \mathrm{Var}(x) = \Pi''(1) + \Pi'(1) - [\Pi'(1)]^2$

Se $x_1 \dots x_n$ sono n variabili indipendenti:

- $\Pi_{\sum_i x_i}(s) = \prod_i \left[\Pi_{x_i}(s)\right]$

e se le x_i sono i.i.d.: $\Pi_{\sum x_i}(s) = \left[\Pi_x(s)\right]^n$

Nei processi di branching la Π è una funzione composta. Ad es.: $y \equiv \sum_{i=1}^{n} x_i$; x_i, n variabili casuali

$\Rightarrow \Pi_y(s) = \Pi_n\left[\Pi_x(s)\right]$

Coefficiente binomiale $\binom{a}{b} = \dfrac{a!}{b!\,(a-b)!}$

Distribuzione binomiale probabilità di r successi in N prove di Bernoulli (solo 2 possibili risultati):

$P(r|N) = \binom{N}{r} p^r (1-p)^{N-r}$ con p = probabilità del singolo successo.

- $\mu = pN = E(r)$
- $Var(r) = Np(1-p)$
- $\Phi_r(t) = (pe^t + 1 - p)^N$

Distribuzione geometrica Probabilità di ottenere un successo dopo N prove di Bernoulli sfavorevoli:

$P(N|p) = (1-p)^{N-1}p$

- $\mu = E(N) = \dfrac{1}{p}$
- $Var(N) = \dfrac{1-p}{p^2}$
- $\Phi_N(t) = \dfrac{p}{e^{-t} - (1-p)}$
- $\Pi_{geom}(s) = \dfrac{ps}{1 - s(1-p)}$

Distribuzione di Pascal È una binomiale inversa: rappresenta la probabilità di dover ripetere una prova di Bermoulli N volte (N è la variabile aleatoria) per ottenere r successi.

$P(N|r) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$

- $\Phi_N(t) = \left(\frac{p}{e^{-t}-(1-p)}\right)^r$
- $E(N) = r/p$
- $Var(N) = \frac{r(1-p)}{p^2}$
- $\Pi_{pascal}(s) = \left(\frac{ps}{1-s(1-p)}\right)^r$

Distribuzione di Poisson $P(r|\mu) = \frac{\mu^r}{r!}e^{-\mu}$

- La probabilità di un evento è costante e indipendente dal tempo
- La probabilità che ci sia UN evento in un tempo dt è $\mu\, dt$
- La probabilità che al tempo t ci siano stati ZERO eventi è: $P(0|t) = e^{-\mu t}$
- La probabilità che al tempo t ci sia stato UN evento è: $P(1|t) = \mu t\, e^{-\mu t}$
- La probabilità di aver avuto r eventi al tempo t è: $P(r|t) = \frac{(\mu t)^r}{r!}e^{-\mu t}$ $\mu t = \mu$
- $E(r) = \mu$
- $Var(r) = \mu$
- $\Phi_r(t) = e^{\mu(e^t-1)}$
- $\Pi_{pois}(s) = e^{\mu(s-1)}$
- La distribuzione di Poisson è additiva: se $r \sim Poiss(\mu_1)$, $t \sim Poiss(\mu_2)$, allora $s \equiv r+t \sim Poiss(\mu_1+\mu_2)$.

Distribuzione di Poisson composta

$r \equiv \sum_{i=1}^{m} r_i \qquad r_i \sim Poiss(\mu) \qquad m \sim Poiss(\nu)$

$$P(r|\mu,\nu) = \sum_{m=1}^{\infty}\left[\frac{1}{r!}(m\mu)^r e^{-m\mu}\right]\frac{1}{m!}\nu^m e^{-\nu}$$

- $E(r) = \mu\nu$
- $Var(r) = \nu\mu(\mu+1)$
- $\Phi_r(t) = e^{\nu[e^{\mu(e^t-1)}-1]}$
- $\Pi_{Poiss.\,comp.}(s) = e^{\nu[e^{\mu(s-1)}-1]}$

Distribuzione ipergeometrica

In una scatola ho N palle bianche ed M nere. Ne estraggo k. La probabilità di aver estratto r palle bianche è $P(r|N,M,k) = \frac{\binom{N}{r}\binom{M}{k-r}}{\binom{N+M}{k}}$

- $E(r) = k\,E(1) = \frac{kN}{N+M}$
- $Var(r) = \frac{kNM}{(M+N)^2}\left(1 - \frac{k-1}{N+M-1}\right)$

Le distribuzioni binomiale, geometrica, di Pascal, di Poisson, ed ipergeometrica sono distribuzioni discrete (≡ la variabile aleatoria può assumere solo valori discreti).

Nel caso che la variabile possa assumere valori continui, definiamo la probabilità che x sia contenuto in un piccolo intervallo dx:

$dF = f(x)\,dx$ dove $f(x)$ è la densità di probabilità ed $F(x)$ è detta distribuzione cumulativa.

Distribuzione uniforme $f(x) = \begin{cases} 0 & \text{se } x < a \vee x > b \\ \frac{1}{b-a} & \text{se } a \leq x \leq b \end{cases}$

- $\Phi_x(t) = \frac{1}{b-a}\,\frac{e^{bt} - e^{at}}{t}$
- $E(x) = \frac{b-a}{2}$
- $Var(x) = \frac{(b-a)^2}{12}$

Distribuzione esponenziale $f(x) = \frac{1}{\tau}e^{-\frac{x}{\tau}}$

- $\Phi_x(t) = \frac{1}{1-t\tau}$
- $E(x) = \tau$
- $Var(x) = \tau^2$

La distribuzione esponenziale, come la distribuzione di Poisson, è memoryless:

$P(X > a+b \mid X > a) = P(X > b) \quad \forall a, b \geq 0$

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

La distribuzione cumulativa è, per una normale standard, $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$ ed ha il significato di probabilità $F(x) = P(X < x \mid \mu = 0, \sigma = 1)$.

De Moivre scoprì la distribuzione normale come limite, per $N \to \infty$, della distribuzione della somma di N variabili poissoniane identicamente distribuite con media μ e varianza σ^2.

- $\Phi_x(t) = e^{\mu t + \frac{(\sigma t)^2}{2}}$
- $E(x) = \mu$
- $Var(x) = \sigma^2$
- $FWHM = 2\sqrt{2 \log 2}\, \sigma$.

Distribuzione Gamma $f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$

con: $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy$ (funzione gamma).

La funzione gamma gode delle proprietà:

- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$
- $n \in \mathbb{N} \Rightarrow \Gamma(n) = (n-1)!$

La distribuzione esponenziale è una Gamma con $\alpha \equiv 1$ $(\beta \equiv \tau)$.

- $\Phi_x(t) = \frac{1}{(1-\beta t)^{\alpha}}$
- $E(x) = \alpha\beta$
- $Var(x) = \alpha\beta^2$

Distribuzione di Cauchy $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$

- $Var(x) = \infty$

Densità di probabilità di funzioni di variabili aleatorie

Se $\vec{x} \sim f(\vec{x})$ e $\vec{h} = \vec{h}(\vec{x})$, $f(\vec{x})\,d\vec{x} = g(\vec{h})\,d\vec{h}$

Segue che $g(\vec{h}) = \sum_{\vec{x}_i : \vec{h}(x_i) = \vec{h}} f[\vec{x}_i(\vec{h})] \left| \frac{\partial \vec{x}}{\partial \vec{h}} \right|_{\vec{x}_i}$

(dove l'ultimo fattore è il valore assoluto dello Jacobiano del cambio di variabile $\vec{x} = \vec{x}(\vec{h})$).

Approssimazione lineare

A volte è conveniente utilizzare quest'approssimazione per calcolare la varianza di funzioni di variabili aleatorie. Se $\vec{y} = \vec{y}(\vec{x})$, tronchiamo al primo ordine uno sviluppo di Taylor attorno alla media:

$\vec{y}(\vec{x}) \simeq \vec{y}(\vec{\mu}) + J \cdot (\vec{x} - \vec{\mu}) + \dots$ dove $\vec{\mu} = E(\vec{x})$ e

J è la matrice Jacobiana $\frac{\partial y_i}{\partial x_j}$. In questa approssimazione, $V(\vec{y}) \simeq J\,V(\vec{x})\,J^T$, dove $V(\vec{y})$ e $V(\vec{x})$ sono le matrici di covarianza di $\vec{y}$ ed $\vec{x}$.

Questa approssimazione è valida se, nel volume in cui varia $\vec{x}$, $\vec{y}$ è ben approssimata dalla sua derivata prima.

Distribuzione del χ^2

$$f(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right) 2^{k/2}} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \equiv \chi^2_k$$

Rappresenta la densità di probabilità della somma di k variabili normali standard elevate al quadrato.

- $\Phi_x(t) = (1-2t)^{-\frac{k}{2}}$
- $E(x) = k$
- $Var(x) = 2k$

Se le k variabili normali sono invece distribuite come $N(0,\sigma^2)$, $\chi^2_k \rightarrow \sigma^2 \chi^2_k$.

Distribuzione di Student $f_{n-1}(x) = \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\left(1+\frac{x^2}{n-1}\right)^{n/2}}$

Per $n = 2$ la Student si riduce alla distribuzione di Cauchy. È usata nei test statistici per studiare la significanza della differenza di medie campionarie estratte da una popolazione normale, quando la deviazione standard della popolazione è sconosciuta. La variabile x rappresenta

$$x \equiv t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \cdot \left(\frac{s}{\sqrt{\sigma^2}}\right)^{-1} = \frac{\bar{x}-\mu}{s/\sqrt{n}} \quad \text{dove} \quad s \equiv \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1},$$

ed è quindi il rapporto tra una $N(0,1)$ ed un χ^2_{n-1}.

- $E(x) = 0$
- $n > 3 \Rightarrow Var(x) = \frac{n-1}{n-3}$
- $n \to \infty \Rightarrow x \sim N(0,1)$

Distribuzione di Fisher-Snedecor

$$f_{n,m}(x) = \frac{\Gamma\left(\frac{n}{2}+\frac{m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \cdot \frac{x^{\frac{n}{2}-1}}{(nx+m)^{\frac{n+m}{2}}}$$

È usata per stabilire la significanza della differenza delle varianze di due campioni indipendenti estratti da una popolazione normale. La variabile x rappresenta il rapporto tra gli estimatori s_n, s_m delle varianze dei due campioni di dimensioni, rispettivamente, $n+1$ ed $m+1$. I test sono condotti sull'ipotesi $\sigma_n^2 = \sigma_m^2 \equiv \sigma^2$

$$\Rightarrow x = \frac{s_n^2}{s_m^2} \sim \frac{\sigma_n^2 \chi_n^2}{n}\left(\frac{m}{\sigma_m^2 \chi_m^2}\right) \sim \frac{\chi_n^2}{n} \frac{m}{\chi_m^2}.$$

- $m \geq 3 \Rightarrow E(x) = \frac{m}{m-2}$
- $m \geq 5 \Rightarrow Var(x) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$
- $f_{n,m}(x) = f_{m,n}\left(\frac{1}{x}\right)$
- $G_{n,m}(x) \equiv \int_{-\infty}^{x} f_{n,m}(t)\,dt \Rightarrow G_{m,n}(x) = 1 - G_{n,m}\left(\frac{1}{x}\right)$

Media campionaria

Se $\{X_i\}_{i=1..n}$ è una collezione di variabili aleatorie i.i.d, si definisce media campionaria $\bar{x} = \frac{\sum_i x_i}{n}$.

- E' anch'essa una variabile aleatoria
- E' un operatore lineare
- $Var(\bar{x}) = \frac{1}{n} Var(x_i)$
- legge dei grandi numeri: $\lim_{n \to \infty} P\left[|\bar{x} - E(x_i)| > \varepsilon \right] = 0$
- teorema del limite centrale: $n \to \infty \Rightarrow \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Per n finito:

- $X_i \sim Cauchy \Rightarrow \bar{x} \sim Cauchy$
- $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $x_i \sim Poiss(\mu_i) \Rightarrow \bar{x} \sim Poiss\left(\frac{\sum_i^n \mu_i}{n}\right)$
- $x_i \sim \binom{N_i}{k} p^k (1-p)^{N_i - k} \Rightarrow \bar{x} \sim \binom{\sum_i^n N_i}{k} p^k (1-p)^{\sum_i^n N_i - k}$
- $x_i \sim exp(\tau) \Rightarrow f(\bar{x} | \tau, n) = \frac{n^n}{\Gamma(n) \tau^n} \bar{x}^{n-1} e^{-\bar{x}/\tau}$

Somma dei quadrati rispetto alla media

Per un campione di dimensione n:

$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$ Ma se μ non è conosciuta:

$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 \sim \chi^2_{n-1}$

Il problema si è ridotto di un grado di liberta' a causa del vincolo $\sum_{i=1}^n (x_i - \bar{x}) = 0$.

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

- $E(s^2) = \sigma^2$
- $x_i \sim N(\mu, \sigma^2) \Rightarrow Var(s^2) = \frac{2\sigma^4}{n-1} \wedge s^2 \sim \frac{\sigma^2}{n-1} \chi^2_{n-1}$

DARIA
Ipotesi test
Z test
Student
ML
parametri stimatori + Dim
(suff + eff)

MAURO + TOMMASO

Prob. secondo kolmogorov, stimatori, decision making, ML, test

MANZI: lim. centr. (dim), grandi numeri (dim), stime: suff, eff, cramer-rao (+ dim), test d'ipotesi decision making, c.l.

Disuguaglianza di Bienaymé - Chebyshev

$\forall$ densità di probabilità $f(x)$ t.c. $E(x)=\mu$ ed

$E(x-\mu)^2 = \sigma^2$ vale:

$\forall k \in \mathbb{N} \quad P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

- Infatti: $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$

$$\geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{+\infty} (x-\mu)^2 f(x)\,dx$$

$$\geq (k\sigma)^2 \cdot \left[\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{+\infty} f(x)\,dx\right]$$

$$= (k\sigma)^2 P(|x-\mu| \geq k\sigma)$$

- e' la disuguaglianza più "forte" che è possibile ottenere senza sapere nulla su $f(x)$.

Legge dei grandi numeri

Dato un insieme aleatorio di variabili aleatorie iid, di dimensione n, tale che $\bar{x} \equiv \frac{\sum_{i=1}^n x_i}{n} \wedge \forall i \; E(x_i)=\mu$:

- forma debole: $\lim_{n\to\infty} P(|\bar{x}-\mu| > \varepsilon) = 0$
- forma forte: $P(\lim_{n\to\infty} |\bar{x}-\mu| > \varepsilon) = 0$
- dimostrazione per la forma debole per distribuzioni a varianza finita:

$$Var(\bar{x}) = \frac{\sigma^2}{n} \Rightarrow P\left(|\bar{x}-\mu| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq k^{-2}$$

$$\varepsilon \equiv k\frac{\sigma}{\sqrt{n}} \Rightarrow k^2 \equiv n\varepsilon^2/\sigma^2$$

$$P(|\bar{x}-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \Rightarrow P(|\bar{x}-\mu| < \varepsilon) > 1 - \frac{\sigma^2}{n\varepsilon^2}$$

$$\Rightarrow \lim_{n\to\infty} P(|\bar{x}-\mu| < \varepsilon) = 1$$

- Si dice che $\bar{x}$ tende in probabilità a μ.

Teorema del limite centrale

$\{x_i\}$ variabili aleatorie indipendenti t.c. $E(x_i) = \mu_i$ e $Var(x_i) = \sigma_i^2 < +\infty$, con almeno uno $\sigma_m^2 > 0$.

Definendo $S_m \equiv \sum_{i=1}^{m} x_i$ e $s_m \equiv \sqrt{Var(S_n)} = \sqrt{\sum_{i=1}^{m} \sigma_i^2}$, allora $Z_m \equiv \frac{S_m - E(S_n)}{s_m}$ converge normalmente (tende in distribuzione ad $N(0,1)$) per $n \to +\infty$.

- $\{x_i\}$ iid $\Rightarrow \frac{S_m/m - \mu}{\sigma/\sqrt{m}} \xrightarrow[m \to +\infty]{Dist} N(0,1)$

  Infatti: $\forall$ distribuzione con momenti $\neq 0$ si ha $\Phi_{x-\mu}(t) = E\left(e^{(x-\mu)t}\right) \simeq 1 + 0 + \sigma^2 \frac{t^2}{2!} + \dots$

  Se definiamo: $w \equiv \frac{x-\mu}{\sigma\sqrt{n}} \Rightarrow \Phi_w(t) = \Phi_{x-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right) = \simeq 1 + \frac{t^2}{2m}$

  Definiamo: $z \equiv \sum_{i=1}^{m} w_i = \sum_{i=1}^{m} \frac{x_i - \mu}{\sigma\sqrt{m}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

  Allora $\Phi_z(t) = \left[\Phi_w(t)\right]^m = \left(1 + \frac{t^2}{2m}\right)^m \xrightarrow[n \to \infty]{} e^{t^2/2}$

  che è la $\Phi_x(t)$ per $x \sim N(0,1)$

- $\bar{x} = \frac{\sigma}{\sqrt{m}} z + \mu \Rightarrow \bar{x} \xrightarrow[m \to \infty]{Dist} N\left(\mu, \frac{\sigma^2}{m}\right)$

- Il teorema non vale per distribuzioni che non hanno momenti finiti (ad es. Cauchy).

P-value

Data una ipotesi nulla H_0, esatta e che specifichi completamente la distribuzione di probabilità di una variabile aleatoria X, e dato un valore x_0 misurato, si definisce p-value la quantità

$P(X > x_0 | H_0)$

- H_0 vera $\Rightarrow P(X > x_0 | H_0) \sim U(0,1)$

Test d'ipotesi (test di significato)

H_0 definisce una distribuzione di probabilità per X.

Viene misurato $X = x_0$. Lo sperimentatore deve definire un Livello di significanza α tale che

$P(X \geq x_0 | H_0) \leq \alpha \Rightarrow$ lo sperimentatore decide di rigettare H_0.

Il p-value è una misura di quanto i dati si accordano con H_0. Se $P(x \geq x_0 | H_0) \leq \alpha$ si dice che il risultato è significativo.

Z - test

Si vuole verificare la media di una distribuzione Gaussiana.

$\forall i \quad z_i \sim N(\mu, \sigma^2) \Rightarrow$ si prende la media campionaria

$\bar{z} = \frac{\sum_i^n z_i}{n}$. L'ipotesi H_0 è:

- $H_0 : \mu = \mu_0$

Se H_0 è vera, $\bar{z} \sim N(\mu_0, \frac{\sigma^2}{n}) \Rightarrow z \equiv \sqrt{n} \, \frac{\bar{z} - \mu_0}{\sigma} \sim N(0,1)$

- $\text{P-value} = 2 \int_{|z_{mis}|}^{+\infty} N(0,1) \, dz$

T- test (test di Student)

Si vuole verificare la media di una distribuzione Gaussiana a varianza sconosciuta. Si stima la varianza con la varianza campionaria $s^2 \equiv \frac{\sum_i^n (z_i - \bar{z})^2}{n-1}$, $z_i \sim N(\mu, \sigma^2) \forall i$.

- $H_0 : \mu = \mu_0$

Se H_0 è vera, $t \equiv \sqrt{n} \, \frac{\bar{z} - \mu_0}{s} \sim S_{n-1}$ (distribuzione di Student con $n-1$ gradi di libertà).

- $\text{P-value} = 2 \int_{|t_{mis}|}^{+\infty} S_{n-1}(t) \, dt$

Distribuzione empirica

Se abbiamo un campione di eventi $\{x_i\}_{i=1...n}$, si definisce distribuzione empirica:

$$F_n(y) = \begin{cases} 0 & \text{se } y < x_1 \\ i/n & \text{se } x_i < y < x_{i+1} \\ 1 & \text{se } y > x_n \end{cases}$$

(si assume il campione ordinato in ordine di x_i crescente)

Se la distribuzione cumulativa di probabilità della popolazione da cui è estratto il campione è $F(x)$, allora $n \cdot F_n(y) \sim Bin\left(F(y), n\right)$.

- $E\left(F_n(x_i)\right) = F(x_i)$
- $Var\left(F_n(x)\right) = \frac{1}{n} F(x)\left[1 - F(x)\right]$

Test di Pearson

- H_0: il campione $\sim f(x)$

Dividiamo gli n elementi del campione in k classi (costruiamo un istogramma), di occupazione $r_1 \dots r_k$. Chiamiamo $m_i = p_i \cdot n$ la predizione di occupazione della classe i-esima se H_0 è vera.

Ciascun r_i ha probabilità binomiale, ma per il T.L.C., per $n \cdot p_i \geq 5$, $r_i \sim N(np_i, np_i)$.

Se H_0 è vera, $u = \sum_{i=1}^{k} \frac{(r_i - m_i)^2}{m_i} \sim \chi^2_{k-1}$.

- $\text{P-value} = \int_{u_M}^{+\infty} \chi^2_{k-1}(u)\, du$

Test ω^2 di Smirnov, Cramer, Von Mieses

Dato un campione $\{x_i\}_{i=1\dots n}$ ordinato in ordine crescente, ne costruiamo la distribuzione empirica $F_n(x)$.

- H_0: la cumulativa della popolazione è $F(x)$.

$\omega^2 \equiv \int_0^1 [F_n(x) - F(x)]^2 dF = \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^{n}\left[F(x_i) - \frac{2i-1}{2n}\right]^2$ rappresenta la varianza di $F_n(x)$.

Se H_0 è vera, la densità di ω^2 non dipende dalla distribuzione di x.

- $E(\omega^2) = \frac{1}{6n}$
- $Var(\omega^2) = \frac{4n-3}{180\, n^3}$
- $P(n \cdot \omega^2 > x \mid H_0) = \frac{1}{\pi\sqrt{x}} \sum_{k=0}^{\infty}\left[\frac{\Gamma(k+\frac{1}{2})}{k!\,\sqrt{\pi}}\sqrt{4k+1}\, e^{-\frac{(4k+1)^2}{16x}} \cdot K_{1/4}\left(\frac{(4k+1)^2}{16x}\right)\right]$

Test di Kolmogorov - Smirnov

- H_0: la cumulativa del campione è $F(x)$.

Usiamo come statistica la distanza tra F e la distribuzione empirica. Se H_0 è vera, la distribuzione di $d_n \equiv \sup_{-\infty<x<+\infty} |F_n(x) - F(x)|$ non dipende dalla forma di H_0. Si ricava: $(t \equiv \sqrt{n} \cdot d_n)$

- P-value $= P(t \geq t_{mis} \mid H_0) = 2\sum_{k=1}^{\infty}(-1)^{k+1} e^{-2k^2 t_{mis}^2}$

Run - test (test di Wald - Wolfowitz)

- H_0: una sequenza di eventi è casuale (non presenta trend).

Mescoliamo due serie di dati mettendoli in ordine crescente. Chiamiamo r il numero di run (passaggi da una serie all'altra). Se n, m sono le dimensioni delle serie:

- $E(r) = \frac{2nm}{n+m} + 1$
- $Var(r) = \frac{2nm(2nm - n - m)}{(n+m)^2(n+m-1)}$

Per $n m > 10$, $r \sim$ gaussiana e si effettua uno z-test su $z \equiv \frac{r - E(r)}{\sqrt{Var(r)}} \sim N(0,1)$.

Normal probability plot

Prendiamo un campione ordinato in modo crescente. Se $\forall i \quad x_i \sim N(\mu, \sigma)$, ovvero $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$, costruiamo una nuova variabile y tale che:

$$F_n(x_i) = \frac{i}{n} = \int_{-\infty}^{y(F_n)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \Rightarrow y_i(F_n) \equiv F^{-1}(F_n(x_i))$$

Si chiama "normal probability plot" il plot dei punti $(x_i, y_i(F_n))$. Se $x_i \sim N(\mu, \sigma)$, i punti si allineano lungo la retta $y = \frac{x-\mu}{\sigma}$.

Combinazione di k test indipendenti sullo stesso H_0

Supponiamo di avere ottenuto k p-value $\{p_i\}_{i=1\dots k}$.
Definiamo $y_i \equiv -\ln p_i$. Ogni $p_i \sim U(0,1)$ se H_0 è vero, quindi:

$$g(y_i)\,dy_i = f(p_i)\left|\frac{dp_i}{dy_i}\right| dy_i = e^{-y_i} dy_i$$

$$\Phi_{y_i}(t) = \frac{1}{1-t}$$

Definiamo: $u \equiv -2\sum_{i=1}^{k} y_i \Rightarrow \Phi_u(t) = \left(\frac{1}{1-2t}\right)^k$

$\Rightarrow u \sim \chi^2_{2k}$. Il p-value globale sarà allora:

$$\text{P-value} = P\left(\chi^2_{2k} \geq -\sum_{i=1}^{k} \ln p_i\right) = \int_{u_{MIS}}^{+\infty} \chi^2_{2k}(t)\,dt$$

Test decisionale

Vogliamo scegliere tra due ipotesi H_0 ed H_1. Definiamo una regione critica w_α per H_0 (w_α = regione di accettanza per H_1). Si dice:

- probabilità di falso rigetto $\alpha = \int_{w_\alpha} f(x|H_0)\,dx$ = "dimensione" del test
- probabilità di falsa accettanza $\beta = \int_{\overline{w_\alpha}} f(x|H_1)\,dx$
- potenza del test $P = 1-\beta = \int_{w_\alpha} f(x|H_1)\,dx$

Costruiamo il test fissando α e massimizzando P (cioè minimizzando β).

Lemma di Neyman - Pearson

Supponiamo: $H_0 : x \sim f_0(x)$; $H_1 : x \sim f_1(x)$ con f_0 ed f_1 completamente specificate. Date le verosimiglianze $L(x|H_0)$ ed $L(x|H_1)$, fissati α e $t_\alpha : P(t \geq t_\alpha | H_0) = \alpha$, usiamo come statistico di test $t = \frac{L(x|H_1)}{L(x|H_0)}$.

Il test più potente si costruisce scegliendo w_α, con il vincolo $P(x \in w_\alpha | H_0) = \alpha$, in modo da massimizzare $E_{w_\alpha}\left(\frac{f_1}{f_0} | H_0\right)$, cioè massimizzando $P(t(x \in w_\alpha) | H_0)$.

Allora rigettiamo H_0 se $t \geq t_\alpha$.

Test decisionale, nel caso di ipotesi complesse, di verosimiglianza

Si usano per testare teorie i cui parametri devono essere stimati a partire dai dati. Si definisce, per una distribuzione $f(x, \vec{\theta})$, la FUNZIONE DI VEROSIMIGLIANZA $L = \prod_{i=1}^{n} f(x_i, \vec{\theta})$.

Supponiamo che l'ipotesi H_0 ponga restrizioni su alcuni parametri: detto Ω lo spazio dei parametri,

- $H_0 : \vec{\theta} \in \omega \subset \Omega$
- $H_1 : \vec{\theta} \in \Omega$

Definiamo il RAPPORTO DI MASSIMA VEROSIMIGLIANZA

$$\lambda = \frac{\sup_{\theta \in \omega} L(\vec{x}|\theta)}{\sup_{\theta \in \Omega} L(\vec{x}|\theta)}$$

La densità $\lambda(H_0)$ non è sempre calcolabile, ma Wilks dimostrò che:

- $n \to \infty \Rightarrow -2\ln(\lambda) \sim \chi^2_r$

con r = numero di parametri fissati da H_0. In questo limite il test diventa:

- $k_\alpha : P(\chi^2_r > k_\alpha) = \alpha$
- $-2\ln\lambda \geq k_\alpha \Rightarrow$ rigettare H_0

Estimatori

Un estimatore è una statistica, funzione delle osservazioni, usata per stimare un parametro sconosciuto.

Una statistica non dipende in alcun modo dai parametri della distribuzione.

CONSISTENZA: la stima tende al valore vero del parametro all'aumentare del numero di osservazioni.

NON DISTORSIONE: per campioni finiti, $E(\hat{\theta}) = \theta$, avendo chiamato $\hat{\theta}$ l'estimatore del parametro θ.

SUFFICIENZA: $\hat{\theta}$ è sufficiente quando nessuna altra funzione dei dati aggiunge informazioni su θ.

EFFICIENZA: si dice efficiente l'estimatore a varianza minima.

Disuguaglianza di Cramer - Rao

Dato un estimatore distorto t di una funzione τ del parametro θ: $E(t) = \tau(\theta) + b(\theta)$

- $$Var(t) \geq \frac{\left(\frac{\partial \tau}{\partial \theta} + \frac{\partial b}{\partial \theta}\right)^2}{E\left(\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right)}$$
- Se $\tau(\theta) = \theta$: $\quad Var(t) \geq \dfrac{1}{E\left(\left(\frac{\partial}{\partial \theta} \ln L\right)^2\right)} \quad$ (se $b = 0$)

Se vale l'uguaglianza, l'estimatore t è efficiente.

Teorema di fattorizzazione di Neyman - Fisher

Se la verosimiglianza $L = \prod_{i=1}^{n} f(x_i, \theta)$ è fattorizzabile nella forma

$$L = g(\hat{\theta}, \theta)\, h(x)$$

allora $\hat{\theta}$ è un estimatore sufficiente per θ.

Rapporto tra sufficienza ed efficienza di un estimatore

Sufficienza $\Leftarrow$ efficienza.

Teorema di Darmois

Se esiste una statistica sufficiente, allora la densità di probabilità appartiene alla famiglia degli esponenziali.

Informazione di Fisher

Si definisce informazione di Fisher: $I_x(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$.

Si dimostra che: $I_x(\theta) = E\left(\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right)$.

- $I_x(\theta)$ è additiva per misure indipendenti.
- In generale, se $\hat{\theta}$ è un estimatore di θ: $I_x(\theta) \geq I_{\hat{\theta}}(\theta)$

  L'informazione si conserva sse $\hat{\theta}$ è sufficiente. Infatti:

  $\hat{\theta}$ suff. $\Leftrightarrow L(x,\theta) = g(\hat{\theta},\theta)\, h(x) \quad \Leftrightarrow$

  $I_x(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = -E\left(\frac{\partial^2 \ln g}{\partial \theta^2}\right) = I_{\hat{\theta}}(\theta)$
- In caso si abbiano più parametri, I ha una forma matriciale: $(I_x(\theta))_{ij} = -E\left(\frac{\partial \ln L}{\partial \theta_i} \cdot \frac{\partial \ln L}{\partial \theta_j}\right)$
- In alcuni casi "regolari": $(I_x(\theta)_{ij}) = -E\left(\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right)$
- Per variabili normali $x \sim N(\mu, \sigma^2)$:

$$I_x(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

- Il metodo della massima verosimiglianza fornisce estimatori asintoticamente consistenti, efficienti e distribuiti normalmente. In questo caso la matrice di covarianza $V(\theta)$ è approssimata da $I_x(\theta)$.

  $$J_n(\theta) \equiv \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \ln[f(x_i, \theta)]$$

  Dalla legge dei grandi numeri e dalla consistenza di $\hat{\theta}$:

  $$\hat{\theta} \xrightarrow{P} \theta \Rightarrow \frac{J_n(\hat{\theta})}{n} \xrightarrow{P} I_x(\theta) \sim n \cdot V^{-1}(\theta)$$

  Per la normalità asintotica di $\hat{\theta}$: $(\hat{\theta} - \theta)\sqrt{J_n(\hat{\theta})} \xrightarrow{\mathcal{L}} N(0,1)$.

Errore quadratico minimo

Per uno stimatore $\hat{\theta}$ di θ tale che $E(\hat{\theta}) = \theta + b(\theta)$, si definisce:

$$\mathrm{mse}(\hat{\theta}) = E[(T-\theta)^2] = E[(T - E(T) + E(T) - \theta)^2] =$$
$$= [E(T) - \theta]^2 + E(T^2) - E^2(T) = b^2(\theta) + Var(\theta)$$

Si dice efficiente lo stimatore a mse minimo.

- Se S è un generico stimatore di θ, e $\hat{\theta}$ uno stimatore efficiente di θ, allora, definendo $\hat{S} = E(S|\hat{\theta})$:

$$\mathrm{mse}(\hat{S}) \leq \mathrm{mse}(S) \quad \text{(teorema di Rao - Blackwell)}$$

e l'uguaglianza vale sse $S = S(\hat{\theta})$.

TEST DI SIGNIFICATO

1) z - test verificare la media di una distrib. Gaussiana:

$x_i \sim N(\mu, \sigma^2)$ $\forall i$ σ^2 nota

$H_0: \mu = \mu_0$

$z \equiv \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu_0) \sim N(0,1)$ se H_0 è vera.

. test a due code.

2) t - test (test di Student)

$x_i \sim N(\mu, \sigma^2)$ $\forall i$, σ^2 sconosciuta

$H_0: \mu = \mu_0$

Si stima la varianza: $\hat{\sigma}^2 = s^2 \equiv \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1} \sim \chi^2_{n-1}$

$t \equiv \frac{\sqrt{n}}{s}(\bar{x} - \mu_0) \sim S_{n-1}$ se H_0 è vera.

. test a due code.

3) Test di Pearson, verificare la densità di probabilità (o la distribuzione) della popolazione da cui è estratto il campione.

- Dividiamo il campione in K classi. Chiamiamo $r_1 \dots r_k$ il numero di elementi in ogni classe.

$H_0: x \sim f(x)$

- Chiamiamo: $n_i \equiv \int_{\substack{\text{classe}\\ i\text{-esima}}} f(x)dx \cdot m$ (m = dimensione campione)

Per $r_i > 5$, $r_i \sim N(n_i, n_i)$.

$u \equiv \sum_{i=1}^{k} \frac{(r_i - n_i)^2}{n_i} \sim \chi^2_{k-1}$ se H_0 è vera.

4) ω^2- test (test di Smirnov, Cramer, Von Mieses) " " " "

$H_0: x \sim f(x)$

- Costruiamo la cumulativa $F(x|H_0) = \int_{-\infty}^{x} f(t)\,dt$ e l'empirica $F_n(x)$.

$\omega^2 \equiv \int_0^1 [F_n(x) - F(x)]^2 dF = \frac{1}{12m^2} + \frac{1}{m}\sum_{i=1}^{m}\left[F(x_i) - \frac{2i-1}{n}\right]^2 = Var(F_n(x))$.

Se H_0 è vera, la densità di ω^2 è indipendente dalla densità di x:

$$\left.\begin{aligned} E(\omega^2) &= 1/6m \\ Var(\omega^2) &= \frac{4m-3}{180\,m^3} \end{aligned}\right\} \text{ se } H_0 \text{ è vera.}$$

5) Test di Kolmogorov-Smirnov " " " "

$H_0 : x \sim f(x)$

. Costruiamo l'empirica $F_n(x)$ e la cumulativa $F(x)$.

$t \equiv \sqrt{n} \cdot \sup_{-\infty < x < +\infty} |F_n(x) - F(x)|$ è indipendente dalla densità di x, se Ho è vera.

$P(t \geq t_M | H_0) = 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 t_M^2}$ se Ho è vera.

6) Run test (test di Wald, Wolfowitz) verifica l'omogeneità di due campioni di dimensioni n, m (provenienza dalla medesima popolazione).

Ho : i due campioni sono estratti dalla stessa popolazione.

. Uniamo i due campioni e li rimescoliamo ordinando i dati in ordine crescente.

r ≡ numero di volte che, nella sequenza, si passa da un elemento di un campione ad un elemento dell'altro campione (numero di run).

$$\left.\begin{aligned} E(r) &= \frac{2nm}{n+m} + 1 \\ Var(r) &= \frac{2nm(2nm-n-m)}{(n+m)^2(n+m-1)} \end{aligned}\right\} \text{ se Ho è vera.}$$

Per $n \cdot m \geq 10$ r ha una distribuzione normale:

$z \equiv \frac{r - E(r)}{\sqrt{Var(r)}} \sim N(0,1)$ (z-test) se Ho è vera.

7) TEST ANOVA (analysis of variance) verifica l'omogeneità di più campioni di dati.

- limite centrale
- legge grandi numeri
- estimatori

} DIMOSTR.

- Dimostrazione disuguaglianza di Cramer - Rao

  - n osservaz. di $x \sim f(x,\theta)$
  - L derivabile 2 volte in θ
  - t estimatore per $\tau(\theta)$

$E(t) = \int t(x_i)\, L(x,\theta)\, dx_1 \cdots dx_n \equiv \tau(\theta) + b(\theta)$

Derivando una volta:

$\int t \frac{\partial}{\partial\theta} L(x,\theta)\, d\vec{x} = \partial\tau/\partial\theta + \partial b/\partial\theta$

Usando la derivata logaritmica:

$\frac{\partial}{\partial\theta} L = L \frac{\partial}{\partial\theta} \ln L \Rightarrow \int t \frac{\partial}{\partial\theta} \ln L\, L\, d\vec{x} = \partial_\theta \tau + \partial_\theta b$ (1)

Usando la condiz. di normalizza. della p.d.f.:

$\int L(x,\theta)\, d\vec{x} = 1$ Derivando a dx e sx:

$\int \frac{\partial}{\partial\theta} \ln L\, L\, d\vec{x} = 0 \Rightarrow 0 = (\tau(\theta) + b(\theta)) \cdot \int \frac{\partial}{\partial\theta} \ln L\, L\, dx$ (2)

Sottraggo la (2) dalla (1) (sto sottraendo 0):

$\int (t - \tau - b)\, \partial_\theta \ln L\, L\, d\vec{x} = \partial_\theta \tau + \partial_\theta b$

Fattorizzando: $\left[\int (t - \tau - b)\, L\, d\vec{x}\right] \cdot \left[\int \partial_\theta \ln L\, L\, d\vec{x}\right] = \partial_\theta \tau + \partial_\theta b$

Posso applicare la disug. di Schwartz: $\|a\| \cdot \|b\| \geq \|a b\|$

la applico al quadrato: $\|a^2\|\, \|b^2\| \geq \|(ab)^2\|$ (considerando come misura $L \cdot d\vec{x}$):

$\left[\int (t - \tau - b)^2 L\, d\vec{x}\right] \cdot \left[\int (\partial_\theta \ln L)^2 L\, dx\right] \geq (\partial_\theta \tau + \partial_\theta b)^2$

Dato che $E(t) = \tau + b$, $\int (t - \tau - b)^2 f(x,\theta)\, d\vec{x} = Var(t)$.

Quindi:

$$Var(t) \geq \frac{(\partial_\theta \tau + \partial_\theta b)^2}{E\left((\partial_\theta \ln L)^2\right)} \qquad \text{cvd.}$$

- Dimostrazione equivalenza definizioni di sufficienza

(A) $P(\vec{X} = \vec{x} \mid \hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = h(\vec{x})$ (h non dip. da θ)

(B) $P(\vec{X} = \vec{x} \mid \theta) = L(\vec{x},\theta) = h(\vec{x}) \cdot g(\hat{\theta}(\vec{x}), \theta)$

- (B) $\Rightarrow$ (A)

$A \equiv \{\vec{x}^* \in S : \hat{\theta}(\vec{x}^*) = \hat{\theta}(\vec{x})\}$ (è un evento!)

Allora $P(\hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = \sum_{\vec{x}^* \in A} P(\vec{X} = \vec{x}^*)$

Avevamo supposto:

$P(\vec{X} = \vec{x} \mid \theta) = h(\vec{x})\, g(\hat{\theta}(\vec{x}), \theta) \Rightarrow P(\vec{X} = \vec{x}) = h(\vec{x})\, g(\hat{\theta}(\vec{x}))$

Per cui:

$P(\hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = \sum_{\vec{x}^* \in A} h(\vec{x}^*)\, g(\hat{\theta}(\vec{x}^*))$ g non dip. espl. da $\vec{x}$, quindi:

$P(\hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = g(\hat{\theta}(\vec{x})) \cdot \sum h(\vec{x}^*)$

E quindi:

$P(\vec{X} = \vec{x} \mid \hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = \frac{P(\vec{X} = \vec{x})}{P(\hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x}))} = \frac{\cancel{g(\hat{\theta})}}{\cancel{g(\hat{\theta})}} \frac{h(\vec{x})}{\sum_{\vec{x}^* \in A} h(\vec{x}^*)}$ che non dip. da θ cvd.

- (A) $\Rightarrow$ (B) $h(\vec{x}) = P(\vec{X} = \vec{x} \mid \hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = \frac{P(\vec{X} = \vec{x})}{P(\hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x}))} = P(\vec{X} = \vec{x}) / g(\hat{\theta}(\vec{x}), \theta)$

Dato che $P(\hat{\theta}(\vec{X}) = \hat{\theta}(\vec{x})) = \sum_{\vec{x}^* \in A} P(\vec{X} = \vec{x}^*) = g(\hat{\theta}(\vec{x}), \theta)$ cvd.

Cercare dim. teor. di Darmois

• Efficienza $\Rightarrow$ Sufficienza

Nella disug. di Schwartz vale l' "=" quando i due vettori (nel nostro caso: $\hat{\theta} - \tau(\theta) - b(\theta)$ e $\partial_\theta \ln L$) sono proporzionali.
Quindi:
condizione di efficienza (minima varianza): $\partial_\theta \ln L = A(\theta) \cdot (\hat{\theta} - \tau(\theta) - b(\theta))$
La condizione di sufficienza è:
$L(x,\theta) = g(\hat{\theta},\theta)\, h(x) \Rightarrow \partial_\theta \ln L = \partial_\theta \ln g(\hat{\theta},\theta)$ (generica funzione di θ e dello stimatore).
$A(\theta) \cdot (\hat{\theta} - \tau(\theta) - b(\theta))$ È una generica funzione di $\hat{\theta}, \theta$
$\Rightarrow$ la cond. di efficienza implica quella di sufficienza.

• Dimostrazione TLC nel caso x_i iid

Ipotesi: x_i iid, $\sigma^2 \neq 0$, $\sigma^2 < +\infty$
$S_n \equiv \sum_{i=1}^n x_i \qquad s_n^2 \equiv \sum_{i=1}^n \sigma_i^2 = n\sigma^2 \qquad \Rightarrow \quad z_n \equiv \dfrac{S_n - E(S_n)}{s_n} \rightsquigarrow N(0,1)$

Dim.: $W_i \equiv \dfrac{x-\mu}{\sqrt{n}\sigma}$ detti μ e σ media e varianza di x.
La f. gen. dei momenti:
$\Phi_x(t) = E(e^{tx}) \Rightarrow \Phi_{x-\mu}(t) = E(e^{t(x-\mu)}) = 1 + E(x-\mu)t + E(x-\mu)^2\frac{t^2}{2} + \dots$
$= 1 + 0 + \sigma^2 t^2/2 + \dots$
Quindi: $\Phi_W(t) = \Phi_{x-\mu}(t/\sigma\sqrt{n}) = 1 + 0 + \dfrac{\sigma^2 t^2}{2\sigma^2 n} + O(n^{-3/2})$
Per $n \to \infty$ il termine $O(n^{-3/2})$ è trascurabile rispetto al resto.
$z_n \equiv \sum_i^n W_i = \sum_i \dfrac{x_i - \mu}{\sqrt{n}\sigma} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
$\Phi_{z_n}(t) = [\Phi_{W_i}(t)]^n = (1 + t^2/2n)^n \xrightarrow[n\to\infty]{} e^{t^2/2} = \Phi_{N(0,1)}(t)$

• Teorema di Darmois

Se $\exists$ una statistica sufficiente, la densità di probabilità è della famiglia degli esponenziali.

Dim.: $\exists$ statistica suff. $\hat{\theta} \Rightarrow L(x,\theta) \equiv h(x)\, g(\hat{\theta},\theta)$
$\Rightarrow \log L = \log h(x) + \log g(\hat{\theta},\theta) \Rightarrow \partial_{x_i}\partial_\theta \ln L = \partial_{x_i}\partial_\theta \ln g(\hat{\theta}(x),\theta)$
$= (\partial_{\hat{\theta}}\partial_\theta g)(\partial_{x_i}\hat{\theta})$
Se $n > 1$ ($\exists\, x_j$): $\dfrac{\partial_{x_i}\partial_\theta \ln L}{\partial_{x_j}\partial_\theta \ln L} = \dfrac{\partial_{\hat{\theta}}\partial_\theta g}{\partial_{\hat{\theta}}\partial_\theta g}\, \dfrac{\partial_{x_i}\hat{\theta}}{\partial_{x_j}\hat{\theta}} = a(x)$ (indip. da θ)

$\Rightarrow \partial_x \partial_\theta \ln L \equiv a(x, \hat{\theta}(x))\, b(\theta)$
Integrando: $\ln L = \ln f(x|\theta) = c_1(x)\, c_2(\theta) + k_1(x) + k_2(\theta)$

Si dimostra anche che la statistica sufficiente è $c_1(x)$.

LEGGE DEI GRANDI NUMERI IN FORMA DEBOLE

n variabili aleatorie X_i i.i.d. se $\exists\, \mu : E(X_i) = \mu, \Rightarrow$
$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) = 0$

Dimostrazione

1 - se le X_i hanno varianza finita:

$Var(\bar{x}) = \frac{\sigma^2}{n}$ si applica la disug. di Chebyshev:

$P(|\bar{x}-\mu| \geq k\,\sigma/\sqrt{n}) \leq \frac{1}{k^2}$ chiamiamo: $\varepsilon \equiv k\frac{\sigma}{\sqrt{n}} \Rightarrow k^2 = n\frac{\varepsilon^2}{\sigma^2}$

$\Rightarrow P(|\bar{x}-\mu| \geq \varepsilon) < \frac{\sigma^2}{n\varepsilon^2}$ che tende a 0 per $n \to \infty$, cvd.

2 - se $\nexists$ o non è finita $Var(x_i)$:

Trasliamo le variabili per avere $\mu \equiv 0$.

Dalle n misure costruiamo due serie di dati:

$\forall k = 1 \dots n, \begin{cases} U_k \equiv X_k & \text{se } |X_k| \leq nc \\ U_k \equiv 0 & \text{se } |X_k| > nc \end{cases} \wedge \begin{cases} V_k \equiv 0 & \text{se } |X_k| \leq nc \\ V_k \equiv X_k & \text{se } |X_k| > nc \end{cases}$

con c costante arbitraria.

Sse vale la LLN, $\lim_{n\to\infty} P\left(\left|\sum_{i=1}^{n} U_i\right| > \frac{n\varepsilon}{2}\right) = \lim_{n\to\infty} P\left(\left|\sum_{i=1}^{n} V_i\right| > \frac{n\varepsilon}{2}\right) = 0.$

$E(|X_k|) \equiv a \Rightarrow E(U_i^2) = \sum_{|x_i|<nc} x_i^2 f(x_i) \leq nc \sum_{|x_i|<nc} |x_i| f(x_i) \leq nca$

X_i iid $\Rightarrow U_i$ iid $\Rightarrow Var\left(\sum_{i=1}^{n} U_i\right) = n\, Var(U_i)$

Inoltre: $Var(U_i) = E(U_i^2) - [E(U_i)]^2$ ma $E(U_i) \xrightarrow[n\to\infty]{} E(x_i) = 0$

$\Rightarrow Var\left(\sum_{i=1}^{n} U_i\right) \xrightarrow[n\to\infty]{} n\, E(U_i^2) \leq n^2 ca$

Applichiamo la disug. di Chebyshev a $\sum U_i$: $P(|y - E(y)| \geq b) \leq \frac{Var(y)}{b^2}$

$\Rightarrow P\left(\left|\sum_{i=1}^{n} U_i\right| > \frac{n\varepsilon}{2}\right) \leq \frac{n^2 ca}{n^2\varepsilon^2/4}$ Questo vale $\forall c \Rightarrow$ possiamo anche scegliere $c \to 0$, dimostrando la prima disuguaglianza.

$P\left(\sum_{i=1}^{n} V_i > 0\right) \leq \sum_{i=1}^{n} P(V_i > 0) = n P(V_i > 0) = n P(|x_i| > nc) =$

$= n\sum_{|x_i|>nc} f(x_i) \leq n \sum_{|x_i|>nc} \frac{|x_i|}{nc} f(x_i) = \frac{1}{c}\sum_{|x_i|>nc} |x_i| f(x_i) \xrightarrow[n\to\infty]{} 0$

poiché il valore di aspettazione di $|x_i|$ è finito, cvd.

Unicità estimatori ML

- Se $\exists$ uno stimatore efficiente, la soluzione ottenuta col metodo della massima verosimiglianza è unica.
- dimostrazione

$\hat{\theta}$ efficiente $\Rightarrow$ $\frac{\partial \ln L}{\partial \theta} = A(\theta)\left(\hat{\theta} - \tau(\theta) - b(\theta)\right)$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \left(\hat{\theta} - \tau(\theta) - b(\theta)\right)\frac{\partial A(\theta)}{\partial \theta} + A(\theta)(-1)\left(\frac{\partial \tau(\theta)}{\partial \theta} + \frac{\partial b(\theta)}{\partial \theta}\right)$$

$\frac{\partial \tau(\theta)}{\partial \theta} + \frac{\partial b(\theta)}{\partial \theta} = A(\theta) \cdot Var(\hat{\theta})$ se $\hat{\theta}$ è efficiente,

quindi:

$$\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = 0 - A^2(\theta) Var(\hat{\theta}) \leq 0$$

$\ln L$ è una funzione convessa a concavità negativa, ed ha un solo massimo (almeno nelle vicinanze di θ).

$\Rightarrow \exists!\ \max(L) \Rightarrow \exists!\ \hat{\theta}$.

$E\left(\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$

- dimostrazione

$\int d\bar{x}\, f(\bar{x}, \theta) = 1 \Leftrightarrow \int d\bar{x}\, L(\theta|x) = 1$ Derivo:

$0 = \int d\bar{x} \frac{\partial L}{\partial \theta} = \int d\bar{x} \frac{\partial \ln L}{\partial \theta} L$ Derivo ancora:

$0 = \int d\bar{x} \frac{\partial^2 L}{\partial \theta^2} = \int d\bar{x} \left[L \frac{\partial^2 \ln L}{\partial \theta^2} + \frac{\partial \ln L}{\partial \theta}\frac{\partial L}{\partial \theta}\right]$ Quindi:

$$\int d\bar{x}\, L\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = -\int d\bar{x}\left(\frac{\partial \ln L}{\partial \theta}\right)\left(L \frac{\partial \ln L}{\partial \theta}\right) = -\int d\bar{x}\, L\left(\frac{\partial \ln L}{\partial \theta}\right)^2$$

$$\Rightarrow E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = -E\left(\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right)$$

Normalità asintotica stimatori ML

Se $\hat{\theta}$ è uno stimatore ML di θ, allora $\hat{\theta} \sim N(\theta, \sigma^2_{\hat{\theta}})$
con $\sigma^2_{\hat{\theta}} = \dfrac{1}{E\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right)}$

- dimostrazione

$$\int L\, dx = 1 \Rightarrow \int L \frac{\partial \ln L}{\partial \theta} dx = 0 \Rightarrow E\left(\frac{\partial \ln L}{\partial \theta}\right) = 0$$

Espandiamo con Taylor $\frac{\partial \ln L}{\partial \theta}$ attorno a θ (attorno al max della verosim.):

$$0 = \left.\frac{\partial \ln L}{\partial \theta}\right|_{\hat{\theta}} = \left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta} + (\hat{\theta}-\theta)\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta' \in (\theta, \hat{\theta})} \Rightarrow \left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta} = (\hat{\theta}-\theta)\left(-\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta'}\right)$$

Calcoliamo la varianza di $\frac{\partial \ln L}{\partial \theta}$:

$$\mathrm{Var}\left(\frac{\partial \ln L}{\partial \theta}\right) = E\left(\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right) - \left(E\left(\frac{\partial \ln L}{\partial \theta}\right)\right)^2 = E\left(\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right) = E\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right)$$

Per il teorema del limite centrale, dato che $\ln L$ è additivo al crescere del numero delle misure,

$$\frac{\partial \ln L - E(\partial \ln L)}{\sqrt{\mathrm{Var}(\partial \ln L)}} \xrightarrow{\text{Distribuzione}} N(0,1) \xleftarrow{\text{dist.}} \frac{\frac{\partial \ln L}{\partial \theta}}{\sqrt{\mathrm{Var}\left(\frac{\partial \ln L}{\partial \theta}\right)}} \quad (E=0)$$

Gli stimatori ML sono consistenti $\Rightarrow \hat{\theta}, \theta' \to \theta$
quindi $\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta'} \to E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$

$\Rightarrow$ Dall'espansione con Taylor $\frac{\partial \ln L}{\partial \theta} = (\hat{\theta}-\theta) E\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right)$

Quindi:

$$\frac{\partial_\theta \ln L}{\sqrt{E(-\partial^2_\theta \ln L)}} = \frac{(\hat{\theta}-\theta)\, E(-\partial^2_\theta \ln L)}{\sqrt{E(-\partial^2_\theta \ln L)}} = (\hat{\theta}-\theta)\sqrt{E(-\partial^2_\theta \ln L)} \sim N(0,1)$$

$$\Rightarrow \hat{\theta} \sim N\left(\theta, \frac{1}{E(-\partial^2_\theta \ln L)}\right)$$

$$\chi^2 \equiv (y-A\theta)^T \frac{1}{\sigma^2}(y-A\theta) \qquad y - A\theta \equiv y - A\theta + A\hat{\theta} - A\hat{\theta} = y - A\hat{\theta} + A(\hat{\theta}-\theta)$$

$$\chi^2 = \frac{1}{\sigma^2}\left[(y-A\hat{\theta})+A(\hat{\theta}-\theta)\right]^T\left[(y-A\hat{\theta})+A(\hat{\theta}-\theta)\right] =$$

$$= \frac{1}{\sigma^2}\Big\{(y-A\hat{\theta})^T(y-A\hat{\theta}) + (\theta-\hat{\theta})^T \cancel{A^T(y-A\hat{\theta})} + \cancel{(y-A\hat{\theta})^T}A(\theta-\hat{\theta}) + (\theta-\hat{\theta})^T A^T A(\theta-\hat{\theta})\Big\}$$

I termini misti:

$$(\theta-\hat{\theta})^T A^T (y-A\hat{\theta}) = (y-A\hat{\theta})^T A(\theta-\hat{\theta}) \qquad \hat{\theta} = (A^TA)^{-1}A^T y$$

$$\cancel{= y^T A(\theta-\hat{\theta}) - \hat{\theta}^T A^T A(\theta-\hat{\theta})} =$$

$$\cancel{= y^T A[\theta-(A^TA)^{-1}A^Ty] - y^T A (A^TA)^{-1} A^T [\theta-(A^TA)^{-1}A^Ty]} =$$

$$\cancel{= y^T A\theta - y^T A(A^TA)^{-1}A^T y - y^T A(A^TA)^{-1}A^T\theta + y^T A(A^TA)^{-1}A^T(A^TA)^{-1}A^T y}$$

$$= \left(y - A(A^TA)^{-1}A^T y\right) A(\theta-\hat{\theta}) = \left(y - AA^{-1}A^{T^{-1}}A^T y\right)A(\theta-\hat{\theta}) = 0$$

Rimane:

$$\chi^2 = \frac{1}{\sigma^2}\Big[\underbrace{(y-A\hat{\theta})^T(y-A\hat{\theta})}_{\equiv \chi^2_{MIN}} + \underbrace{(\theta-\hat{\theta})^T A^T A(\theta-\hat{\theta})}_{\equiv \chi^2_{\theta}}\Big]$$

indipendente da θ

- caratteristiche stimatori
- " " ML
- dimostr. unicità
- min χ^2 (e ML)
- costruzione int. conf