STATISTICA

MODELLO STATISTICO

> $X_1 \dots X_n$ campione casuale v.a. iid
> Immagino che la legge della variabile
> $\mathcal{L}(X_i)$ dipenda da $\vec{\vartheta} \in \mathbb{R}^k$ $\quad \vec{\vartheta}$ incognito

($\mathcal{L}(X_i)$ è in corrispondenza biunivoca con $F_{X_i}(t)$ f.d.ripartizione,
- ci mettiamo in un sottospazio, quindi ci basta fare inferenza su un numero finito)

$X_1 \dots X_n \quad X_i \sim Be(p) \quad \vartheta = p \in [0,1]$

$X_1 \dots X_n \quad X_i \sim N(\mu, \sigma^2) \quad \vec{\vartheta} = (\vartheta_1, \vartheta_2) = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$

COS'È UN **MODELLO STATISTICO** IN QUESTO AMBITO PARAMETRICO?

> È LA TRIPLETTA
> $$(\mathbb{R}^n, \underbrace{\mathcal{B}(\mathbb{R}^n)}_{\sigma\text{ algebra dei boreliani}}, \{P_{\vec{\vartheta}}, \vec{\vartheta} \in \Theta \subset \mathbb{R}^k\})$$

per ogni scelta dei possibili valori dei parametri ho lo spazio dove vive il mio modello

lo sperimentatore ha a disposizione la realizzazione del vettore
$x_1 \dots x_n$ dati $\rightsquigarrow$ con questi dati vogliamo dare info su $\vec{\vartheta}$.

|| lo statistico deve utilizzare la realizzazione casuale x dare informazione sul modello statistico.

Def STATISTICA

| $Y = T(X_1 \dots X_n)$ FUNZIONE DELLE SOLE V.A. È UNA STATISTICA

es
campione gaussiano
$X_1 \dots X_n \sim N(\mu, \sigma^2)$
$\bar{X}_n$ È UNA STATISTICA

$\dfrac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$ non è una STATISTICA perché dipende anche da μ e σ che sono parametri incogniti

Come faccio a conoscere i parametri incogniti? Usando la statistica

$T(X_1 \dots X_n) = \frac{1}{n} \sum X_i$ STATISTICA MEDIA CAMPIONARIA

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X - \bar{X}_n)^2$ STATISTICA VARIANZA CAMPIONARIA

$\sqrt{S^2} = S$

$X_{(1)} = \min(X_1 \dots X_n)$ STATISTICA: minimo

$X_{(n)} = \max(X_1 \dots X_n)$ STATISTICA: massimo

(come si calcolano min e max conoscendo una v.a?)

$T(X_1 \dots X_n): \Omega \to \mathbb{R}$
ma potremmo in realtà anche avere delle statistiche bidimens.
$(T_1, T_2): \Omega \to \mathbb{R}^2$

- Il campione stesso è una statistica, ma mentre le altre statistiche "summarize" i dati, riproducono una sintesi, il campione stesso non lo fa.

Dobbiamo scegliere la statistica in modo da riuscire a definire il parametro θ.

PRINCIPIO DI SUFFICIENZA

DATO UN CAMPIONE CASUALE $X_1 \dots X_n$
LA CUI LEGGE $\mathcal{L}(X_i)$ DIPENDE DA $\vec{\theta}$
LA STATISTICA $T(\vec{x})$ È SUFFICIENTE PER $\vec{\theta}$
QUANDO OGNI INFERENZA SU $\vec{\theta}$ DIPENDE DA $\vec{x}$
SOLO TRAMITE $T(\vec{x})$.

Ossia se ho due realizzazioni del campione $\vec{x}, \vec{y}$ diverse
$T(\vec{x}) = T(\vec{y})$ l'inferenza che faccio su $\vec{\theta}$ è la stessa
se $T(\vec{x})$ è sufficiente
Se ho due campioni diversi ma la media aritmetica è la stessa
e la statistica è sufficiente allora la media è uguale
la + banale statistica sufficiente è il campione intero

Def STATISTICA SUFFICIENTE

$T(\vec{x})$ È STATISTICA SUFFICIENTE PER $\vec{\theta}$
SE LA DISTRIBUZIONE CONDIZIONATA a $T(\vec{x}) = t$
NON DIPENDE DA $\vec{\theta}$ QUALUNQUE SIA t.

LA LEGGE di $(X_1 \dots X_n)$ condizionata a $T(x) = t$

ESEMPIO

X_1, X_2 iid $Be(p)$
prendiamo $T(\vec{x}) = X_1 + X_2$
La legge congiunta del campione dipende solo da p.
$T(\vec{x}) = t \qquad t = 0, 1, 2$

$P(X_1, X_2 \mid t = 0)$
diventa

$P(X_1 = 0, X_2 = 0 \mid t = 0) = 1$

$P(X_1 = 1, X_2 = 1 \mid t = 2) = 1$

$$P(X_1 = 1, X_2 = 0 \mid t = 1) = \frac{\mathbb{P}(X_1 = 1, X_2 = 0, T = 1)}{P(T = 1)} = \frac{p(1-p)}{\binom{2}{1} p(1-p)} = \frac{1}{2}$$

idem per

$P(X_1 = 0, X_2 = 1 \mid t = 1) = \frac{1}{2}$

$T \sim Bi(2, p)$
xk è somma di 2 Bernoulli

la legge condizionata alla somma è degenere agli estremi,
vale $\frac{1}{2}$ ~~Questa~~ per $t=1$.

Ma non dipende da p $\Rightarrow$ La legge è sufficiente

Ho tutte le info necessarie x parlare di θ

Ma mettersi a calcolare la legge condizionata potrebbe essere un bel pasticcio
quindi di fatto come controllo se la statistica è sufficiente?

TEOREMA di FATTORIZZAZIONE

SIA $f(\vec{x}, \vec{\theta})$ LA DENSITÀ CONGIUNTA (discreta o continua)
DEL CAMPIONE $\vec{X}$.

UNA STATISTICA $T(\vec{x})$ È SUFFICIENTE PER $\vec{\theta}$

$$\Updownarrow$$

$\exists$ DUE FUNZIONI $g(t, \theta)$, $h(\vec{x})$

t.c.

$\forall \vec{x}, \vec{\theta}$ LA DENSITÀ CONGIUNTA

$$f(\vec{x}, \vec{\theta}) = g(T(x), \theta) \cdot h(\vec{x})$$

$\uparrow$ g dipende da $\vec{x}$ solo tramite T

Interessante è che è un $\Leftrightarrow$

la densità $f(\vec{x}, \theta): \mathbb{R}^n \to \mathbb{R}$
- dentro la f ho dipendenza da θ, è equivalente a scrivere $f_\theta(\vec{x})$,

la densità, quindi la probabilità quindi la media dipende da θ

$E_\theta[\vec{x}]$ (Potrebbe essere che come notazione si trovi $f(\vec{x}|\theta)$, Approccio Bayesiano che vede θ come v.a)

<u>dim</u> Nel CASO DISCRETO.

- SIA $T(\vec{x})$ STATISTICA SUFFICIENTE

  devo mostrare che posso fattorizzare la legge congiunta.

$$f(\vec{x}, \vec{\theta}) = P_\theta(\vec{X} = \vec{x}) \underset{\text{TEO BAYES}}{=} \underbrace{P_\theta(\vec{X} = \vec{x} \mid T(\vec{X}) = T(\vec{x})}_{\substack{\text{Poiché la statistica} \\ \text{è sufficiente, non dipende} \\ \text{da } \theta \\ h(\vec{x})}} \overset{\text{osservazione}}{)} \underbrace{P(T(\vec{X}) = T(\vec{x}))}_{\substack{g(T(x), \theta) \\ \text{dipende} \\ \text{da } \theta, \text{ ma} \\ \text{tramite } T(x). \\ g \text{ è la legge di } T.}}$$

- SAPENDO che VALE LA FATTORIZZAZIONE DOBBIAMO DIMOSTRARE CHE $T(x)$ È SUFFICIENTE

  Assumiamo che valga la fattorizzazione
  $$f(\vec{x}, \theta) = h(\vec{x}) \, g(T(\vec{x}), \theta)$$
  Chiamiamo
  $$A_{T(\vec{x})} = \{\vec{y} : T(\vec{y}) = T(\vec{x})\}$$
  Calcolo la legge di
  $$\vec{X} | T = \frac{f(\vec{x}, \theta)}{q(T(\vec{x}), \theta)} = \frac{h(\vec{x}) g(T(\vec{x}), \theta)}{q(T(\vec{x}), \theta)} =$$

$p(X=x, T=t)$ non è nulla solo quando $T=T(x)$

$\leftarrow$ marginale (legge di T)

$$= \frac{g(T(\vec{x}),\theta)\, h(x)}{\sum_{y \in A_{T(x)}} f(\vec{y},\theta)} = \frac{g(T(\vec{x}),\theta)\, h(x)}{\underbrace{\sum_{y \in A_{T(x)}} g(T(\vec{y}),\theta)\, h(\vec{y})}}$$

ma sommando su questo insieme ho che $g(T(y),\theta)$ sono tutti $= g(T(x),\theta)$

$$= \frac{g(T(\vec{x}),\theta)\, h(\vec{x})}{g(T(\vec{x}),\theta) \sum_{y \in A_{T(\vec{x})}} h(\vec{y})}$$

ma $h(\vec{x})$ e $h(\vec{y})$ non dipendono da θ

$\Rightarrow$ la legge del campione condizionata a T non dipende da θ e quindi è sufficiente.

■

es

$T(\vec{x})$ = somma facce

$\vec{x} = (x_1, x_2)$

$x_i = 1 \dots 6$

$P(T(\vec{x}) = 4) = P(\vec{x} = (1,3)) + P(\vec{x} = (2,2)) + \dots$

↖ qst è la f.

Esempi

1) $\boxed{X_1 \dots X_n \sim U_{[0,\theta]}}$

È uniforme, quindi $f(x,\theta) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)$

La congiunta è

$$f(\vec{x},\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(x_i)$$

ma vorrei scriverlo come statistica...

Un prodotto di funzioni indicatrici è una f. indicatrice

$$\prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(x_i) = \begin{cases} 1 & 0 \le x_i \le \theta \ \forall i \\ 0 & \text{altrimenti} \end{cases}$$

$$\parallel$$

$$\mathbb{1}_{[0,+\infty)}(X_{(1)}) \cdot \mathbb{1}_{[-\infty,\theta]}(X_{(n)}) \quad \circ \quad \mathbb{1}_{[0,\theta]} \max(x_i)$$

$$= \frac{1}{\theta^n} \mathbb{1}_{[0,\theta]}(\max(x_i))$$

$\Rightarrow$ la densità congiunta è la fattorizzazione di max e funzione di θ

$\Downarrow$

$X_{(n)}$, il max, è statistica sufficiente.

2) $\boxed{X_1 \dots X_n \sim P(\lambda)}$ $\quad f = \dfrac{e^{-\lambda}\lambda^x}{x!}$

la legge congiunta

$$f(\vec{x},\lambda) = \frac{e^{-\lambda n}\prod_{i=1}^{n}\lambda^{x_i}}{\prod x_i!}\prod_{i=1}\mathbb{1}_{[N]}(x_i) =$$

$$= \frac{1}{\prod x_i!}\,e^{-\lambda n}\,\lambda^{\sum x_i}\prod \mathbb{1}_{(N)}(x_i)$$

$h(x)$

$T(x) = \sum x_i$ è sufficiente.

3) $\boxed{X_1 \dots X_n \sim N(\mu,\sigma^2)}$ $\quad \vec{\theta} = (\mu,\sigma^2)$

densità congiunta

$$f(\vec{x},\vec{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} =$$

$\pm \bar{X}_n$ media aritmetica *

$$= \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum x_i^2 - 2\mu\sum x_i + \mu^2\right]\right\}$$

$T(\vec{x}) = (\sum x_i^2, \sum x_i)$ È SUFFICIENTE

*

$$= \frac{1}{\sqrt{2\pi\sigma^2}^{\,n}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_i (x_i-\bar{x}_n)^2 + n(\bar{x}_n-\mu)^2 + \underbrace{2(\bar{x}_n-\mu)\sum(x_i-\bar{x}_n)}_{=0}\right.\right.$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}^{\,n}}\exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)S^2 + n(\bar{x}_n-\mu)^2\right]\right\}$$

Per il criterio di fattorizzazione

$T(\vec{x}) = (S^2, \bar{X}_n)$ È SUFFICIENTE

CASPITA MA ALLORA di statistiche e di statistiche sufficienti ce ne sono tante!

È perché vale qst risultato:

$T(\vec{x})$ sufficiente e r funzione biunivoca

$$\Downarrow$$

$T^*(\vec{x}) = r(T(\vec{x}))$ È sufficiente

xk $T(\vec{x}) = h(\vec{x})\,g(T(\vec{x}),\theta)$ ma r biunivoca $\Rightarrow T(\vec{x}) = r^{-1}(T^*(\vec{x}))$

$r^{-1}(T^*)$ $= h(\vec{x})\,g(r^{-1}(T^*(\vec{x})),\theta) = h(\vec{x})\,g^*(T^*(\vec{x}),\theta)$

Quindi di fatto posso provare la sufficienza dell'una o dell'altra coppia e poi trovando la n dimostrare che sono suff entrambe

Def GENERALE DI DENSITÀ PER UNA FAMIGLIA ESPONENZIALE

X HA LEGGE CHE APPARTIENE ALLA FAMIGLIA ESPONENZIALE se

$$f(x,\vec{\theta}) = h(x)\, c(\vec{\theta}) \exp\left[\sum_{j=1}^{k} w_j(\vec{\theta})\, t_j(x)\right]$$

Se presa una legge, posso scriverla così, allora la variabile appartiene alla famiglia esponenziale.

SE VALE QUESTO

$$\vec{T}(\vec{X}) = \left(\sum_{i=1}^{n} t_1(x_i), \dots, \sum_{i=1}^{n} t_k(x_i)\right) \text{ È SUFFICIENTE}$$

la congiunta sarà un prodotto di $h(x)$, un prodotto di c, poi

$$f(\vec{x}, \ ) = \dots \exp\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{k} w_j(\theta)\cdot t_j(x_i)\right)\right]$$

scambio le sommatorie e ho le mie statistiche.

ESEMPI

1) $\boxed{X \sim Be(p)}$

$\exp(\cancel{x} \ln p) + \exp((1-\cancel{x}) \ln(1-\cancel{p}))$

$$f(x,p) = p^x (1-p)^{1-x} \, \mathbb{1}_{\{0,1\}}(x)$$

$$= \underbrace{\mathbb{1}_{\{0,1\}}(x)}_{h(x)} \underbrace{(1-p)}_{c(p)} \exp\left\{ \underbrace{x}_{t_1(x)} \underbrace{\log \frac{p}{1-p}}_{w_1(p)} \right\}$$

$T(x) = \sum_{i=1}^{n} t_1(x_i) = \sum x_i$ la somma è una statistica sufficiente.

$\log \frac{p}{1-p} = \log(\text{ODDS})$

ODDS $= 1$ qnd l'esperimento è bilanciato

$$\text{ODDS} = \frac{\text{prob successo}}{\text{prob insuccesso}}$$

2) $\boxed{X \sim P(\lambda)}$

$$f_X(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \mathbb{1}_{\mathbb{N}}(x) = \frac{\mathbb{1}_{\mathbb{N}}(x)}{x!} e^{-\lambda} \exp(x \log \lambda)$$

$T(\vec{x}) = \sum x_i$ è statistica suff

3) $\boxed{X \sim N(\mu, \sigma^2)}$

$$f_X(x, \vec{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left[x_i^2 - 2\mu x_i + \mu^2\right]\right\}$$

isoliamo qll che dipende solo dal parametro

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}}}_{c(\theta)} \exp\left[-\underbrace{x_i^2}_{t_1(x_i)} \underbrace{\frac{1}{2\sigma^2}}_{w_1(\vec{\theta})} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\vec{\theta})} \underbrace{x_i}_{t_2(x_i)}\right]$$

$\sum x_i$ e $\sum x_i^2$ sono soff.

4) $\boxed{X \sim U[0, \theta]}$

$$f(x, \theta) = \frac{1}{\theta} \underbrace{\mathbb{1}_{[0,\theta]}(x)}_{\text{non riesco mai a scriverla come exp.}}$$

Le leggi il cui supporto dipende da θ non stanno nella famiglia esponenziale.

Le statistiche sufficienti ci raccontano bene il nostro parametro θ.
Ma magari ce ne sono di migliori

Def STATISTICA SUFFICIENTE E MINIMALE

UNA STATISTICA SUFFICIENTE $T(\vec{x})$ È DETTA STATISTICA SUFFICIENTE MINIMALE SE PER OGNI ALTRA STATISTICA SUFFICIENTE $T'(x)$

$T(x)$ È FUNZIONE DI $T'(x)$

Quindi se $T'(\bar{x}) = T'(y) \Rightarrow T(\bar{x}) = T(\bar{y})$

- Tutto il campione è statistica soff. ma non minimale.

Ma non è una definizione applicabile!

TEOREMA di LEMANN - SCHEFFE

SIA $f(\vec{x}, \vec{\theta})$ la LEGGE CONGIUNTA del VETTORE $\vec{x}$.
SUPPONIAMO CHE $\exists$ una funzione $T(\vec{x})$
tc $\forall \vec{x}, \vec{y}$ coppia di vettori
IL RAPPORTO $\dfrac{f(\vec{x},\vec{\theta})}{f(\vec{y},\vec{\theta})}$
È COSTANTE in θ $\iff T(\vec{x}) = T(\vec{y})$
$\Rightarrow T(\vec{x})$ È STATISTICA SUFFICIENTE MINIMALE PER $\vec{\theta}$.

dim

- Assumiamo di essere nel supporto della densità $f(\vec{x},\vec{\theta}) > 0 \quad \forall \vec{x}, \forall \vec{\theta}$.

Sia $A_t = \{ \vec{x} : T(\vec{x}) = t \}$
$\forall A_t$ scegliamo un elemento $\vec{x}_t \in A_t$.
Ovviamente $\vec{x}$ e $\vec{x}_{T(\vec{x})}$ sono nello stesso A_t.
sta nella curva di livello di $T(x)$ nella stessa curva di livello
$\Rightarrow$ per l'Hp del teorema
$\dfrac{f(\vec{x},\theta)}{f(\vec{x}_{T(\vec{x})},\vec{\theta})}$ è costante rispetto a θ.
Quindi posso dire

$$f(\vec{x},\theta) = \underbrace{f(\vec{x}_{T(x)},\vec{\theta})}_{g(T(\vec{x}),\theta)} \cdot \underbrace{\frac{f(\vec{x},\vec{\theta})}{f(\vec{x}_{T(x)},\vec{\theta})}}_{\text{è x Hp solo funzione della } \vec{x}, \text{ è cost in } \theta}$$

Per teo di fatt
ho dimostrato
che la statistica è sufficiente.

- SIA $T'(\vec{x})$ un'ALTRA statistica sufficiente. Allora
$\exists h', g'$ tc $f(\vec{x},\vec{\theta}) = h'(\vec{x})\, g(T'(\vec{x}),\vec{\theta})$.
Prendiamo
$\vec{x}, \vec{y}$ tc $T'(\vec{x}) = T'(y)$

$$\frac{f(\vec{x},\vec{\theta})}{f(\vec{y},\vec{\theta})} = \frac{g'(\cancel{T'(\vec{x})},\theta)\, h'(\vec{x})}{g'(\cancel{T'(y)},\theta)\, h'(y)} = \frac{h'(\vec{x})}{h'(\vec{y})}$$

si semplifica
XK $T'(x) = T'(y)$ non dipende da θ

Allora x l'Hp del teorema
$T'(x) = T(y)$
Ma allora $T(x)$ è funzione di $T'(x) \Rightarrow T$ minimale

(T' è funzione di T se $T'(x) = T'(y)$
$\Downarrow$
$T(x) = T(y)$)

$\blacksquare$

Supponiamo che T sia sufficiente minimale
r biunivoca $\Rightarrow r(T(x))$ è suff. minimale

- la suff l'abbiamo vista
- per la minimalità:

sia W stat. suff. Allora $T(\vec{x})$ è funzione di W.
$\Rightarrow r(T(\vec{x}))$ è funzione di W

ESEMPIO

Consideriamo $X_1 \dots X_n \sim N(\mu, \sigma^2)$
Per applicare il teo devo fare il quoziente e vedere se riesco a non farlo dipendere da ϑ.

$$\frac{f(\vec{x},\vec{\vartheta})}{f(\vec{y},\vec{\vartheta})} = \frac{\exp\left[-\frac{1}{2\sigma^2}\left((n-1)S^2_x + n(\bar{x}_n-\mu)^2\right)\right]}{\exp\left[-\frac{1}{2\sigma^2}\left((n-1)S^2_y + n(\bar{y}_n-\mu)^2\right)\right]}$$

Se $S_x = S_y$ e $\bar{X}_n = \bar{Y}_n$
num e den sono uguali e così la frazione non dipende da ϑ.
Ma è l'unico modo possibile.
Quindi $\Leftrightarrow S_x = S_y$ e $\bar{X}_n = \bar{Y}_n$ $\frac{f(\vec{x},\vec{\vartheta})}{f(\vec{y},\vec{\vartheta})}$ non dipende da ϑ.
e quindi
$(S^2, \bar{X}_n)$ è sufficiente e minimale.
Posso dire
$(\sum x_i, \sum x_i^2)$ è suff minimale xk è funz. biunivoca.

Def ANCILLARITÀ

$\boxed{\text{UNA STATISTICA } S(\vec{x}) \text{ LA CUI LEGGE NON DIPENDE DA } \vec{\vartheta} \text{ È DETTA ANCILLARE}}$

Esempio
Si può dimostrare che prese
$X_1 \dots X_n \sim U_{[\vartheta, \vartheta+1]}$
$R = \text{Range} = X_{(n)} - X_{(1)} \sim Beta((n-1), 2)$
- che sia una Beta è confontante, xk è $U_{[\vartheta,\vartheta+1]}$

Quindi il Range non dipende da ϑ, ma solo dal campione.
Quindi è una statistica di ϑ, ma ancillare.
Il massimo e il minimo da soli ci parlano sì di ϑ
(sono tendenti a $\vartheta+1$ e ϑ)
ma la loro differenza invece mi parla di 1.

Guardiamo
$T(\vec{x}) = (X_{(1)}, X_{(n)})$

$f_X(x,\vartheta) = \mathbb{1}_{[\vartheta,\vartheta+1]}(x)$ ← non sta nella fam. exp.

$f(\vec{x},\vec{\vartheta}) = \prod_{i=1}^{n} \mathbb{1}_{[\vartheta,\vartheta+1]}(x_i)$ ← è 1 se $\forall i$ $\vartheta \le x_i \le \vartheta+1$
Quindi $\min x_i \geq \vartheta$
e $\max x_i \le \vartheta + 1$

Quindi
$f(\vec{x},\vartheta) = \mathbb{1}_{[x_{(n)}-1, x_{(1)}]}(\vartheta)$ ← il max −1 è x forza + piccolo del min.

Quindi $\vec{T}(\vec{x}) = (X_{(1)}, X_{(n)})$ è statistica sufficiente.

Per teo Scheffé

$$\frac{f(\vec{x},\theta)}{f(\vec{y},\theta)} = \frac{\mathbb{1}_{[X_{(n)}-1,\, X_{(1)}]}(\theta)}{\mathbb{1}_{[Y_{(n)}-1,\, Y_{(1)}]}(\theta)}$$

L'unico modo per buttare via θ è che max e min coincidano.
In tal caso avrei statistica sufficiente e minimale.
Verrebbe da pensare che due ancillari siano ⊥ (???)
∃ condizioni per cui sufficienti minimali ⊥̸ ancillari.

IL TEOREMA di Scheffé da in un botto sufficienza e minimalità.
SE verifica il teo di Scheffé ho sia minimalità che sufficienza

Def COMPLETEZZA

Consideriamo una famiglia di densità di probabilità $f(t,\theta)$ la famiglia di densità di $T(\vec{x})$ statistica
$T(\vec{x})$ è completa se $f(t,\theta)$ è completa, cioè se

$$E_\theta[g(T)] = 0 \ \forall\theta \Rightarrow P(g(T)=0) = 1$$

Completezza: $\int g(t) f(t,\theta) = 0 \ \forall\theta \iff g = 0$

la funzione deve essere la funzione identicamente nulla qualora la media della funzione sia nulla.

ESEMPIO

1) $X_1 \dots X_n \sim Be(p)$

$\sum X_i$ È SUFFICIENTE.
È completa?
$T(\vec{x}) = \sum X_i \sim Bi(n,p)$
La famiglia delle binomiali è completa?
se una funzione a media nulla è 0

Sia g tc $E_p(g(T)) = 0 \quad \forall p \in [0,1]$
$\forall p \in (0,1)$

$$\sum_{k=0}^{p} g(k)\binom{n}{k} p^k (1-p)^{n-k} = 0$$

$$\Updownarrow$$

$$(1-p)^n \sum_{k=0}^{n} g(k)\binom{n}{k} p^k (1-p)^{n-k} = 0$$

Quindi sia $n = \frac{p}{1-p}$

ottengo $\sum_{k=0}^{n} g(k)\binom{n}{k} r^k = 0 \quad \forall p$, quindi $\forall\, 0<r<\infty$

↑ sto dicendo che un polinomio è 0 sempre
Ciò accade $\Leftrightarrow$ sono 0 tutti i suoi coeff

$\Downarrow$

$g(k)=0 \quad \forall k = 0\ldots n$

Quindi g deve essere 0 su tutti gli interi.

Se $E_p(g(T))=0 \quad \forall p$

$g(k) = 0 \quad \forall k = 0 \ldots n$

$P(g(T)=0) = 1$

xk il supporto di T sono solo gli interi.

$g(T)$ sarà solo $g(0), g(1), \ldots$

Quindi devo guardare le masse solo sugli interi.

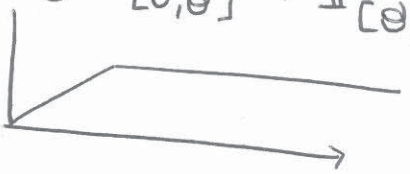
2) $X_1 \ldots X_n \sim U_{[0,\theta]}$ iid

dimostrare che $X_{(n)}$ è suff, min, compl.

Mi manca solo completo.

Devo scrivere la densità del massimo.

$F_{X_i}(t) = \frac{t}{\theta}\mathbb{1}_{[0,\theta]} + \mathbb{1}_{[\theta,+\infty)}$ fdr del massimo?



$F_{X_{(n)}}(t) = P(X_{(n)} \le t) = P(\bigcap_{i=1}^{n} X_i \le t) = \prod_{i=1}^{n} F_{X_i}(t) = (F_X(t))^n$

↑ il max è $\le t$ $\Downarrow$ lo sono tutte

Per il massimo si usa la fdr xk è lei elevata alla n nel caso iid

(Per il minimo si usa $1-F_X(t)$)

Quindi

$F_{X_{(n)}}(t) = \left(\frac{t}{\theta}\right)^n \mathbb{1}_{[0,\theta]} + \mathbb{1}_{[\theta,+\infty)}$

[$X_{(n)} \xrightarrow[P]{\mathcal{L}} \theta$ xk la fdr di $X_{(n)}$ converge a qll di θ]

$f_{X_{(n)}}(t) = \frac{nt^{n-1}}{\theta^n}\mathbb{1}_{[0,\theta]}$

$E_\theta[g(T)] = \int_0^\theta g(T)\frac{nt^{n-1}}{\theta^n}\,dt = 0 \quad \forall\theta$

voglio ottenere qualcosa che dipende da θ.

$\frac{d}{d\theta}E[g(T)] = 0 \quad \leftarrow$ xk $E[g(T)] = \text{cost}.$

$\Downarrow$

$\frac{d}{d\theta}\int_0^\theta g(T)\frac{nt^{n-1}}{\theta^n}\,dt = 0$

Quindi

$$= \underbrace{\left(\frac{d}{d\theta}\theta^{-n}\right)\theta^{n}\mathbb{E}[g(T)]}_{=0 \text{ xk } E(g(T))=0\ \forall\theta} + \frac{1}{\theta^{n}} n\theta^{n-1} g(\theta) = 0 \quad \forall\theta$$

Quindi

$$\frac{n}{\theta} g(\theta) = 0 \quad \forall\theta$$

$$\Downarrow$$

$P[g(T)=0] = 1$ Quindi $X_{(n)}$ è completa.

L'idea è che voglio sfruttare che l'integrale sia nullo e uso la derivata.

Quindi il massimo è stat suff minimale completa.

TEOREMA di BAHADOR

$\boxed{\text{UNA STATISTICA SUFFICIENTE E COMPLETA È ANCHE SUFFICIENTE E MINIMALE}}$

sufficienza e completezza mi regalano la minimalità

TEOREMA di BASU

$\boxed{\text{SE } T(\vec{x}) \text{ È STATISTICA SUFFICIENTE MINIMALE E COMPLETA} \Downarrow T(\vec{x}) \text{ È INDIPENDENTE DA OGNI STATISTICA ANCILLARE}}$

dim $S(\vec{x})$ È ANCILLARE ($\Rightarrow$ la sua legge non dipende da θ)

$\Downarrow$

$P(S(\vec{x})=s)$ non dipende da θ.

$P(S(\vec{x})=s \mid T(x)=t)$ non dipende da θ xk $T(x)$ è suff.

Per dimostrarlo voglio mostrare

$$S(\vec{x}) \perp T(\vec{x}) \iff \mathbb{P}(S(\bar{x})=s) = \mathbb{P}(S(x)=s \mid T(x)=t) \quad \forall t$$

$$\mathbb{P}(S(x)=s \mid T(x)=t) = \frac{\mathbb{P}(S=s, T=t)}{\mathbb{P}(T=t)}$$

(se la legge di S dipendesse da θ non potrei scriverlo)

Teorema delle probabilità totali

$$\mathbb{P}(S(\vec{x})=s) = \sum_t \mathbb{P}(S(\vec{x})=s \mid T(\vec{x})=t)\,\mathbb{P}(T(\vec{x})=t)$$

$$\mathbb{P}(S(\vec{x})=s) = \underbrace{\sum_t \mathbb{P}(S(\vec{x})=s)\cdot \mathbb{P}(T(\vec{x})=t)}_{=P(S(x)=s)\ \forall t}$$

xk $\sum_t P(T(x)=t) = 1$.

dunque ho che

$$\sum_t P(S(\vec{x})=s|T(\vec{x})=t)\cdot P(T(\vec{x})=t) =$$
$$= \sum_t P(S(\vec{x})=s) P(T(\vec{x})=t)$$

$$\sum g(t) P(T(\vec{x})=t)=0$$ ponendo $g(t)=\mathbb{P}(S(\vec{x})=s|T(x)=t)-\mathbb{P}(S(x)=s)$

$\Downarrow$

$E_\theta[g(T)]=0$

ma T è completa $\Rightarrow P(S(\vec{x})=s|T(x)=t)=\mathbb{P}(S(x)=s))$ $\forall t$

Quindi ho l'indipendenza di una stat minimale completa da una ancillare. ■

Esempio

$X_1 \dots X_n \sim N(\mu,\sigma^2)$

σ^2 noto

$\bar{X}_n$ è suff minimale completa per μ

S^2 è ancillare per μ, xk la sua legge è una χ^2, non entra μ nella legge di S.

Quindi $\bar{X}_n \perp S^2$

È PIÙ DIFFICILE DIMOSTRARE L'$\perp$ O LA COMPLETEZZA?

TEOREMA

$X_1 \dots X_n$ iid nella famiglia esponenziale

$\left(f(x,\theta)=h(x)g(\theta)\exp\left[\sum_{j=1}^{k} t_j(x)w_j(\theta)\right]\right)$

$\Rightarrow \vec{T}=\left(\sum_{i=1}^{n} t_1(x_i),\dots,\sum_{i=1}^{n} t_k(x_i)\right)$

È COMPLETA SE $\{w_1(\theta),\dots,w_k(\theta),\vec{\theta}\in\Theta\}$

CONTIENE ALMENO UN APERTO DI $\mathbb{R}^k$

$\Rightarrow$ T È SUFFICIENTE MINIMALE COMPLETO

Controllare questa richiesta è facile

Esempi

$X_1 \dots X_n \sim N(\mu,\sigma_0^2)$ σ_0^2 noto

$$f(x,\mu)=\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_0^2}\right]=$$

$$=\underbrace{\frac{1}{\sqrt{2\pi\sigma_0^2}}\underbrace{\exp\left[-\frac{1}{2}\frac{x^2}{\sigma_0^2}\right]}_{\substack{\text{dipende solo da x}\\ \text{x k } \sigma_0 \text{ noto}}}}_{\text{dipende solo da x}}\underbrace{\exp\left[-\frac{1}{2}\frac{\mu^2}{\sigma_0^2}\right]}_{\text{dipende solo da }\mu}\underbrace{\exp\left[\frac{\mu x}{\sigma_0^2}\right]}_{\substack{k=1\\ t_1(x)=x\\ w_1(\mu)=\mu/\sigma_0^2}}$$

$T(\bar{X}) = \sum_{i=1} x_i$ è sufficiente

$W_1(\mu)$ mappa $\mathbb{R}$ in $\mathbb{R}$

$W_1(\mu) : \mathbb{R} \to \mathbb{R}$

$\uparrow \mu \in \mathbb{R}$.

Quindi $Im(W_1(\mu)) = \mathbb{R}$ contiene aperto di $\mathbb{R}$

Quindi $\sum x_i$ è completo

Quindi $\sum x_i$ è sufficiente minimale completa

Ma io avevo la media campionaria.

È ancora sufficiente minimale completa?

Sì, xk la legge di $\bar{X}_n$ è ancora una Gaussiana, ho ~~non~~ solo un n a denominatore nella $f(x,\mu)$

UNA TRASFORMAZIONE BIUNIVOCA DI UNA STATISTICA COMPLETA È COMPLETA?

T completa $T^* = r(T)$

$\mathbb{E}[g(T^*)] = 0 = \mathbb{E}[g(r(T))] = 0$

$\Downarrow$

$P(g(r(T)) = 0) = 1 \Rightarrow P(g(T^*) = 0) = 1$

Quindi $E[g(T^*)] = 0 \Rightarrow P[g(T^*) = 0] = 1$

ESEMPIO

$X_1 \dots X_n \sim P(\lambda)$

trovare statistica suff minimale completa

$$f(x,\lambda) = \frac{\mathbb{1}_{\mathbb{N}}(x)}{x!} e^{-\lambda} \exp\{x \log \lambda\}$$

So che $\sum x_i$ è sufficiente.

$\log \lambda : \mathbb{R}^+ \to \mathbb{R}$

Quindi contiene almeno un aperto di $\mathbb{R}$

$\Rightarrow$ È sufficiente, completa, minimale.

POTREBBE SUCCEDERE CHE PER ES. L'IMMAGINE SIA UNA CURVA, SIA VINCOLATA, se ho $\mathbb{R} \to \mathbb{R}^2$

QUINDI HO CHE NON HO UN APERTO DI $\mathbb{R}^2$.

es: $N(\mu, \mu^2)$ mappa in una PARABOLA, che non contiene un aperto di $\mathbb{R}^2$

ESEMPIO

$X_1 \dots X_n \sim Be(p)$

$$f(x,p) = \mathbb{1}_{\{0,1\}}(x)(1-p)\exp\left\{x \log \frac{p}{1-p}\right\}$$

$$\log \frac{p}{1-p} : [0,1] \to (-\infty, +\infty)$$

$\uparrow$ contiene almeno un aperto di $\mathbb{R}$.

STIMARE i PARAMETRI

STIMA PUNTUALE

1) METODI PER STIMARE I PARAMETRI (stimatori)
2) METODI PER CONFRONTARE GLI STIMATORI.

Def STIMATORE PUNTUALE

UNO STIMATORE PUNTUALE PER θ È UNA QUALUNQUE FUNZIONE $W(X_1 \dots X_n) = W(\vec{X})$ del CAMPIONE

(è una qualunque statistica)

la statistica mi parla in generale ("tifoso")

lo stimatore puntuale mi parla bene solo di qualcosa ("tifoso della squadra X")

$$W(\vec{X}) \text{ È } V.A: \Omega \to \mathbb{R}$$

STIMA

dati $(x_1 \dots x_n) = (X_1, \dots, X_n)(\omega)$

realizzazione del campione in ω

LA STIMA È

$$W(\vec{X})(\omega) = W(x_1 \dots x_n) \leftarrow \text{È un numero}$$

Ho realizzato i dati campionando un modello.

La stima è la realizzazione dell'operatore usando quei dati

METODO dei MOMENTI (Pearson)

$X_1 \dots X_n$ iid

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \mu_1 = \mathbb{E}[X]$$

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \qquad \mu_2 = E[X^2]$$

$$\vdots$$

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k \qquad \mu_k = E[X^k]$$

MEDIE O MOMENTI EMPIRICI

MOMENTI TEORICI (tipicamente μ_j è funzione di $\vec{\theta}$)

SI OTTENGONO GLI STIMATORI $\hat{\theta}$ RISOLVENDO LE EQUAZIONI

$$m_1 = \mu_1(\theta_1 \dots \theta_k)$$

$$\vdots$$

$$m_k = \mu_k(\theta_1 \dots \theta_k)$$

Perché dovrebbe essere vero che i momenti campionari... portano ai quelli teorici?

ESEMPIO

$X_1 \dots X_n \sim N(\mu, \sigma^2) \qquad \vec{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2)$

$m_1 = \frac{1}{n} \sum X_i = \bar{X}_n \qquad \mu_1 = \mu$

$m_2 = \frac{1}{n} \sum X_i^2 \qquad \mu_2 = E[X^2] = \sigma^2 + \mu^2 = \theta_1^2 + \theta_2$

Devo risolvere le equazioni in θ_1 e θ_2

$m_1 = \mu_1 = \mu \Rightarrow \hat{\mu} = \hat{\theta}_1 = \bar{X}_n \leftarrow$ LO STIMATORE della MEDIA È LA MEDIA CAMPIONARIA

$m_2 = \sigma^2 + \mu^2 \Rightarrow \hat{\sigma}^2 = \hat{\theta}_2 = m_2 - \hat{\mu}^2$

$= \frac{1}{n} \sum X_i^2 - \bar{X}_n^2 = \frac{1}{n} \sum (X_i - \bar{X}_n)^2$

non è la varianza campionaria! Qst è lo stimatore proposto da Pearson.

IL NUMERO di UGUAGLIANZE UTILI DEVE EGUAGLIARE IL N° di PARAMETRI DA STIMARE

Ma l'ordine (il grado) potrebbe non coincidere con l'ordine dello spazio dei parametri:

ESEMPIO

$X_1 \dots X_n \sim U_{[-\theta, \theta]}$ devo stimare θ.

$m_1 = \frac{1}{n} \sum X_i = E[X] = 0 \leftarrow$ il momento primo non è più funzione di $\theta \Rightarrow$ devo andare al momento secondo

$m_2 = \frac{1}{n} \sum X_i^2 = E[X^2] = Var[X] = \frac{1}{12} 4\theta^2 = \frac{\theta^2}{3}$

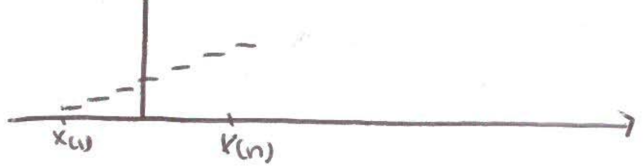
lo stimatore proposto è

$\hat{\theta} = \sqrt{3 \frac{1}{n} \sum X_i^2}$

lo spazio dei parametri θ ora in $\mathbb{R}$, ma devo andare al 2^ momento per trovare lo stimatore, e probabilmente è x qst che è bestto.

PERCHÉ PEARSON HA AVUTO QUESTA IDEA?

> FUNZIONE di RIPARTIZIONE EMPIRICA (DISCRETA)
>
> $$F_n(x) = \frac{1}{n} \sum \mathbb{1}_{[X_i \le x]}$$



ogni volta che ho x_i ho funzione che mi fa salire di $\frac{1}{n}$

Quindi assegna una probabilità $\frac{1}{n}$ a ogni punto del salto (È DISCRETA) una distribuzione discreta così ha

$\mathbb{E} = \frac{1}{n}\sum x_i$ ← media che corrisponde alla fdr di $F_n(x)$

↑ $E\left[\frac{1}{n} X(\omega)\right]$

E perché la media della fdr empirica dovrebbe assomigliare alla media?

PERCHÉ VALE IL TEOREMA di GLIVENKO-CANTELLI

$$\sup_x |F_n(x) - F(x)| \xrightarrow[qc]{} 0$$

[se si assomigliano le fdr si assomiglieranno tutte le altre cose]

mi aspetto che il valore atteso dell'empirica sia vicino alla teorica $\forall \omega$.

E così vale per i momenti

La fdr empirica va a schiacciarsi sulla fdr teorica.
E quindi così fanno la media e i momenti.
Se io ho v.a. discreta
X ha legge

x	p_i
x_1	$1/n$
$\vdots$	
x_n	$1/n$

$\mathbb{E}[X] = \sum x_i p_i = \frac{1}{n}\sum x_i$

Fdr empirica:

$F_{(n)}(x)$ È ALEATORIO, XK $\mathbb{1}_{[X_i \le x]}$ ← aleatoria

Con questo metodo riusciamo a risolvere operazioni essenzialmente algebriche

VEROSIMIGLIANZA (LIKELIHOOD)

DENSITÀ CONGIUNTA di θ VISTA COME PARAMETRO

$$\boxed{L(\theta, \vec{X}) = f(\vec{x}, \theta)}$$

nel discreto:

$P(\vec{X} = \vec{x}, \theta) \leftarrow f(\vec{x}, \theta)$

$L(\theta, \vec{X}) = P(\vec{X} = \vec{x}, \theta)$

↑ sia realizzato x, vogliamo vedere come al variare di θ cambia la probabilità.

Andrò a cercare i θ che aumentano la probabilità di osservare quello che ho osservato.

SE OSSERVO UNA COSA IMMAGINO CHE SIA PROBABILE che FOSSE POSSIBILE.

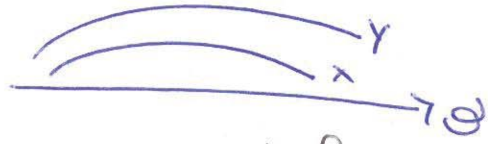
Per X fissato guardo come varia la probabilità di quell'X al variare di θ.

PRINCIPIO DI VEROSIMIGLIANZA

Se $\vec{x}, \vec{y}$ osservazioni del campione sono tc
$L(\theta, \vec{x})$ è proporzionale (in funzione di θ)
alla $L(\theta, \vec{y})$
(cioè $\exists\, c(\vec{x}, \vec{y})$ tc $L(\theta, \vec{x}) = c(\vec{x}, \vec{y})\, L(\theta, \vec{y}) \quad \forall \theta$

SI DICE CHE
$\vec{x}, \vec{y}$ SONO IN ACCORDO CON IL PRINCIPIO di VEROSIMIGLIANZA

PERCHÉ LE CONCLUSIONI CHE OTTENIAMO CONOSCENDO $\vec{x}$
SONO LE STESSE CHE OTTENIAMO CONOSCENDO $\vec{y}$



se ho un fattore di scala
le info su quanto è
probabile x, al variare di θ
è uguale a quanto è probabile y

l'inferenza su θ è la stessa.

Competitor del metodo dei momenti?

METODO di MASSIMA VEROSIMIGLIANZA

Def MLE = Maximum Likelihood Estimator

$\hat{\theta}_{MLE}(\vec{x})$ È IL VALORE del PARAMETRO θ
in cui $L(\theta, \vec{x})$ RAGGIUNGE IL MASSIMO

$$\hat{\theta}_{MLE}(\vec{x}) = \underset{\theta \in \Theta}{\operatorname{Arg\,Sup}}\, L(\theta, \vec{x})$$

non sempre riesco a trovare una soluzione chiusa al problema
Altrimenti devo andare a usare metodi numerici

Per definizione
$\hat{\theta}(\vec{x})$ ASSUME VALORI nel RANGE di AMMISSIBILITÀ,
cioè dove vive θ $\quad \theta \in \Theta$

Io ho $L(\theta, \vec{x})$ e trovo un $\hat{\theta}(\vec{x})$ che realizza il massimo.
stima.

Se ho per es. $\frac{1}{n} \sum x_i$
avrei
$\hat{\theta}(\vec{x}) = \frac{1}{n} \sum x_i$ ← stimatore

Può essere che non riesco a trovare il massimo in forma chiusa

Problema di massimizzazione ⇝ metodi numerici
che possono essere sensibili all'inizializzazione
⇒ metodi instabili
oppure potrebbero non convergere

Il primo tentativo che si fa è

$\frac{\partial}{\partial \theta} L(\theta, \vec{x}) = 0$ $\qquad$ $\frac{\partial L}{\partial \theta_i}(\theta_i, \vec{x}) = 0 \quad \forall i = 1 \dots k$

$\vec{\theta} = (\theta_1 \dots \theta_k)$

candidato a essere punto di max

Ma io ho

$L(\theta, \vec{x}) = \prod_{i=1}^{n} f(x_i, \theta)$

$\Rightarrow$ che casino fare le derivate di un prodotto!

Passo a $\log L(\theta, \vec{x}) = \ell(\theta, \vec{x}) = \sum_{i=1}^{n} \log f(x_i, \theta)$

$$\boxed{\hat{\theta}_{MLE} = \operatorname{Arg\,Sup}_{\theta} L(\theta, \vec{x}) = \operatorname{Arg\,Sup}_{\theta} \ell(\theta, \vec{x})}$$

xk il log è funzione monotona

Ma ci sono casi in cui non posso fare il logaritmo!

ESEMPI

1) Supponiamo di avere un'urna da cui estraiamo con reimissione 3 palline
p palline bianche
$1-p$ palline nere

Ottengo 2 palline bianche su 3.

$p = 1/2$ o $p = 1/3$?

Come faccio a prendere questa decisione?

CALCOLIAMO LA PROBABILITÀ CHE SUCCEDA CIÒ CHE È SUCCESSO.

$P_{p=\frac{1}{2}}(\sum X_i = 2)$ $\qquad$ $\sum X_i \sim Bi(3, \frac{1}{2})$

$= \binom{3}{2}\left(\frac{1}{2}\right)^3 = \frac{3 \cdot 2}{2} \cdot \frac{1}{8} = \frac{3}{8}$ $\leftarrow$ probabilità di vedere 2 palline bianche con reimissione

$P_{p=\frac{1}{3}}(\sum X_i = 2)$ $\qquad$ $\sum X_i \sim Bi(3, \frac{1}{3})$

$= \binom{3}{2}\left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{2}{9}$

$\frac{3}{8} > \frac{2}{9}$ $\Rightarrow$ la mia opinione mi dice di scommettere su 1/2, xk quel parametro, tra quei due è quello che mi rende + probabile quello che ho osservato.

Ma se invece dico ok tu hai visto $\frac{2}{3}$, quanto dici che vale p?

$\downarrow$

$X_1 \dots X_n$ iid $Be(p)$

- con il metodo dei momenti $\hat{p}_{momenti} = \frac{1}{n}\sum X_i = \frac{2}{3}$
- con la verosomiglianza

$L(p, x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \mathbb{1}_{\{0,1\}}(x_i) = p^y (1-p)^{n-y} \prod_{i=1}^{n} \mathbb{1}_{\{0,1\}}(x_i)$

$\ell(p, \vec{x}) \propto y \log p + (n-y)\log(1-p)$ $\qquad$ (derivo per p quindi tralascio ciò che non dipende da p)

$\frac{\partial \ell}{\partial p} = \frac{y}{p} - (n-y)\frac{1}{1-p}$

$$\frac{\partial \ell}{\partial p} = \frac{y - yp - np + yp}{p(1-p)} = 0 \iff p = \frac{y}{n}$$

- se $y = 0$ Avrei avuto la derivata sempre decrescente $\Downarrow$ $\hat{p} = 0$
- $y = n$ avrei avuto la derivata sempre crescente $\Downarrow$ $\hat{p} = 1$

Per $0 < y < n$ ho $p = \frac{y}{n}$, $\hat{p}_{MLE} = \frac{\sum x_i}{n}$ $(y = \sum x_i)$

Quindi di fatto $\hat{p}$ è la media campionaria delle mie osservazioni

2) $X_1 \dots X_n \sim U_{[0,\vartheta]}$

$\hat{\vartheta}_{MOM} = 2\frac{\sum x_i}{n}$ $\quad (\mu = \frac{\vartheta}{2})$

$\hat{\vartheta}_{MLE} = ?$

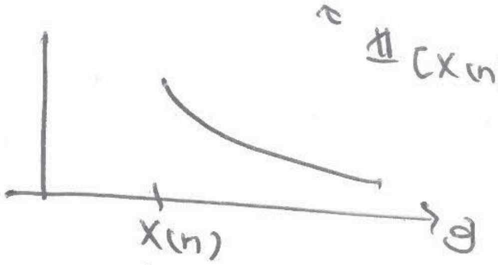
$$L(\vartheta, \vec{x}) = \frac{1}{\vartheta^n} \prod_{i=1}^{n} \mathbb{1}_{[0,\vartheta]}(x_i) = \frac{1}{\vartheta^n} \underbrace{\mathbb{1}_{[0,\vartheta]}(x_{(n)})}_{\text{qui c'è dentro } \vartheta!}$$

non ce la faccio con la derivata! ← Non posso toglierlo neanche con la logverosimiglianza!

Allora lavorando,

è $\frac{1}{\vartheta^n}$ se $0 \le x_{(n)} \le \vartheta$

$\mathbb{1}_{[x_{(n)}, +\infty)}(\vartheta)$



Quindi $\hat{\vartheta}_{MLE} = X_{(n)}$

$\hat{\vartheta}_{MLE} \neq \hat{\vartheta}_{MOM}$

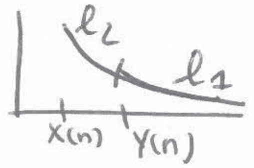
So che il massimo è suff. minimale completa

$\Rightarrow$ prendo $\hat{\vartheta}_{MLE}$

Se due campioni sono d'accordo con il principio di verosimiglianza trovo lo stesso $\hat{\vartheta}_{MLE}$, ed è anche x qst che decido qst di stimatore.

Se due punti sono in accordo con il PVS ho che le 2 curve sono proporzionali a una costante che non dipende da ϑ:

se avessi



$$l_1 = l_2 \underbrace{\mathbb{1}_{(Y_{(n)}, +\infty)}(\vartheta)}$$

la cost dipende da ϑ
$\Rightarrow$ non vale PVS

$\vec{x}$ e $\vec{y}$ sono in accordo con il principio di verosimiglianza
$$\Updownarrow$$
$$X_{(n)} = Y_{(n)}$$

Quindi $\hat{\theta}_{MLE}(\vec{x}) = \hat{\theta}_{MLE}(\vec{y})$
ma $\hat{\theta}_{MoM}(\vec{x}) \overset{?}{=} \hat{\theta}_{MoM}(\vec{y})$?
non necessariamente,
potrei avere $x = (1, 5)$ e $y = (2, 5)$
il massimo è $=$ ma la media aritmetica è $\neq$

3) $X_1 \dots X_n \sim N(\mu, \sigma_0^2)$

$$L(\mu, \vec{x}) = \frac{1}{(\sqrt{2\pi\sigma_0^2})^n} \exp\left[-\frac{1}{2\sigma_0^2} \sum (x_i - \mu)^2\right]$$

$$\ell(\mu, \vec{x}) \propto -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\frac{d\ell}{d\mu} = \frac{1}{2\sigma_0^2} \cdot 2 \sum_{i=1}^{n} (x_i - \mu) = 0 \iff \mu = \frac{\sum x_i}{n}$$

Quindi

$$\hat{\mu}_{MLE} = \frac{\sum x_i}{n} = \bar{X}_n$$

(calcolare $(\mu, \sigma^2)_{MLE}$)

Pro e Contro di MoM e MLE

1) $X_1 \dots X_K \sim Bin(K, p)$ $\quad \vec{\theta} = (K, p)$

$$\Theta = \{\mathbb{N} \times [0,1]\}$$

METODO DEI MOMENTI

$$\begin{cases} \bar{X}_n = Kp \\ \frac{1}{n} \sum x_i^2 = Kp(1-p) + K^2 p^2 \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \hat{p} = \frac{\bar{X}_n}{K} \\ \frac{1}{n} \sum x_i^2 = K \underbrace{\frac{\bar{X}_n}{K}}_{p} - K \underbrace{\frac{\bar{X}_n^2}{K^2}}_{p^2} + K^2 \underbrace{\frac{\bar{X}_n^2}{K^2}}_{p^2} \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \hat{p} = \frac{\bar{X}_n}{\hat{K}} \\ \bar{X}_n + \bar{X}_n^2 - \frac{1}{n} \sum x_i^2 = \frac{\bar{X}_n^2}{K} \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \hat{p} = \frac{\bar{X}_n}{\hat{K}} \\ \hat{K} = \frac{\bar{X}_n^2}{\bar{X}_n - \frac{1}{n} \sum (x_i - \bar{X}_n)^2} \end{cases}$$

stime...:
$\hat{k}$ deve darmi stime intere >0, xk $k \in \mathbb{N}$.
Ma non è detto che ciò accada!
Il problema è il range di ammissibilità.
Le stime possono $\notin \Theta$.

Anche con il metodo di verosimiglianza potrei avere problemi xk k una discreta quindi non avrei vera derivata

2) $X \sim U[0,\theta]$
$Y \sim U[0,\theta]$ Pro MOM

Non conosco il valore di X e Y,
ma conosco $T = (X-Y)^2$, ho la distanza presa da 2 valori a caso pescati da una uniforme

Calcolo la legge di $-Y \sim U[-\theta, 0]$
di $X-Y \sim$ △
e poi di $(X-Y)^2$

$$f_T(t) = \frac{\theta - \sqrt{t}}{\theta^2 \sqrt{t}} \mathbb{1}_{[0,\theta^2]}(t)$$

$$\mathbb{E}[T] = \frac{\theta^2}{6} \rightarrow \hat{\theta} = \sqrt{6 \frac{\sum T_i}{n}}$$

con la verosimiglianza

$$L(\theta, t) = \frac{1}{\theta^{2n}} \frac{1}{\prod_{i=1} \sqrt{t_i}} \prod_{i=1}^{n} (\theta - \sqrt{t_i}) \mathbb{1}_{[\max t_i, +\infty)}(\theta)$$

dovrei guardare il massimo. $\forall i$ $0 \leq t_i \leq \theta^2$
Guardo la logverosimiglianza

$$\ell(\theta, t) = -2n \log \theta + \sum_{i=1}^{n} \log(\theta - \sqrt{t_i})$$

$$\frac{\partial \ell}{\partial \theta} = -\frac{2n}{\theta} + \sum \frac{1}{(\theta - \sqrt{t_i})}$$

ma non riesco a calcolare il punto di massimo in forma chiusa!

In generale le stime le faremo con il metodo di verosimiglianza.

PROPRIETÀ di INVARIANZA degli MLE

SE $\hat{\theta}$ È MLE PER θ
ALLORA OGNI FUNZIONE $\tau(\theta)$ L'MLE di $\tau(\theta)$ è $\tau(\hat{\theta}_{MLE})$
$\forall$ τ biunivoca o univoca
(dim sul quad)

Θ voglio restringere
l'insieme dei possibili parametri Θ_0 restrizione di Θ

$$\hat{\Theta}^{0}_{MLE} = \underset{\theta \in \Theta^0}{\operatorname{Arg\,sup}} \; L(\theta, \vec{x})$$

$\hat{\theta}_{MLE}$ può essere uguale o $\neq$ da $\hat{\Theta}^{0}_{MLE}$

METODI per VALUTARE gli STIMATORI

> **MSE** Mean Square Error di uno Stimatore T di θ

$$\boxed{MSE_\theta(T) = E_\theta(T-\theta)^2}$$

Aggiungo e tolgo $E[T]$ — funzione di perdita.

$$MSE_\theta(T) = E_\theta\left[(T - E_\theta(T))^2\right] + E_\theta\left((E_\theta(T)-\theta)^2\right) + \underbrace{2E_\theta\left[(T-E(T))(E(T)-\theta)\right]}$$
$$= Var\theta + (E_\theta(T) - \theta)^2$$

fa 0 xk
$E[T]-\theta$ è un numero.
Mi resta
$2(E(T)-\theta)E[T-E(T)] = 0$

Def DISTORSIONE o BIAS
$E_\theta(T) - \theta$

$$\boxed{MSE_\theta(T) = Var_\theta(T) + \text{distorsione}^2}$$

se $E_\vartheta(T)=\vartheta$, ossia la distorsione è nulla,
allora si dice che T è un operatore non distorto
per ϑ: è UNO STIMATORE UNBIASED.

Bisogna cercare di minimizzare l'intero MSE, ma di solito
si ha l'effetto
"BIAS-VARIANCE TRADE OFF", ossia cerco di
minimizzare la varianza e mi aumenta la
distorsione, minimizzo la distorsione e mi aumenta
la variabilità.

Non sempre la soluzione giusta è andare a caccia di stimatori
non distorti

T_1 È PREFERIBILE a T_2 SE
$$MSE_\vartheta(T_1) \le MSE_\vartheta(T_2) \quad \forall \vartheta$$
è disuguaglianza tra 2 funzioni di ϑ!

In alcuni casi non riusciremo a decidere xk può essere
che per alcuni ϑ vada meglio uno, per altri ϑ meglio l'altro

ESEMPI

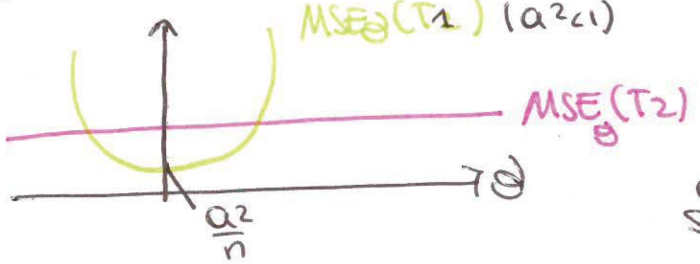
(1) $X_1 \dots X_n$ $\quad \vartheta = E[X_i] \quad \sigma^2 = Var[X_i] = 1$

$T_1 = a\bar{X}_n$

$T_2 = \bar{X}_n$

- $MSE_\vartheta(T_1) = Var(a\bar{X}_n) + (E[a\bar{X}_n]-\vartheta)^2 =$
$$= \frac{a^2}{n} + (a\vartheta - \vartheta)^2 =$$
$Var\,\bar{X}_n = \frac{Var[X_i]}{n}$ $\qquad E[\bar{X}_n] = E[X_i] = \vartheta$
$$= \frac{a^2}{n} + (a-1)^2\vartheta^2$$

- $MSE_\vartheta(T_2) = Var(\bar{X}_n) + 0 = \frac{1}{n}$



Se $a^2>1$ ok
vince T_2.
Ma per $a^2<1$
dipende da ϑ
quale stimatore
sia meglio.

L'ordinamento non è totale, nello spazio degli stimatori,
xk x qst 2 stimatori non so scegliere.

Questo in realtà è un caso un po' particolare xk all'aumentare
di n di fatto mi va quasi sempre meglio $\bar{X}_n$
e di fatto per stimare la media si usa sempre la
media campionaria

② $X_1 \dots X_n \sim N(\mu, \sigma^2)$

$$L(\mu, \sigma^2, \vec{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right)$$

$$\ell(\mu, \sigma^2, \vec{x}) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$$

tanto deriveremo rispetto σ^2

Problema di massimizzazione in 2 variabili.

$\Downarrow$

CERCO I PUNTI in cui si Annulla il Gradiente

$$\begin{cases} \frac{\partial \ell}{\partial \mu} = 0 & \to \frac{1}{\sigma^2} \sum (x_i - \mu) = 0 \\ \frac{\partial \ell}{\partial \sigma^2} = 0 & \to -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\mu} = \bar{X}_n \\ \hat{\sigma}^2 = \frac{\sum (x_i - \bar{X}_n)^2}{n} \end{cases}$$ $\leftarrow$ non ottengo la varianza campionaria!

Per avere un massimo almeno una delle due derivate seconde deve essere < 0 e il determinante dell'Hessiano > 0.

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n}{\sigma^2}$$ $\leftarrow$ ok! Almeno una delle due derivate è < 0!

$$\frac{\partial^2 \ell}{\partial \sigma^{2\,2}} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum (x_i - \mu)^2$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4} \sum (x_i - \mu)$$

$$\det \begin{bmatrix} \frac{\partial^2 \ell}{\partial \mu^2} & \frac{\partial^2 \ell}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \ell}{\partial \mu \partial \sigma^2} & \frac{\partial^2 \ell}{\partial \sigma^{2\,2}} \end{bmatrix} = \begin{vmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum (x_i - \mu) \\ -\frac{1}{\sigma^4} \sum (x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum (x_i - \mu)^2 \end{vmatrix}$$

$$= \frac{1}{\sigma^6} \left[-\frac{n^2}{2} + \frac{n}{\sigma^2} \sum (x_i - \mu)^2 - \frac{1}{\sigma^2} \left(\sum (x_i - \mu) \right)^2 \right] =$$

$$\left(\mu = \bar{x} \quad \sigma^2 = S_0^2 = \frac{1}{n} \sum (x_i - \bar{X}_n)^2 \right)$$

$$= \frac{1}{(S_0^2)^3} \left[-\frac{n}{2} + \frac{n^2}{S_0^2} \cdot S_0^2 - \frac{1}{S_0^2} \left(\underbrace{\sum (x_i - \bar{X}_n)}_{=0} \right)^2 \right]$$

$(\sum x_i - n\bar{x} = n\bar{x} - n\bar{x})$

$$= \frac{n^2}{2(S_0^2)^3} > 0$$

$\Rightarrow$ Ho Punto di max interno - non vado a vedere @ bordo

$$(\hat{\mu}, \hat{\sigma}^2) = \left(\bar{X}_n, \frac{1}{n} \sum (X_i - \bar{X}_n)^2\right)$$

E la buona vecchia varianza campionaria?

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$$

VOGLIO CONFRONTARLI CON L'MSE.

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \quad - \ xk \ S^2 \sim N$$

- $\mathbb{E}[S^2] = \frac{\sigma^2}{n-1} \mathbb{E}[\chi^2(n-1)] = \frac{\sigma^2}{\cancel{n-1}} \cdot \cancel{(n-1)}$ $\quad \left(\chi^2(\gamma) = \Gamma\left(\frac{\gamma}{2}, \frac{1}{2}\right), \right.$

$\Rightarrow$ È non distorto! $\quad \Gamma(\alpha, \beta):$ $\mathbb{E}[\alpha, \beta] = \frac{\alpha}{\beta}$

$$Var[S^2] = \frac{\sigma^4}{(n-1)^4} Var[\chi^2(n-1)] = \qquad \left. Var[\alpha, \beta] = \frac{\alpha}{\beta^2} \right)$$

$$= \frac{\sigma^4}{(n-1)^4} \, 2(n-1) = \frac{2\sigma^4}{n-1}$$

$\Downarrow$

$$MSE_{\sigma^2}[S^2] = Var[S^2] = \frac{2\sigma^4}{n-1}$$

- $MSE_{\sigma^2}[\hat{\sigma}^2_{MLE}] = MSE_{\sigma^2}\left[\frac{1}{n} \sum (X_i - \bar{X}_n)^2\right] =$

$$= MSE_{\sigma^2}\left[\frac{n-1}{n} S^2\right] =$$

$$= Var_{\sigma^2}\left[\frac{n-1}{n} S^2\right] + \left[\mathbb{E}_{\sigma^2}\left[\frac{n-1}{n} S^2\right] - \sigma^2\right]^2 =$$

$$= \frac{(n-1)^2}{n^2} \frac{2\sigma^4}{n-1} + \left[-\frac{\sigma^2}{n}\right]^2$$

$$= \frac{2\sigma^4(n-1)}{n^2} + \frac{\sigma^4}{n^2} = \frac{(2n-1)}{n^2}\sigma^4$$

>> Ora confronto i due MSE.

$$\frac{(2n-1)}{n^2}\sigma^4 \ ? \ \frac{2\sigma^4}{n-1} = MSE_{\sigma}[S^2]$$

$$(2n-1)(n-1) \ ? \ 2n^2$$

$$2n^2 - 3n + 1 < 2n^2 \qquad (\text{tanto } n > 1)$$

$\Downarrow$

$$\frac{2n-1}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1} \qquad \forall \sigma^2$$

$\Downarrow$

$MSE[\hat{\sigma}^2_{MLE}] < MSE[S^2]$

↑ È quello distorto.

Quindi qui è meglio uno stimatore distorto.
Perché? E perché di solito prendiamo cmq S^2?
Dato che ho funzione quadratica tratto allo stesso modo sovrastime e sottostime.
$\hat{\sigma}^2_{MLE}$ sottostima la varianza.
Dato che la varianza è un parametro di precisione non posso permettermi di sottostimarlo.
Quindi non va bene che sottostimo e sovrastimo la stessa cosa, non va bene sottostimare la varianza

LA FAMIGLIA DEGLI STIMATORI È TROPPO GRANDE x RIUSCIRE A DEFINIRE L'OTTIMALITÀ LÌ DENTRO;
NON È PROPRIO CHE BUTTO VIA GLI STIMATORI DISTORTI ...

(3) $X_1 \sim P(\lambda)$ con una variabile voglio stimare $e^{-2\lambda}$

$W(x_1) = (-1)^{x_1}$

$$E[W] = \sum_{k=0}^{+\infty} (-1)^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} = e^{-2\lambda}$$

W stimatore non distorto di $e^{-2\lambda}$
Qualunque sia λ, io devo stimare $e^{-2\lambda}$, che è >0.
Con la stessa probabilità con cui una Poisson assume un valore dispari io ~~stimo~~ stimo $e^{-2\lambda}$ con qlcs <0.
(Preferisco centrare con la stessa probabilità l'albero a dx e quello a sx e in media entrare nel box o sfiorare con la fiancata i 2 alberi?)

UMVUE = Uniform Minimum Variance Unbiased Estimator

Def T^* È UMVUE PER θ SE
$E_\theta[T^*] = \theta$ e $\forall T$ STIMATORE NON DISTORTO di $E(T)=\theta$
$Var[T^*] \leq Var[T] \quad \forall \theta$

Cerco lo stimatore non distorto che batte in termini di varianza gli altri stimatori non distorti.

Non sempre è facile confrontare le varianze!

ES: $X_1 \sim P(\lambda)$

$\lambda = E[X_i] = Var[X_i]$

$\bar{X}_n$ è non distorto per λ

S^2 è non distorto per λ.

Arriveremo a dire che è meglio $\bar{X}_n$, che è all non distorto a varianza minima.

$X_i \quad E[X_i] = \mu \quad Var[X_i] = \sigma^2$

$$E\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum E[X_i] = \frac{n\mu}{n} = \mu$$

$$E[S^2] = \sigma^2$$

$$E\left[\frac{1}{n-1}\sum (X_i - \bar{X}_n)^2\right] = \frac{1}{n-1} E\left[\sum X_i^2 + n\bar{X}_n - 2n\bar{X}_n^2\right]$$

$$= \frac{1}{n-1}\left[n E[X_i^2] - n E[\bar{X}_n^2]\right] =$$

$$= \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - n\left[\frac{\sigma^2}{n} + \mu^2\right]\right] =$$

$$= \frac{1}{n-1}\left[(n-1)\sigma^2 + n\mu^2 - n\mu^2\right] =$$

$$= \sigma^2$$

Quindi per una v.a. qualsiasi S^2 è stimatore non distorto.

Andiamo alla ricerca di criteri x stimare l'UMVUE!

DISUGUAGLIANZA di CRAMER-RAO

$X_1 \dots X_n \sim f(\vec{x}, \theta)$

$T = T(\vec{x})$ stimatore di θ

Assumiamo che

1) IL SUPPORTO DI $f(\vec{x}, \theta)$ NON DIPENDA DA θ (es: famiglia exp)

2) $0 = \frac{d}{d\theta} \underbrace{\int_{\mathbb{R}^n} f(\vec{x}, \theta) dx}_{=1} = \int_{\mathbb{R}^n} \frac{\partial f}{\partial \theta}(\vec{x}, \theta) d\vec{x}$

3) $\frac{d}{d\theta} E_\theta[T] = \frac{d}{d\theta} \int_{\mathbb{R}^n} T(\vec{x}) f(\vec{x}, \theta) dx = \int_{\mathbb{R}^n} T(\vec{x}) \frac{\partial}{\partial \theta} f(\vec{x}, \theta) dx$

4) $E[T^2] < +\infty \quad \forall \theta$ (stimatore di varianza finita.

SE

$$0 < E_\theta\left[\left(\frac{\partial}{\partial \theta} \log f(\vec{x}, \theta)\right)^2\right] < +\infty$$

$$\Rightarrow Var_\theta[T] \geq \frac{\left[\frac{d}{d\theta} E_\theta[T]\right]^2}{E_\theta\left[\left(\frac{\partial}{\partial \theta} \log f(\vec{x}, \theta)\right)^2\right]}$$

informazione di Fisher $I_n(\theta)$

LIMITE INFERIORE di CRAMER-RAO

Se lo stimatore è non distorto

$E_\theta[T] = \theta \Rightarrow \frac{d}{d\theta} E_\theta[T] = 1$

Perchè può essere utile x la stima dell'UMVUE?
Perchè mi dà un limite x la varianza.
↓
Con qst RISPONDIAMO A se $x \sim P(\lambda)$ è meglio prendere la media o la varianza campionaria?

IL Limite di Cramer-Rao dipende dalla logverosimiglianza.

$X_1 \dots X_n \sim P(\lambda)$

$$\frac{\partial}{\partial \lambda} \log f(\vec{x}, \lambda) = \frac{\partial}{\partial \lambda} \log\left[\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}\right] = \frac{\partial}{\partial \lambda} \sum \log \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} =$$

$$= \sum \frac{\partial}{\partial \lambda}(-\lambda + x_i \log \lambda) = \sum\left(-1 + \frac{x_i}{\lambda}\right)$$

$$I_n(\lambda) = E_\lambda\left[\left(\frac{\partial}{\partial \lambda} \log f(\vec{x}, \theta)\right)^2\right] = E_\lambda\left[\left(-n + \frac{\sum x_i}{\cancel{\lambda}}\right)^2\right] =$$

$$= \frac{1}{\lambda^2} E_\lambda[(-n\lambda + \underbrace{\sum x_i}_{\text{somma di Poisson} \Rightarrow \sim P(n\lambda)})^2]$$

Quindi dato che in questo caso $\mathbb{E}[\sum X_i] = n\lambda$

$$= \frac{1}{\lambda^2} Var_\lambda (\sum X_i) = \frac{1}{\lambda^2} n\lambda = \frac{n}{\lambda}$$

(in genere cerco di ricondurmi a una varianza)

lo avevo $\bar{X}_n, S^2$ non distorti

$$Var[\bar{X}_n] = \frac{\lambda}{n} = \frac{1}{I_n(\lambda)}$$

$\Rightarrow$ È L'UMVUE!

QUINDI IN QUESTO CASO LA MEDIA CAMPIONARIA RAGGIUNGE IL LIMITE DI CRAMER-RAO

dim

- DISUGUAGLIANZA DI CAUCHY-SCHWARTZ.

X, Y $\quad |Cov(X,Y)|^2 \le Var(X) \cdot Var(Y)$

- il prodotto scalare tra 2 vettori è controllato dalla norma -

dim

$a \in \mathbb{R}$

$$0 \le Var[aX+Y] = a^2 Var[X] + 2a\, Cov(X,Y) + Var[Y]$$

una parabola è sempre ≥ 0

$\Downarrow$

$\Delta \le 0$ discriminante

$$Cov^2(X,Y) - Var(X)Var(Y) \le 0$$

$\Downarrow$

$$Cov^2(X,Y) \le Var(X)Var(Y)$$

- VOGLIAMO DIMOSTRARE CHE CRAMER-RAO È UNA PARTICOLARE DISUG DI C-S.

$X = T(X)$

$Y = \frac{\partial}{\partial\theta} \log f(\vec{x}, \theta)$

$\boxed{\text{OSSERVAZIONE}}$

$$\frac{d}{d\theta} \mathbb{E}_\theta[T(X)] = \frac{d}{d\theta} \int_{\mathbb{R}^n} T(x) f(\vec{x}, \theta) d\vec{x} \overset{Hp}{=} \int_{\mathbb{R}^n} T(x) \frac{\partial}{\partial\theta} f(\vec{x}, \theta) dx$$

$$= \int_{\mathbb{R}^n} T(\vec{x}) \frac{\frac{\partial}{\partial\theta} f(\vec{x},\theta)}{f(\vec{x},\theta)} f(\vec{x},\theta) dx =$$

$$= \int_{\mathbb{R}^n} T(x) \frac{\partial}{\partial\theta} \log f(\vec{x},\theta)\, f(\vec{x},\theta)\, dx = \mathbb{E}\left[T(x) \frac{\partial}{\partial\theta} \log f(\vec{x},\theta)\right]$$

OSSERVAZIONE. 1

La applichiamo scegliendo $T(x)=1$.

$$0 = \frac{d}{d\theta}\int_{\mathbb{R}^n} f(\vec{x},\theta)dx \underset{\substack{\uparrow \\ HP}}{=} \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(\vec{x},\theta)d\vec{x} = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \log f(\vec{x},\theta)\cdot f(\vec{x},\theta)\,dx$$

$$= \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(\vec{x},\theta)\right]$$

che è la mia Y

$\Rightarrow Y$ HA VALORE medio nullo.

Allora abbiamo scoperto $\mathbb{E}[Y]=0$

$$\mathrm{Cov}(T(x),Y) \overset{*}{=} E[T(\vec{x})Y] \underset{\substack{\uparrow \\ OSS.}}{=} \frac{d}{d\theta}\mathbb{E}_\theta[T(\vec{x})]$$

forse volevo mettere Y?

e

$$\mathrm{Var}[Y] = \mathbb{E}[Y^2]$$

$\hookrightarrow$ dis C-S.

$$\mathrm{Var}[T(x)] \geq |\mathrm{Cov}(T(\vec{x}),Y)|^2 \frac{1}{\mathrm{Var}[Y]}$$

SOSTITUISCO

$$\mathrm{Var}[T(x)] \geq \frac{\left[\frac{d}{d\theta}E_\theta(T(\vec{x}))\right]^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(\vec{x},\theta)\right)^2\right]}$$

$(*)$ $\mathrm{Cov}(X,Y) = E[XY]-E[X]E[Y]$

Quando la disuguaglianza di C-S diventa uguaglianza?

QUANDO $T(\vec{x})$ È LINEARMENTE DIPENDENTE DA $\frac{\partial}{\partial\theta}\log f(\vec{x},\theta)$

Se troviamo stimatore non distorto che RAGGIUNGE il limite di CR È L'UMVUE

A CACCIA dell'UMVUE

TEOREMA di RAO - BLOCKWELL

T STIMATORE NON DISTORTO di θ e
W STATISTICA SUFFICIENTE PER θ
$\Rightarrow M = E[T|W]$ È STIMATORE NON DISTORTO per θ
e $Var(M) \leq Var_\theta(T)$

dim

1) che M È STIMATORE

$$\mathbb{E}[T|W=w] = \int T(x_1 \dots x_n) \underbrace{f_\theta(x_1 \dots x_n | W=w)}_{\text{dato che la statistica è sufficiente non dipende da } \theta} dx_1 \dots dx_n$$

Quindi ho funzione di w che non dipende da θ
$\Rightarrow$ È STIMATORE

2) $$\mathbb{E}[M] = \mathbb{E}\Big[\mathbb{E}[T|W]\Big] \underset{\text{regola della doppia media}}{=} \mathbb{E}[T] \underset{\text{xk } T \text{ è stimatore non distorto di } \theta}{=} \theta$$

Quello che non dipende da θ è la legge di $T|W$

3) $$Var[T] \underset{\text{decomp. varianza}}{=} Var\big[\mathbb{E}[T|W]\big] + \mathbb{E}\big[Var[T|W]\big]$$
$$= Var[M] + \underbrace{\mathbb{E}\big[Var[T|W]\big]}_{\geq 0}$$
$$\Downarrow$$
$$Var[M] \leq Var[T]$$

■

ESEMPIO

$X_1, X_2 \sim N(\mu, 1)$

$\bar{X}_n = \frac{1}{2}(X_1 + X_2)$ non distorto $E[\bar{X}] = \mu$ $Var[\bar{X}] = \frac{1}{2}$

Uso il teorema senza pormi il problema della sufficienza

$M = E[\bar{X} | X_1]$

← X_1 non è sufficiente per μ, xk la condizionata mi resta X_2, che dipende da μ.

$E[M] = \mu$

$Var[M] \leq Var[\bar{X}]$

$$M = E[\bar{X}|X_1] \underset{\text{linearità}}{=} \frac{1}{2}\left[E[X_1|X_1] + E[X_2|X_1]\right] =$$

$$= \frac{1}{2}[X_1 + \mu]$$

⇒ non è una statistica né uno stimatore xk dipende da μ.

TEOREMA UNICITÀ

SIA $T(\vec{x})$ UNO STIMATORE UMVUE PER θ.

⇓

$T(\vec{x})$ È L'UNICO UMVUE

dim supponiamo che ∃ T' un altro UMVUE per θ.

$T^* = \frac{1}{2}(T + T')$ è ancora uno stimatore ed è ancora non distorto per θ.

$$\left(E[T^*] = \frac{1}{2}\left[E[T] + E[T']\right] = \frac{1}{2} \cdot 2\theta = \theta\right)$$

$$Var[T^*] = \frac{1}{4}\left[Var[T] + Var[T'] + 2\,Cov(T, T')\right]$$

$$\leq \frac{1}{4} Var[T] + \frac{1}{4} Var[T'] + \frac{1}{2}\sqrt{Var[T]Var[T']}$$

vale anche senza modulo (vedi sopra)

Dato che T e T' sono UMVUE ⇒ $Var[T] = Var[T']$

Quindi $Var[T^*] \leq Var[T]$

ma dato che T è UMVUE e nella disug. di C-S deve valere =

$T' = aT + b$ xk la disuguaglianza vale quando sono lin. dipendenti

Per non avere assurdo mi serve qst, con $a, b \in \mathbb{R}$.

Però ho T e T' con stessa media e varianza.

$E[T'] = \theta = E[T]$.

$\theta = a\theta + b$) applico E a $T' = aT + b$.

⇓

$b = \theta(1 - a)$ ← 1^ VINCOLO

$$\underbrace{Cov(T, T')}_{= Var(T)} = Cov(T, aT+b) = a\, Var T$$

$$\Downarrow$$

$$Var(T) = a\, Var T \Rightarrow a = 1 \quad \leftarrow 2^\wedge \text{vincolo}$$

$$\Downarrow$$

$$b = 0$$

Quindi $T' = T$

TEOREMA di LEMANN - SCHEFFÈ UMVUE.

W STATISTICA SUFFICIENTE e COMPLETA.
U STIMATORE NON DISTORTO per θ
ALLORA
$M = \mathbb{E}[U|W]$ È L'UNICO UMVUE per θ

dim M È STIMATORE.
$\mathbb{E}[M] = \mathbb{E}[U] = \theta$
Supponiamo che M non sia l'UMVUE e quindi $\exists$ T' non distorto con $Var[T'] < Var[M]$.
Potrei prendere $M' = \mathbb{E}[T'|W]$ stimatore non distorto e per Rao-Black...

$$Var[M'] \leq Var[T'] < Var[M]$$

Se così fosse
M e M' sono funzioni di W
Quindi
$M - M'$ è funzione di W

$$\Downarrow$$

$$M - M' = g(W)$$

$\mathbb{E}[M - M'] = 0$ xk sono tutti stimatori non distorti

$$\Downarrow$$

W è completa $P(g(W) = 0) = 1$

$$\Downarrow$$

$M - M' = 0$ qc
$M = M'$ qc
Ma se è così $Var(M) < Var(M)$ ASSURDO!
Allora ho contraddetto che potesse essere uno stimatore con varianza + piccola ∎

VARIANTE

T STIMATORE NON DISTORTO PER $\tau(\theta)$.
W STATISTICA SUFFICIENTE MINIMALE COMPLETA per θ

$$\Downarrow$$

$$M = \mathbb{E}[T|W] \text{ È L'UNICO UMVUE PER } \tau(\theta)$$

ESEMPIO

$X_1 \dots X_n$ iid $\sim Bi(K, \theta)$ K NOTO.

Voglio trovare l'UMVUE per $\tau(\theta) = P_\theta(X_i = 1)$ ← probabilità di avere esattam. 1 successo

$= K\theta(1-\theta)^{K-1}$

Cosa so della binomiale?

$$f(x, \theta) = \binom{K}{x} \theta^x (1-\theta)^{K-x} \mathbb{1}_{\{0 \dots K\}}(x)$$

sta nella famiglia exp?

$$= \underbrace{\binom{K}{x} \mathbb{1}_{\{0 \dots K\}}(x)}_{\text{dipende solo da } x} \underbrace{(1-\theta)^K}_{\text{dipende solo da } \theta} \exp\Big\{ \underbrace{x \ln \frac{\theta}{1-\theta}}_{w, t} \Big\}$$

Quindi sì, sono nella famiglia exp.

$$w_1(\theta) = \log \frac{\theta}{1-\theta}$$

$$t_1(x) = X$$

Quindi osserviamo che $w_1 : [0,1] \to \mathbb{R}$

$$\Rightarrow T(x) = \sum_{i=1}^n x_i \quad \text{È SUFFICIENTE, completa, minimale.}$$

↑ Ha media $K\theta$, quindi non è uno stimatore non distorto.

Quando devo stimare una probabilità x trovare un operatore non distorto si cercano

$$y_i = \begin{cases} 1 & x_i = 1 \\ 0 & \text{altrimenti} \end{cases}$$

ossia $y_i = \mathbb{1}_{\{x_i = 1\}}$ ← le y_i sono funzioni delle x_i
$\Rightarrow$ le y_i sono tra loro indipendenti e hanno la stessa legge: y_i i id.

$$y_i \sim Be(\tau(\theta)) = Be(K\theta(1-\theta)^{K-1})$$

↑ valore 1 and $x_i = 1$

se prendo

$$\bar{Y}_n = \frac{1}{n} \sum Y_i$$

$$\mathbb{E}[\bar{Y}_n] = \tau(\theta)$$

Quindi ho costruito uno stimatore BASATO SU TUTTO IL CAMPIONE X che è NON DISTORTO per $\tau(\theta)$.

Quindi per il teorema trovo l'UMVUE condizionando:

$$M = \mathbb{E}[\bar{Y}_n \mid \sum X_i] \text{ è UMVUE per } \tau(\theta)$$

$$\mathbb{E}[\bar{Y}_n \mid \sum X_i] \underset{\uparrow}{=} \frac{1}{n} \sum E[Y_i \mid \sum X_i]$$

x la linearità della media

$$E[Y_1 \mid \sum X_i = t] = 1 \cdot P(Y=1 \mid \sum X_i = t) + 0 \cdot P(Y=0 \mid \sum X_i = t)$$

$(t = 0 \ldots, nK)$

$$= \mathbb{P}(Y_1 = 1 \mid \sum X_i = t) =$$

Y_1 vale 1 solo se $X_1 = 1$

$$= \frac{\mathbb{P}(X_1 = 1 \cap \sum X_i = t)}{\mathbb{P}(\sum X_i = t)} =$$

$\sum_{i=2}^{n} X_i \sim Bi((n-1)K, \theta)$

$\sum_{i=1}^{n} X_i \sim Bi(nK, \theta)$

$$= \frac{\mathbb{P}(X_1 = 1 \cap \sum_{i=2}^{n} X_i = t-1)}{\mathbb{P}(\sum X_i = t)}$$

$$= \frac{(K\theta(1-\theta)^{K-1}) \binom{(n-1)K}{t-1} \theta^{t-1} (1-\theta)^{(n-1)K-(t-1)}}{\binom{nK}{t} \theta^t (1-\theta)^{nK-t}}$$

θ è sparito!

se non fosse sparito ci saremmo dovuti allarmare xk la statistica è sufficiente e non dipende da θ.

$$\rightarrow \quad = \frac{K \binom{(n-1)K}{t-1}}{\binom{nK}{t}}$$

Quindi

$$M = \mathbb{E}[\bar{Y}_n \mid \sum X_i] = K \frac{\binom{(n-1)K}{\sum X_i - 1}}{\binom{nK}{\sum X_i}}$$

v.a. funzione della statistica suff. min. completa

è IL MIO UMVUE.

1. Guardo in faccia il modello, la legge congiunta, e cerco statistica sufficiente minimale e completa.
2. Ho uno stimatore per $\tau(\theta)$
   Lo posso trovare con il metodo dei momenti o con la verosimiglianza.
   Se ~~τ~~ $\tau(\theta)$ è una probabilità faccio cm qui
3. Condiziono
   Se trovo già qualcosa che è funzione di $\tau(\theta)$ sono apposto, altrimenti maneggio

INFORMAZIONE di FISHER

$$I_n(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(\vec{x},\theta)\right)^2\right]$$

$$= n\,\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right] = n I_1(\theta)$$

se $X_1 \dots X_n$ iid

dim

$$I_n(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i,\theta)\right)^2\right] =$$

$$= \mathbb{E}_\theta\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i,\theta)\right)^2\right] =$$

RICORDO CHE:

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(x_i,\theta)\right] = \int\frac{d}{d\theta}\log f(x_i,\theta)\, f(x_i,\theta) = \int\frac{d}{d\theta} f(x_i,\theta)\,dx$$

$$= \frac{d}{d\theta}\int f(x_i,\theta)\,dx = 0$$

Quindi anche la somma dei logaritmi ha media nulla.
Ma allora quello che ~~[illegible]~~ ho lì è la varianza.

la varianza della Σ è Σ Var se iid ↳

$$= \mathrm{Var}\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i,\theta)\right) =$$

$$= n\,\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right) =$$

xk la media è 0 $\rightarrow$

$$= n\,\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right] = n I_1(\theta)$$

■

Lemma (Quando lavoro con I_n devono valere le hp di Cramer-Rao)

SE IN AGGIUNTA ALLE IPOTESI PRECEDENTI SI HA CHE

$$\frac{d}{d\theta}\int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(\vec{x},\theta)\,d\vec{x} = \int_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2} f(x,\theta)\,dx$$

$$\Downarrow$$

$$I_n(\theta) = -\,\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(\vec{x},\theta)\right]$$

dim

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\vec{x},\theta)\right] = \int_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2}\log f(\vec{x},\theta)\, f(\vec{x},\theta)\,d\vec{x} =$$

$$= \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right) f(x,\theta)\,d\vec{x} =$$

$$= \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}\left(\frac{\partial f}{\partial\theta}\frac{1}{f}\right) f(x,\theta)\,d\vec{x} =$$

$$= \int_{\mathbb{R}^n} \frac{\partial^2 f}{\partial\theta^2}\frac{1}{f(x,\theta)} f(x,\theta)\,d\vec{x} - \int_{\mathbb{R}^n} \frac{1}{f^2(x,\theta)}\left(\frac{\partial f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\,dx$$

Hp

$$= \frac{d}{d\theta}\int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(x,\theta)\,d\vec{x} - \int_{\mathbb{R}^n} \frac{1}{f^2(x,\theta)}\left(\frac{\partial f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\,dx$$

$$= \frac{d}{d\theta}\Big(\underbrace{\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(x,\theta)\right]}_{=0}\Big) - \int_{\mathbb{R}^n} \frac{1}{f^2(x,\theta)}\left(\frac{\partial f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\,dx$$

$$\Downarrow$$

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)\right] = -\int_{\mathbb{R}^n}\left(\frac{\partial f(x,\theta)}{\partial\theta}\frac{1}{f(x,\theta)}\right)^2 f(x,\theta)\,dx$$

$$= -\int_{\mathbb{R}^n}\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2 f(x,\theta)\,dx$$

$$= -\,I_n(\theta)$$

$\blacksquare$

E COSA MI DICE?

Guardiamolo per il campione di ampiezza 1. (tanto poi basta che moltiplico x n)

$$I_1(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)\right]$$

la derivata 2^ mi parla della concavità.

Se ho verosimiglianze unidimensionali molto concave forniscono informazione di Fisher alta

⇓

limite di Cramer Rao basso ∩

Se invece ho verosimiglianza piatta in pratica non ho informazione di Fisher e il limite di CR è alto

LA VARIABILITÀ del MIO STIMATORE È DATA dalla CONCAVITÀ della LOG VEROSIMIGLIANZA

TEOREMA

$X_1 \dots X_n$ da $f(x,\theta)$ ∈ famiglia esponenziale

$f(x,\theta) = h(x)\,c(\theta)\exp[w(\theta)\,t(x)]$

TALE CHE ∃

$$\frac{d}{d\theta}w(\theta) \neq 0 \text{ e continua}$$

Sia $T(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} t(x_i)$ ← QUINDI SUFF MIN COMPLETA

Le condizioni per la dis di CR valgono e

$$Var(T(\vec{x})) = \frac{\left(\frac{d}{d\theta}E_\theta[T(\vec{x})]\right)^2}{I_n(\theta)}$$

↑ ossia raggiunge il limite di Cramer Rao

$\Rightarrow T(\vec{x})$ È L'UMVUE per $\mathbb{E}[t(x)]$

(è l'umvue x la sua media, quindi x la media di $t(x)$)

★ OSSERVAZIONE

$$0 = \frac{d}{d\theta}\int f(x,\theta)\,dx = \frac{d}{d\theta}\int h(x)\,c(\theta)\exp\{w(\theta)t(x)\}\,dx =$$

$$= \int h(x)\,c'(\theta)\exp\{w(\theta)t(x)\} + \int h(x)c(\theta)\exp\{w(\theta)t(x)\}w'(\theta)t(x)$$

moltiplico e divido per $c(\theta)$

$$= \frac{c'(\theta)}{c(\theta)}\underbrace{\int h(x)c(\theta)\exp\{w(\theta)t(x)\}dx}_{1} + w'(\theta)\int t(x) f(x,\theta)dx$$

$$= \frac{d}{d\theta}\log c(\theta) + w'(\theta)\,\mathbb{E}_\theta(t(x)) = 0$$

$$\Downarrow$$

$$\boxed{\mathbb{E}_\theta(t(x)) = -\frac{1}{w'(\theta)}\frac{d}{d\theta}\log c(\theta)}$$

dim

$$I_n(\theta) = n I(\theta)$$

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right] =$$

$$= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\left(\log h(x) + \log c(\theta) + w(\theta)t(x)\right)\right)^2\right] =$$

$$= \mathbb{E}_\theta\left[\left(\underbrace{\frac{d}{d\theta}\log c(\theta)}_{\text{prendo l'oss}\,\star} + w'(\theta)t(x)\right)^2\right]$$

$$I(\theta) = \mathbb{E}_\theta\left[\left(-w'(\theta)\,\mathbb{E}_\theta(t(x)) + w'(\theta)t(x)\right)^2\right]$$

$$= (w'(\theta))^2\,\mathbb{E}_\theta\left[\left(t(x) - \mathbb{E}_\theta(t(x))\right)^2\right]$$

$$= (w'(\theta))^2 \operatorname{Var} T(x)$$

Quindi a den del limite di CR ho

$$n\,(w'(\theta))^2 \operatorname{Var} T(x)$$

Ora studio il numeratore:

$$\frac{d}{d\theta}\mathbb{E}[t(x)] = \frac{d}{d\theta}\int t(x)h(x)c(\theta)\exp\{w(\theta)t(x)\}dx =$$

tanto $T = \frac{1}{n}\sum t$ Quindi la media è la stessa

$$= \int t(x)h(x)\frac{c'(\theta)}{c(\theta)}c(\theta)\exp\{w(\theta)t(x)\}dx$$

$$+$$

$$\underbrace{\int t^2(x)h(x)c(\theta)\exp\{w(\theta)t(x)\}w'(\theta)dx}_{\mathbb{E}[t^2(x)]w'(\theta)}$$

$$\frac{d}{d\theta}\mathbb{E}_\theta[t(x)] = w'(\theta)\left[\mathbb{E}[t^2(x)] - \mathbb{E}[t(x)]^2\right]$$
$$= w'(\theta)\,Var[t(x)]$$

$\frac{d}{d\theta}\log c(\theta)\mathbb{E}[t(x)]$ x l'oss. è $w'(\theta)E[t(x)]$

Quindi ho ottenuto

- $\frac{d}{d\theta}\mathbb{E}_\theta[t(x)] = w'(\theta)Var(t(x))$
- $I_n(\theta) = n\,Var(t(x))\,w'(\theta)^2$

$$\Rightarrow \text{limite CR} = \frac{w'(\theta)^2 Var^2(t(x))}{n\,Var(t(x))\,w'(\theta)^2} = \frac{Var(t(x))}{n}$$

Il mio stimatore

$$T(x) = \frac{1}{n}\sum t(x_i)$$

$$Var(T(x)) = \frac{Var(t(x))}{n} = \text{limite CR}$$

Quindi il mio stimatore è l'umvue x la media.

ESEMPIO

$X_1 \dots X_n \sim P(\lambda)$

$$f(x,\lambda) = \underbrace{\mathbb{1}_N(x)\frac{1}{x!}}_{h(x)} \underbrace{e^{-\lambda}}_{c(\lambda)} \exp(\underbrace{\log(\lambda)}_{w(\lambda)} \underbrace{x}_{t(x)})$$

il teo mi dice che

$\frac{1}{n}\sum X_i$ è l'UMVUE per $E[X] = \lambda$

Verifichiamo $\mathbb{E}[t_i] = -\frac{1}{w'(\theta)} \frac{d}{d\theta} \log c(\theta)$: $-\lambda \frac{d}{d\lambda}(\log e^{-\lambda})$
$= \lambda$

OSSERVAZIONE

Supponiamo di avere uno stimatore $T(x)$ che raggiunge il limite di CR.

- Se $|Cov(X,Y)|^2 \leq Var(X) Var(Y)$
vale l'$=$ $\Leftrightarrow$ $Y = aX + b$.

La disuguaglianza di CR è sostanzialmente la disuguaglianza di CS applicata a
$T(x)$ e $\frac{\partial}{\partial\theta}\log f(\vec{x},\theta)$

$T(\vec{x})$ raggiunge il limite di CR

$\Updownarrow$

$$T(\vec{x}) - \frac{\partial}{\partial\theta}\log f(\vec{x},\theta) = a(\theta)T(x) + b(\theta)$$

$\Updownarrow$

$f(\vec{x},\theta)$ appartiene alla famiglia esponenziale

TEOREMA

W è l'UMVUE per $\tau(\theta)$

$\Updownarrow$

$E[W] = \tau(\theta)$

e

$Cov(W,U) = 0$

$\forall$ U statistica con $\mathbb{E}_\theta(U) = 0$

Ossia l'UMVUE è uno stimatore non distorto e scorrelato da qualsiasi stimatore non distorto di zero!

dim

- $\Rightarrow$ W è UMVUE per $\tau(\theta)$ e U è $E_\theta[U]=0$

$W' = W + aU$

$E[W'] = E[W]$ $\rightarrow$ W' è ancora non distorto per $\tau(\theta)$.

$Var[W'] = Var[W] + a^2 Var[U] + 2a\, Cov(W,U)$

Se $Cov(W,U) \neq 0$ posso trovare a tc

$a^2 Var[U] + 2a\, Cov(W,U) < 0$

$\left[a \in \left[0, -2\frac{Cov(W,U)}{Var W}\right] \right]$

(o il simmetrico se la cov è >0

Allora saremmo a un assurdo: otterrei

$Var(W') < Var(W)$

Quindi perché non valga l'assurdo deve essere $Cov(W,U)=0$

- $\Leftarrow$ Se W non fosse l'UMVUE

$\Downarrow$

$\exists\, W'$ non distorto per $\tau(\theta)$ con $Var(W') < Var\, W$

Ma $W' = W + (W' - W)$

$E[W'] = E[W]$ xk sono entrambi non distorti

Di conseguenza

$E[W' - W] = 0$

Ma

$$Var[W'] = Var[W] + Var[W'-W] + \underbrace{2Cov(W, W'-W)}_{\text{è } 0 \text{ per Hp.}}$$

$\geq Var[W]$

contraddizione

$\blacksquare$

APPROCCIO BAYESIANO: ESEMPIO

IL PARAMETRO θ DIVENTA ALEATORIO CON LEGGE $\pi(\theta)$ chiamata LEGGE A PRIORI.

Dopo il campionamento ho

$$f(\vec{x}|\theta) \xrightarrow{\text{teo Bayes}} \pi(\theta|\vec{x})$$

DISTRIBUZIONE A POSTERIORI

$f(\vec{x}|\theta)$: condizionatamente al valore che assume θ ho la mia legge congiunta.

$\pi(\theta|\vec{x})$: nell'approccio Bayesiano la si usa x costruire stimatori

$$\underbrace{\pi(\theta|\vec{x})}_{\text{teo Bayes}} = \frac{f(\vec{x}|\theta)\,\pi(\theta)}{\underbrace{m(\vec{x})}_{\text{marginale}}} =$$

$$= \frac{f(\vec{x}|\theta)\,\pi(\theta)}{\underbrace{\int_\theta f(\vec{x}|\theta)\,\pi(\theta)\,d\theta}_{\text{marginalizzazione di x.}}}$$

Tipicamente
$E[\pi(\theta|\vec{x})]$ stima θ

~~Guardiamo un caso fortunato~~

modello Beta - Binomiale

$X_1 \dots X_n | p \sim Be(p)$

$Y = \sum_{i=1}^{n} X_i \sim Bi(n,p)$ sempre condizionatamente a p

$\pi(p) \sim Beta(\alpha, \beta)$ ← p lo sto prendendo a caso da una Beta, che ha supporto tra 0 e 1.

Quindi
$f(y|p) \sim Bi(n,p)$

congiunta: $$f(y,p) = \underbrace{\left[\binom{n}{y} p^y (1-p)^{n-y}\right]}_{f(y|p)} \underbrace{\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\right]}_{\pi(p)}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha+y-1}(1-p)^{n-y+\beta-1}$$

ora devo trovare la marginale.

$$f(y) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \underbrace{p^{\alpha+y-1}(1-p)^{n-y+\beta-1}}_{\text{ha la forma della densità di una Beta.}} dp$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)} \underbrace{\int_0^1 cost \ldots dp}_{1}$$

Posso calcolare la distribuzione a posteriori di p:

$$f(p|y) = \frac{f(y,p)}{f(y)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+y)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}$$

$$\sim Beta(\alpha+y, n-y+\beta)$$

È UN CASO FAVOREVOLE, parto da una beta, il coniugato è Binomiale, la distribuzione a posteriori è ancora una Beta con nuovi parametri (parametri AGGIORNATI)

UNO STIMATORE di p può ESSERE LA MEDIA della LEGGE di p.

$\mathbb{E}_{priori}[p] = \frac{\alpha}{\alpha+\beta}$ $\quad \alpha$ regola la massa in 1

$\mathbb{E}[p|y] = \frac{\alpha+y}{n+\alpha+\beta}$ $\leftarrow$ media della legge a posteriori
- stimatore bayesiano

$$\mathbb{E}[p|y] = \frac{\alpha+y}{n+\alpha+\beta} = \frac{n}{\alpha+\beta+n} \underbrace{\left(\frac{y}{n}\right)}_{\substack{\text{stima frequentista,} \\ \bar{x}_n}} + \frac{\alpha+\beta}{\alpha+\beta+n} \underbrace{\frac{\alpha}{\alpha+\beta}}_{\text{stima a priori}}$$

combinazione di 2 quantità:
di fatto sto pesando un'info fornita dai dati e quella che sapevo a priori.
Più n cresce e più ho importanza quello che mi dicono i dati.

TEST d'IPOTESI

DEF $X_1 \dots X_n$ iid $f(\vec{x}, \theta)$

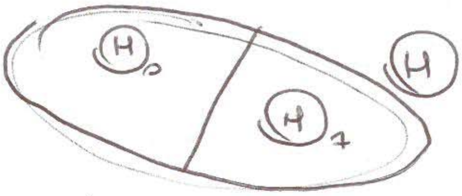
È UN PROCESSO DECISIONALE per DECIDERE A FAVORE O CONTRO L'IPOTESI STATISTICA: affermazione sui parametri θ del campione

In un problema di verifica delle ipotesi chiamiamo

H_0 IPOTESI NULLA
H_1 IPOTESI ALTERNATIVA.

Dovrò scegliere tra H_0 e H_1, che in generale ripartiscono lo spazio dei parametri

$$\begin{cases} H_0 & \theta \in \Theta_0 \\ H_1 & \theta \in \Theta_1 \end{cases}$$



$$\Theta_1 = \Theta_0^C$$

cosa metto in H_0 e cosa in H_1?

In H_0 metto quello che è vero fino a prova contraria,

in H_1 metto quello che voglio avere le prove x poterlo affermare.

In H_0 metto la cosa conservativa

Il test d'ipotesi è un tribunale

Def UN TEST d'IPOTESI È UNA REGOLA CHE SPECIFICA
- a) Per QUALI VALORI del CAMPIONE SI ACCETTA H_0 (H_0 vera)
- b) Per QUALI VALORI del CAMPIONE SI RIFIUTA H_0 (H_0 FALSA)
  ↳ REGIONE CRITICA: Quella x cui RIFIUTO
- → REGIONE di ACCETTAZIONE

Si parla in generale di "accetto H_0" o "rifiuto H_0", xk magari ho anche H_2.

RC, RA SONO COSTRUITE TRAMITE UNA STATISTICA $T(\vec{x})$ CHIAMATA STATISTICA TEST

METODI PER COSTRUIRE I TEST

★ TEST del RAPPORTO di VEROSIMIGLIANZA (LIKELIHOOD RATIO TEST) (LRT)

Def LA STATISTICA SU CUI SI BASA IL LRT per

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases} \quad \text{è}$$

campione di massima verosimiglianza esplorato solo in un range

$$\lambda(\vec{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta, \vec{x})}{\sup_{\theta \in \Theta} L(\theta, \vec{x})} = \frac{L(\hat{\theta}^0_{MLE}, \vec{x})}{L(\hat{\theta}_{MLE}, \vec{x})}$$

xk il sup viene raggiunto in $\hat{\theta}_{MLE}$

REGIONE CRITICA:

$$\{\lambda(\vec{x}) \leq c\} \qquad 0 \leq c \leq 1$$

($\lambda(\vec{x})$ è sicuramente < 1)

Se la quantità λ è piccola significa che la massima verosimiglianza cresce molto di più se non sto in Θ_0, quindi è + probabile che θ stia in Θ_0^c

ESEMPIO

$X_1 \dots X_n \sim N(\mu, 1)$

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

θ è il parametro incognito, θ_0 è noto.

$$\lambda(\vec{x}) = \frac{L(\theta_0, \vec{x})}{L(\hat{\theta}_{MLE}, \vec{x})}$$

⇐ Non ha senso fare il Sup se ho solo 1 punto!

- La regione critica dovrà dipendere dal campione e da θ_0, NON DA θ, non posso prendere una decisione basata su θ che non conosco.

$$= \frac{\frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum (x_i - \theta_0)^2\right\}}{\frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum (x_i - \bar{x}_n)^2\right\}} =$$

$$= \exp\left\{ \frac{1}{2} \left(- \sum (x_i - \theta_0)^2 + \sum (x_i - x_n)^2 \right) \right\}$$

OSS

$$\sum (x_i \underset{\pm \bar{x}_n}{-} \theta_0)^2 = \sum (x_i - \bar{x}_n)^2 + \sum (\bar{x}_n - \theta_0)^2 + 2 \underbrace{\sum (x_i - \bar{x}_n)(\bar{x}_n - \theta_0)}_{=0}$$

$$= \exp\left\{ \frac{1}{2} \left(- \sum (x_i - x_n)^2 - n (\bar{x}_n - \theta_0)^2 + \sum (x - \bar{x}_n)^2 \right) \right\}$$

$$\lambda(\vec{x}) = \exp\left\{ -\frac{n}{2} (\bar{x}_n - \theta_0)^2 \right\}$$

ok è una statistica xk dipende da $\bar{x}_n$ e θ.

COSTRUIAMO LA REGIONE CRITICA

$0 \leq c \leq 1$

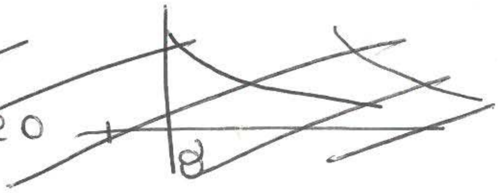
$$RC = \{\lambda(x) \leq c\} = \left\{ \exp\left\{ -\frac{n}{2} (\bar{x}_n - \theta_0)^2 \right\} \leq c \right\}$$

$$= \left\{ |\bar{x}_n - \theta_0|^2 \geq \sqrt{\frac{-2 \log c}{n}} \right\} = \{ |\bar{x}_n - \theta_0| \geq k \}$$

$k \in \mathbb{R}$

più la soglia c è vicino a 1 e più rifiuto sempre.

~~È RAGIONEVOLE LA MIA RC?~~
~~La densità è qualcosa~~
~~Devo decidere se θ è + grande o + piccolo di θ_0~~



~~Finché vedo~~

Voglio capire come è fatta la Regione Critica.

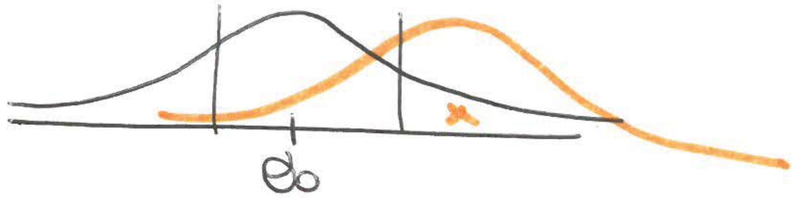
È ragionevole la forma della RC?

Rifiuto H_0 quando la media campionaria è distante dalla mia θ_0.

Immagino di osservare una cosa probabile rispetto al fatto di avere osservato un miracolo.

Quando campiono $X_1 \dots X_n$ da $N(\theta, 1)$

<u>SE È VERA H_0</u> $\bar{X}_n$ sotto H_0



mi dice di rifiutare quando sono in modulo lontana da H_0.

Perché se sono qui × è molto + probabile che il campione sia generato da qst

Quella arancio mi genera il campione pallino arancio con probabilità molto maggiore.

Quando vorremo dare un valore a k guarderemo la deviazione standard.

Ho ottenuto una regione bilatera, $\times k$

H_0 era $\theta = \theta_0$

H_1 $\neq$

La forma della rc dipende molto da cosa metto in H_0 e H_1.

ESEMPIO

$X_1 \dots X_n$ iid

$f(x,\theta) = \exp(-(x-\theta)) \mathbb{1}_{[\theta,+\infty)}(x)$; $\theta \in \mathbb{R}$

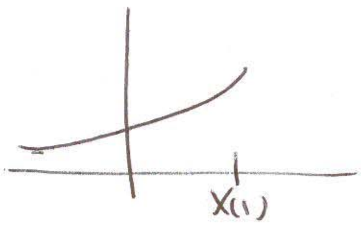
esponenziale traslato

Voglio costruire LRT

Come è fatta la verosimiglianza?

$$L(\theta,\vec{x}) = \exp\left(-\sum x_i + n\theta\right) \prod_{i=1}^{n} \mathbb{1}_{[\theta,+\infty)}(x_i)$$

$$= \exp\left(-\sum x_i + n\theta\right) \mathbb{1}_{(-\infty, x_{(1)})}(\theta)$$

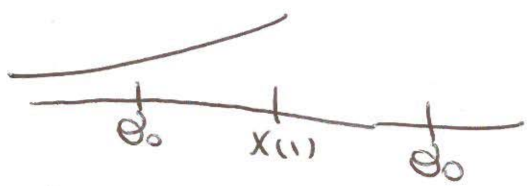


$\hat{\theta}_{MLE} = X_{(1)}$

$$\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

$$\lambda(\vec{x}) = \frac{\sup_{\theta \leq \theta_0} L(\theta,\vec{x})}{\exp\left\{-\sum x_i + n x_{(1)}\right\}} =$$

Se il minimo delle mie osservazioni cade prima di θ_0 il Sup ristretto è ancora il minimo, altrimenti è θ_0

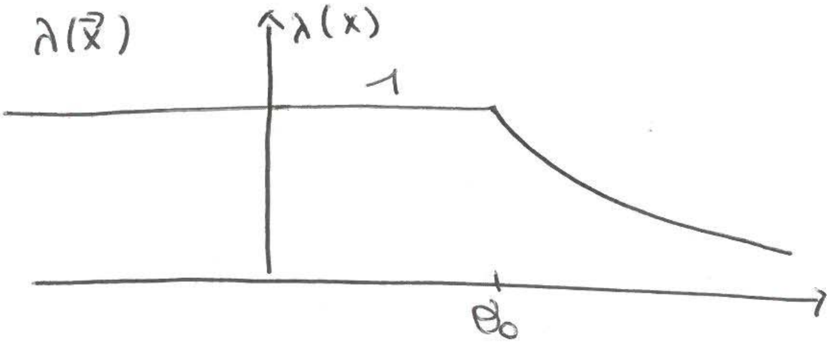


$$= \begin{cases} \dfrac{L(x_{(1)},\cancel{x})}{L(x_{(1)},x)} & x_{(1)} \leq \theta_0 \\[2ex] \dfrac{L(\theta_0,x)}{L(x_{(1)},x)} & x_{(1)} > \theta_0 \end{cases}$$

Quindi

$$\lambda(\vec{x}) = \begin{cases} 1 & x_{(1)} \leq \theta_0 \\ \exp\{-n(x_{(1)}-\theta_0)\} & x_{(1)} > \theta_0 \end{cases}$$

La statistica test qui è il minimo:



Se il minimo è $\leq \theta_0$ accetto sempre

Rifiuto se

$$RC = \{ \lambda(x) \leq c \} = \{ e^{-n(x_{(1)} - \theta_0)} \leq c \}$$

$$\Updownarrow$$

È una RC?
una partizione che posso campionare se ho i dati e c? sì. $\rightarrow \{ x_{(1)} \geq \theta_0 - \log c \cdot \frac{1}{n} \}$

È ragionevole come RC?
La densità è

Devo decidere se θ è + grande o + piccolo di θ_0

se osservo minimo qui accetto, xk se rifiutassi potrei avere densità tipo

Finché vedo minimi sotto θ_0 accetto H_0, se trovo minimi $> \theta_0$ di una certa soglia mi immagino che i dati siano dati da qualcosa che ha θ più grande, xk genera quei dati con alto probabilità.

Quindi la forma della mia RC è ragionevole.

TEOREMA

Sia $T(\vec{X})$ UNA STATISTICA SOFFICIENTE per θ e $\lambda(\vec{x})$ e $\lambda^*(t)$ le statistiche LRT su $\vec{X}$ e $\vec{T}$ RISPETTIVAMENTE.

$$\Downarrow$$

$$\lambda^*(T(\vec{x})) = \lambda(\vec{x})$$

dim

$$\lambda(\vec{x}) = \left\{ \frac{\sup_{\theta \in \Theta_0} L(\theta, \vec{x})}{\sup_{\theta \in \Theta} L(\theta, x)} \right\} = \frac{\sup_{\theta \in \Theta_0} f(\vec{x}, \theta)}{\sup_{\theta \in \Theta} f(\vec{x}, \theta)} =$$

essendo T suff. vale teo fatt. ↓

$$= \frac{\sup_{\theta \in \Theta_0} g(T(\vec{x}), \theta)\, h(\vec{x})}{\sup_{\theta \in \Theta} g(T(\vec{x}), \theta)\, h(\vec{x})}$$

il sup dipende da θ quindi qst lo tiro fuori dal sup e semplifico

$$= \frac{\sup_{\theta \in \Theta_0} g(T(\vec{x}), \theta)}{\sup_{\theta \in \Theta} g(T(\vec{x}), \theta)} = \lambda^*(T(\vec{x}))$$

Se sono uguali le due statistiche test saranno uguali le due regioni critiche.

■

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

ABBIAMO UNA STATISTICA TEST e QUINDI ABBIAMO UNA RC.

$RC = \{ \vec{x} \text{ tc rifiuto } H_0 \}$

		DECISIONE	
		ACCETTO H_0	RIFIUTO H_0
VERO	H_0	OK	Errore 1^ tipo
	H_1	Errore 2^ tipo	OK

Non è possibile avere sia errore 1^ tipo che 2^ tipo a 0.

ES: $X_1 \dots X_n \sim Be(p)$

Voglio fare un test del tipo
$H_0: p = p_0 = \frac{1}{2}$
$H_1: p \neq p_0$

La moneta è equilibrata? Bo!

 se ho una scatola che contiene tante palline bianche quante nere

Potrei, aprendo la scatola, conoscere il vero p.
Ma in generale, come nella moneta, il parametro è davvero incognito!

$P_\theta(\vec{X} \in RC)$ = PROBABILITÀ di commettere un errore di PRIMO TIPO, SE $\theta \in \Theta_0$.

$P_\theta(\vec{X} \in RC^c)$ = PROBABILITÀ di commettere un errore di SECONDO TIPO SE $\theta \in \Theta_0^c$

$$P_\theta(\vec{X} \in RC^c) = 1 - P_\theta(\vec{X} \in RC)$$

Quindi

$$P_\theta(\vec{X} \in RC) = \begin{cases} \text{probabilità di errore di } 1^\wedge \text{ tipo } \theta \in \Theta_0 \\ 1 - \text{probabilità di errore di } 2^\wedge \text{ tipo } \theta \in \Theta_0^c \end{cases}$$

<u>Def</u>

SI DICE FUNZIONE POTENZA di UN TEST d'IPOTESI
$H_0: \theta \in \Theta_0$ VS $H_1: \theta \in \Theta_0^c$,
di Regione critica RC

$$\beta(\theta) = P_\theta(\vec{X} \in RC)$$

↑ mappa lo spazio dei parametri in $[0,1]$

A seconda di dove sia θ mi dice l'errore di $1^\wedge$ tipo o $1 -$ errore $2^\wedge$ tipo

Mi piacerebbe valesse
0 in $\theta \in \Theta_0$
1 in $\theta \in \Theta_0^c$

But it's pretty difficult to find it!

Per confrontare i vari test confronteremo β.

(54) Idealmente avrei ottimalità se avessi prob di commettere 1^ tipo = 0
2^ tipo = 0

OSSERVAZIONE

$X \sim Bi(5, \vartheta)$ (ho un'unica osservazione)

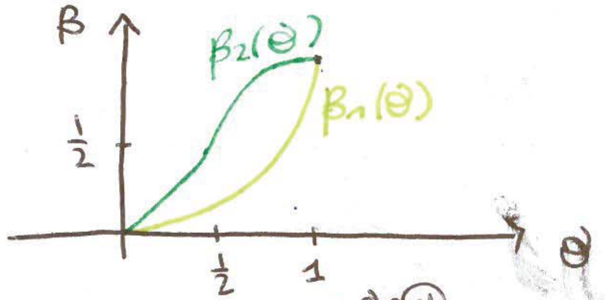
$H_0: \vartheta \leq \frac{1}{2}$ vs $H_1: \vartheta > \frac{1}{2}$

test 1
voglio capire se la p è > o < rispetto alla soglia equa.
Rifiuto H_0 se osservo solo successi: $X = 5$
test 2
Rifiuto H_0 se $X \geq 3$

Come sono questi due test in termini di potenza?

$\beta_1(\vartheta) = P_\vartheta(X=5) = \vartheta^5$

$\beta_2(\vartheta) = P_\vartheta(X \geq 3) = \binom{5}{3}\vartheta^3(1-\vartheta)^2 + 5\vartheta^4(1-\vartheta) + \vartheta^5 > \beta_1(\vartheta) \quad \forall \vartheta$



test 1: $P(\text{errore } 1^\wedge \text{ tipo}) \leq \underbrace{\beta_1\left(\tfrac{1}{2}\right)}_{\text{con } \vartheta \in H_0} = \left(\frac{1}{2}\right)^5 = 0{,}03125$
$P(\text{errore } 2^\wedge \text{ tipo})$ è alta! potenza $\geq 0{,}8$ quindi $\vartheta \geq 0{,}96$
$\uparrow 1-P = \vartheta^5$

test 2: $P(\text{errore } 1^\wedge \text{ tipo}) \leq \beta_2\left(\frac{1}{2}\right) = (10+5+1)\frac{1}{2^5} = \frac{1}{2}$ (50% → troppo alto!)

ESERCIZIO

$X_1 \dots X_n \sim N(\mu, \sigma^2)$ $H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$

LRT provo a fare x es... arrivo a $\left\{\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > c\right\}$ con c che è una costante

$\downarrow$

$\left\{\bar{X}_n > \mu_0 + \frac{\sigma}{\sqrt{n}}c\right\}$

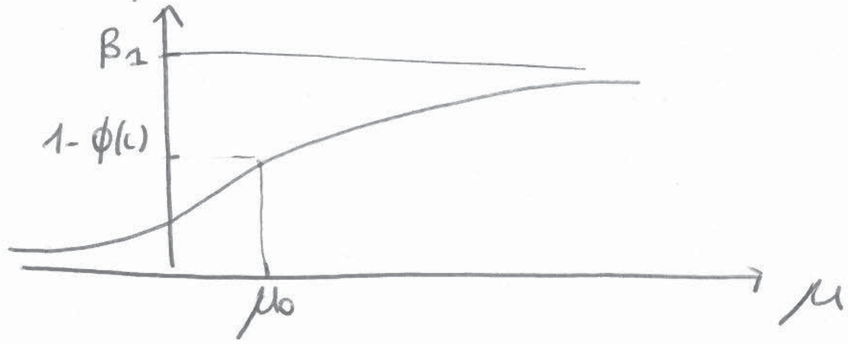
Calcolo ora $\beta(\mu) = P_\mu\left(\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > c\right)$

$= P_\mu\left(\frac{\bar{X}_n - \mu_0 + \mu - \mu}{\sigma/\sqrt{n}} > c\right) \stackrel{\text{così ho standardizzato}}{=}$

$= P_\mu\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$

Quindi $= P_\mu\left(Z > c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) = 1 - \phi\left(c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$

Graficchiamo la funzione potenza



$\lim_{\mu \to +\infty} \beta(\mu) = 1$

$\lim_{\mu \to -\infty} \beta(\mu) = 0$

funzione monotona crescente in μ.

- voglio che $P_\mu(RC)$ per $\mu \leq \mu_0$ sia ≤ 0.1

$\beta(\mu) \leq \beta(\mu_0) \quad \forall \mu \leq \mu_0$ (dato di fatto)

Quindi basta scegliere che

$1 - \phi(c) = 0.1$ cioè $P_\mu(RC) = 0.1$

$\Rightarrow \phi(c) = 0.9 \rightarrow c = z_{0.9} = 1{,}28$

NB Posso modificare n (aumentandolo) in modo che la funzione potenza cresca

- voglio che $\beta(\mu) \geq 0.8 \quad \forall \mu \geq \mu_0 + \sigma$ (probabilità di commettere errore di 2^ tipo: 20%)

Quanto vale β in $\mu_0 + \sigma$?

$\beta(\mu_0 + \sigma) = 1 - \phi(1.28 - \sqrt{n}) = 0.8$

$1.28 - \sqrt{n} = z_{0.2} = -0.84$

$\downarrow$

$n = 4.49 \Rightarrow n \geq 5$

da $\mu_0 + c \quad \beta(\mu) \geq 0.8$

cioè se faccio almeno 5 osservazioni posso avere probabilità di commettere errore 2^ tipo ~ 20%.

56 **DEF** $\forall\, 0 \le \alpha \le 1$, un test con funzione potenza $\beta(\theta)$ è detto di DIMENSIONE α SE

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$$

DEF $\forall\, 0 \le \alpha \le 1$, un test con funzione potenza $\beta(\theta)$ è DETTO di LIVELLO α SE

$$\sup_{\theta \in \Theta_0} \beta(\theta) \le \alpha$$

OSS

Laddove la $\beta(\theta)$ è continua non ho problemi a determinare livello/dimensione (collassano)

In certi casi però non si riuscirà ad ottenere la dimensione α, comunque quasi sempre noi parleremo di livello α.

α lo cercheremo piccolo, tendenzialmente.

Cosa succede se voglio fissare il livello/dimensione di un test?

ESEMPIO 1 (di PRIMA)

$X_1 \dots X_n \sim N(\mu, \sigma^2)$

$H_0 : \mu \le \mu_0$ vs $H_1 : \mu > \mu_0$

LRT: $RC \left\{ \bar{X}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} c \right\}$

$$\beta(\mu) = 1 - \phi\left(c + \frac{\mu_0 - \mu}{\sigma + \sqrt{n}}\right)$$

ora voglio calcolare α:

$$\sup_{\mu \le \mu_0} \beta(\mu) = \beta(\mu_0) = 1 - \phi(c)$$

per come ha fatto il disegno si vede!

Quindi dimensione: $\alpha = 1 - \phi(c)$

(voglio il livello del quantile) $\rightarrow \phi(c) = 1 - \alpha$

$c = z_{1-\alpha}$ (rifiuto se ... < quantile di ordine $1-\alpha$)

ESEMPIO 2 (già fatto)

$X_1 \dots X_n$ iid $\quad f(x|\theta) = e^{-(x-\theta)} \mathbb{1}_{[\theta,+\infty)}(x)$

(esp. traslato) $\quad X_i \sim \underbrace{T_i}_{\mathcal{E}(1)} + \theta$

$H_0: \theta \leq \theta_0$

$H_1: \theta > \theta_0$

LRT: $RC = \left\{ X_{(1)} \geq \theta_0 - \frac{\log c}{n} \right\}$ (+ c è piccolo + rifiuto quasi mai)

NB

- con LRT (o altre tecniche) ho la forma della regione critica
- imponendo il livello si vincola la soglia (c) della RC

$\Downarrow$

$$\sup_{\theta \leq \theta_0} \beta(\theta) = \alpha$$

↑ vediamo se riusciamo

(α è tipicamente 1% o 5% ma in generale scelgo io)

OSS

$P(X_{(1)} \geq t)$

in generale $X \sim \mathcal{E}(\lambda)$

$$\left[\prod_{i=1}^{n} P(X_i \geq t) = (1 - F_{X_i}(t))^n = (1 - 1 + e^{-\lambda t})^n = e^{-\lambda t n} \right]$$

$X_{(1)} \sim T_{(1)} + \theta \sim \mathcal{E}(n\lambda) + \theta \sim \mathcal{E}(n) + \theta$

$P(X_{(1)} \geq t) = e^{-n(t-\theta)}$

Voglio calcolare

1) $\beta(\theta) = P_\theta\left(X_{(1)} \geq \theta - \frac{\log c}{n}\right)$ → sostituisco a quella sopra $t = \theta_0 - \frac{\log c}{n}$

e imporre che

2) $\sup_{\theta \leq \theta_0} \beta(\theta) = \alpha$

$\beta(\theta)$ cresce in θ (xk ho $e^{\overset{\text{positivo}}{n\theta} - nt}$)

Quindi

$$\sup_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = e^{-n\left[\theta_0 - \frac{\log c}{n} - \theta_0\right]} = \alpha$$

↑ per il fatto che la funzione $\beta(\theta)$ è crescente in θ

↓

$e^{\log c} = \alpha \rightarrow \underline{c = \alpha}$

Quindi $\beta(\theta)$ è di dimensione α

e $RC = \left\{ X_{(1)} \geq \theta_0 - \frac{\log \alpha}{n} \right\}$

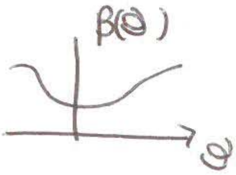
$\beta(\theta) = e^{-n\left(\theta_0 - \theta - \frac{\log \alpha}{n}\right)}$

Agendo su n posso modificare $\beta(\theta)$ - vedi prima -

(3) DEF un test con funzione di potenza $\beta(\theta)$ è NON DISTORTO SE $\beta(\theta') \geq \beta(\theta'')$

$\forall\, \theta' \in \Theta_0^c$ e $\theta'' \in \Theta_0$

(provo LRT $\mu = \mu_0$
$\mu \neq \mu_0$



Verrebbe da dire ~~se~~ è monotona è sempre vero, ma 2 esempio non monotona

→ cioè non è che le funzioni potenze sono sempre monotone!)

ESERCIZIO 1

$X_1 \dots X_5 \sim Be(p) \qquad 0 \leq p \leq 1$

$$\begin{cases} H_0: p = 1/2 \\ H_1: p \neq 1/2 \end{cases} \qquad RC = \left\{ \left|\bar{x} - \tfrac{1}{2}\right| > c \right\}$$

OSS L'IPOTESI È UN PONTO QUINDI SI DICE IPOTESI SEMPLICE

- trovare c tc quel test abbia dimensione $\alpha = 0.1$
- " livello $\alpha = 0.1$

Dato che Ho è 1 punto non ho il sup, ma la prob in quel ponto.

$$P_{\frac{1}{2}}\left(\left|\bar{X}_5 - \tfrac{1}{2}\right| > c\right) = P_{\frac{1}{2}}\left(\left|\sum X_i - \tfrac{5}{2}\right| > 5c\right) =$$

$$= 1 - P_{\frac{1}{2}}\left(-5c \leq \sum X_i \leq 5c\right)$$

NB $\sum X_i \sim Bi(5, \tfrac{1}{2}) \qquad E[X_i] = \frac{5}{2}$

Valori assunti

$\sum X_i - \frac{5}{2}$	P_k
$-5/2$	0,03125
$-3/2$	0,15625
$-1/2$	0,31250
$1/2$	0,31250
$3/2$	0,15625
$5/2$	0,03125

Quindi

$$P_{\frac{1}{2}}\left(\left|\bar{X}_5 - \tfrac{1}{2}\right| > c\right) = \begin{cases} 1 & 5c < 1/2 \\ 1 - 2 \cdot 0,03125 & \frac{1}{2} < 5c < 3/2 \\ 2 \cdot 0,03125 & 3/2 < 5c < 5/2 \\ 0 & 5c \geq 5/2 \end{cases}$$

Perciò le dimensioni possibili sono: 1; 0,375; 0,0625; 0

Quindi non c'è c per cui ho dimensione α, ma posso trovare c tc la $\beta(\theta)$ ha livello α, basta scegliere

$5c > 3/2$, così ho $0 + 0{,}625$ che è $< 10\%$.

Il fatto che $\nexists$ dimensione è xk è discreto

ESERCIZIO 2

$X_1 \sim P(\lambda)$

$$\begin{cases} H_0 : \lambda = 1 \\ H_1 : \lambda = 2 \end{cases} \qquad \beta(1) = P_{\lambda=1}(X > 3) = 0.019$$

RC $\{X > 3\}$ voglio un test di dimensione $\alpha = 5\%$.

Introduco allora dei TEST RANDOMIZZATI

$$\phi(x) = \begin{cases} 1 & x \geq 4 \\ \gamma & x = 3 \\ 0 & x \leq 2 \end{cases}$$

(funzione di γ con la quale scelgo)

Allora decido il valore di γ tc

$$1 \cdot P(X \geq 4) + \gamma P(X = 3) = 0.05 \qquad \gamma = 0{,}506$$

Quindi posso costruire test che non sono tribonali con 2 possibilità, ma aggiungo un'ipotesi (qui γ) al fine di ottenere una funzione potenza di una certa dimensione α che voglio io.

(60) Come massimizzare la potenza?

Def Sia $\mathcal{C}$ una classe di test per verificare

$$H_0: \vartheta \in \Theta_0 \quad vs \quad H_1: \vartheta \in \Theta_0^c$$

un test $\in \mathcal{C}$ con funzione potenza $\beta(\vartheta)$ è detto UMP (uniformly most powerful uniformemente + potente) nella classe $\mathcal{C}$ se

$$\beta(\vartheta) \geq \beta'(\vartheta) \quad \forall \vartheta \in \Theta_0^c$$

$\forall \beta'(\vartheta)$ funzione di un test $\in \mathcal{C}$

Cerchiamo il test che performa meglio, batte gli altri in termini di potenza quando $\vartheta \in \Theta_0^c$

Prima di tutto fissiamo il livello, poi all'interno dei test di livello α cerchiamo l'UMP

$\mathcal{C}$ = test di livello α.

Per trovare l'UMP ci servono dei risultati.

LEMMA di NEYMANN-PEARSON

CNS x caratterizzare l'UMP, tranne che x A a misura nulla

$$H_0: \vartheta = \vartheta_0 \quad vs \quad H_1: \vartheta = \vartheta_1 \quad \text{(IPOTESI SEMPLICI)}$$

$f(\vec{x}, \vartheta_0)$ legge di $\vec{x}$ sotto H_0

$f(\vec{x}, \vartheta_1)$ legge di $\vec{x}$ sotto H_1

Se usiamo una RC tc

(1) $\vec{x} \in RC$ se $f(\vec{x}, \vartheta_1) > K f(\vec{x}, \vartheta_0)$

$\vec{x} \in RC^c$ se $f(\vec{x}, \vartheta_1) < K f(\vec{x}, \vartheta_0)$

- ϑ_1 rende + probabile l'osservazione rispetto a ϑ_0

(2) $\alpha = P_{\vartheta_0}(\vec{x} \in Rc)$ $\quad K \geq 0$

$\Downarrow$

a) Qualunque test che soddisfi (1) e (2) è UMP di livello α

b) Se $\exists$ un test che soddisfa (1) e (2) con $K > 0$ allora ogni test UMP di livello α ha anche dimensione α (vale la (2)) e soddisfa la (1) tranne che per un insieme di misura nulla, cioè

$$P_{\vartheta_0}(\vec{x} \in A) = P_{\vartheta_1}(\vec{x} \in A) = 0$$

<u>dim</u> Se vale la (2), dato che le ipotesi sono semplici, il test ha dimensione α (anche livello α)

Chiamiamo $\phi(\vec{x})$ la funzione test:

$$\underbrace{\phi(\vec{x})}_{\mathbb{1}_{(RC)}} = \begin{cases} 1 & \vec{x} \in RC \\ 0 & \vec{x} \in RC^c \end{cases}$$

• a) Sia $\phi(\vec{x})$ la funzione test di un test che soddisfa (1) e (2)
Sia $\phi'(\vec{x})$ la funzione test di un qualunque altro test di livello α
Siano $\beta(\theta)$ e $\beta'(\theta)$ le relative funzioni potenza.

$$(\phi(\vec{x}) - \phi'(\vec{x}))\underbrace{[f(\vec{x},\theta_1) - K f(\vec{x},\theta_0)]}_{\star} \geq 0 \quad \forall \vec{x} \in \mathbb{R}^n$$

- se $\vec{x} \in RC$:

$$\underbrace{[1 - \phi'(\vec{x})]}_{>=0} \star \geq 0$$

1: xk $\phi(\vec{x})$ è $\mathbb{1}$ RC; $\phi'(\vec{x})$: non lo conosco

ma $\star$ nella RC è >0
xk è costruita proprio qnd $f(x,\theta_1) > K f(x,\theta_0)$

- se $\vec{x} \in RC^c$

$$\underbrace{[0 - \phi'(\vec{x})]}_{\leq 0} \underbrace{\star}_{<0} \geq 0 \quad \text{yo}$$

Quindi

$$0 \leq \int_{\mathbb{R}^n} (\phi(x) - \phi'(x))[f(\vec{x},\theta_1) - K f(\vec{x},\theta_0)]\,d\vec{x}$$

$$0 \leq \beta(\theta_1) - \beta'(\theta_1) - K[\beta(\theta_0) - \beta'(\theta_0)]$$

$\beta(\theta_1)$ è $\int \phi(x) f(\vec{x},\theta_1) = \int_{RC} f(\vec{x},\theta_1) = P_{\theta_1}(\vec{x} \in RC) = \beta(\theta_1)$

ma dato che ϕ candidato UMP SODDISFA LA (2),

$$\beta(\theta_0) = \alpha \Rightarrow \beta(\theta_0) - \beta'(\theta_0) = \alpha - \underbrace{\beta'(\theta_0)}_{\leq \alpha} \geq 0$$

$\leq \alpha$ xk β' è di livello α

(*) $$0 \leq \beta(\theta_1) - \beta'(\theta_1) - K\underbrace{[\beta(\theta_0) - \beta'(\theta_0)]}_{\geq 0} \leq \beta(\theta_1) - \beta'(\theta_1)$$

$$\Rightarrow \beta(\theta_1) \geq \beta'(\theta_1)$$

• b) ϕ' funzione test di un qualunque test UMP di livello α.
Per il punto a) anche ϕ (del test che soddisfa (1) e (2)) è UMP di livello α.
Allora $\beta(\theta_1) = \beta'(\theta_1)$ (altrimenti non sarebbero entrambi UMP.

$\Downarrow$ da (*)

$$\alpha - \beta'(\theta_0) \leq 0 \Rightarrow \beta'(\theta_0) \geq \alpha$$

Però ϕ' è UMP di livello $\alpha \Rightarrow \beta'(\theta_0) \leq \alpha \Rightarrow \beta'(\theta_0) = \alpha$

Ma allora (*) $= 0$

$\Rightarrow \int [\phi(x) - \phi'(x)][f(\vec{x},\theta_1) - K f(\vec{x},\theta_0)] = 0$ - l'unico modo x avere 0 è $\phi' = \phi$ tranne che su insieme di misura nulla A: $P_{\theta_1}(\vec{x} \in A) = P_{\theta_0}(\vec{x} \in A) = 0$ ∎

COROLLARIO

TEST H_0 $\theta=\theta_0$ VS H_1 $\theta=\theta_1$
$T(\vec{x})$ STAT SUFF per θ.
$g(T,\theta)$ la legge di T.
Allora ogni test basato su T con regione di rifiuto S è UMP di livello α SE

$$t \in S \text{ se } g(t,\theta_1) > K g(t,\theta_0)$$
$$t \in S^c \text{ se } g(t,\theta_1) < K g(t,\theta_0)$$
e
$$\alpha = P_{\theta_0}(T \in S)$$

lemma di N-S costruito sulla densità di una stat. suff.

dim

$S: RC = \{\vec{x} : T(\vec{x}) \in S\}$
vale il criterio di fattorizzazione: $f(\vec{x},\theta_i) = h(\vec{x})\, g(T(\vec{x}),\theta_i)$

$$\vec{x} \in RC \text{ se } f(\vec{x},\theta_1) = g(T(x),\theta_1)h(x) \geq K g(T(x),\theta_0)h(x) = K f(x,\theta_0)$$
$$\vec{x} \in RC^c \text{ se } f(\vec{x},\theta_1) = g(T(x),\theta_1)h(x) \leq K g(T(x),\theta_0)h(x) = K f(x,\theta_0)$$

■

ESEMPIO

① $X \sim Bi(2,\theta)$
$H_0: \theta = 1/2$ vs $H_1: \theta = 3/4$

$\dfrac{f(x,\theta_1)}{f(x,\theta_0)}$ per $x=0,1,2$ ← la densità qui è probabilità discreta.

- $\dfrac{f(0,\frac{3}{4})}{f(0,\frac{1}{2})} = \dfrac{(\frac{1}{4})^2}{(\frac{1}{2})^2} = \dfrac{1}{4}$
- $\dfrac{f(1,\frac{3}{4})}{f(1,\frac{1}{2})} = \dfrac{2\cdot\frac{3}{4}\cdot\frac{1}{4}}{2\cdot\frac{1}{2}\cdot\frac{1}{2}} = \dfrac{3}{4}$
- $\dfrac{f(2,\frac{3}{4})}{f(2,\frac{1}{2})} = \dfrac{(\frac{3}{4})^2}{(\frac{1}{2})^2} = \dfrac{9}{4}$

USO NP

$\Rightarrow$
- $\frac{3}{4} < K < \frac{9}{4}$ $RC = \{X=2\}$ è UMP di livello $\alpha = P(X=2 \mid \theta=\frac{1}{2}) = \frac{1}{4}$
- $\frac{1}{4} < K < \frac{3}{4}$ $RC = \{X \geq 1\}$ UMP di livello $\alpha = 3/4$
- $K < \frac{1}{4}$ $RC = \{X \geq 0\}$ $\alpha = 1$ ← rifiuto sempre
- $K > \frac{9}{4}$ non rifiuto mai, $RC = \{\emptyset\}$ $\alpha = 0$

② $X_1 \dots X_n \sim N(\mu, \sigma^2)$ σ^2 noto.

$H_0: \mu = \mu_0$ vs $H_1: \mu = \mu_1 > \mu_0$

Ci ricordiamo che $\bar{X}_n$ è suff. per μ.

$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ Quindi $g(t, \mu) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left\{-\frac{n}{2\sigma^2}(t-\mu)^2\right\}$

Applicando il corollario di NP

$g(t, \theta_1) > k g(t, \theta_0)$ ← SE È COSÌ RIFIUTO

$\Updownarrow$

$$\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(t-\mu_1)^2 \frac{1}{\sigma^2/n}\right) > k \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(t-\mu_0)^2 \frac{1}{\sigma^2/n}\right)$$

$\Updownarrow$ prendo il log

$$-\frac{1}{2}\frac{(t-\mu_1)^2}{\sigma^2/n} > -\frac{1}{2}\frac{(t-\mu_0)^2}{\sigma^2/n} + \log k$$

$\Updownarrow$

$$-(t-\mu_1)^2 > \frac{2\sigma^2}{n}\log k - (t-\mu_0)^2$$

$\Updownarrow$

$$-t^2 + 2t\mu_1 - \mu_1^2 > \frac{2\sigma^2}{n}\log k - t^2 - \mu_0^2 + 2t\mu_0$$

$\Updownarrow$

$$t > \left(\frac{2\sigma^2}{n}\log k - \mu_0^2 + \mu_1^2\right)\frac{1}{2(\mu_1 - \mu_0)} \quad \text{è UMP.}$$

Il test rifiuta quando t è > di una certa quantità

Se avessi avuto $\mu_1 < \mu_0$ avrei avuto $t < \dots$

RC

$$\left\{\vec{x} : \bar{x}_n > \frac{\frac{2\sigma^2}{n}\log k - \mu_0^2 + \mu_1^2}{2(\mu_1 - \mu_0)}\right\}$$

al crescere di n il limite si abbassa, la gauss si stringe.

Mi manca il livello.

Lo posso imporre giocando con K. potrei riscrivere k in funzione di qst.

Il livello è

$P_{\mu=\mu_0}(RC)$ $\left\{\vec{x} : \bar{x}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right\}$

L'UMP mi dà la forma del livello, poi impongo il BOUND.

In realtà non mi interessa ritornare al K iniziale.

Via NP ho dimostrato che $\left\{\vec{x} : \bar{x}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right\}$ è UMP di livello (e dimensione) α per

$$H_0: \mu = \mu_0 \text{ vs } H_1: \mu = \mu_1 \quad \mu_1 > \mu_0.$$

Se non avessi pensato a $\bar{X}_n$ ma avessi usato la densità del campione avrei ottenuto la stessa RC xk di fatto tanto otterrei la media campionaria.

Se avessi

$H_0: \mu = 0$ vs $H_1: \tilde{\mu} = 1$ $\tilde{\mu}_1 > \mu_0$ ← finché vale qst non cambia nulla.

$RC = \{\bar{X}_n > \tilde{k}\}$ IL LIVELLO GUARDA SOLO μ_0!

64) ⇒ IL TEST È UMP per $H_0: \mu=\mu_0$ vs $H_1: \mu>\mu_0$
Se la RC non dipende da ϑ_1
$\Downarrow$
Posso dire che UMP lo è anche x Hp composte come questo

$\begin{cases} H_0: \vartheta \le \vartheta_0 \\ H_1: \vartheta > \vartheta_0 \end{cases}$ test unilateri $\begin{cases} H_0: \vartheta = \vartheta_0 \\ H_1: \vartheta \neq \vartheta_0 \end{cases}$ test bilateri

$\Downarrow$
$h_0 \neq$ RC

$g \sim X: \Omega \to \mathbb{R}$ 1 5 aprile 2015

Def Una famiglia di leggi $\{g(t,\vartheta), \vartheta \in \Theta\}$ univariata
Ha Likelihood Ratio Monotono (MLR) - LEGGE di una qualche statistica
se $\forall\ \vartheta_2 > \vartheta_1$
$\frac{g(t,\vartheta_2)}{g(t,\vartheta_1)}$ è una funzione monotona non decrescente ↑ di t

ESEMPIO

$g(t,\vartheta) \sim N(\vartheta, \sigma^2 \text{ noto})$

$\frac{g(t,\vartheta_2)}{g(t,\vartheta_1)}$ $\vartheta_2 > \vartheta_1$

$$= \exp\left[-\frac{1}{2\sigma^2}(t-\vartheta_2)^2 + \frac{1}{2\sigma^2}(t-\vartheta_1)^2\right] = \exp\left\{-\frac{1}{2\sigma^2}\left[(t^2-2\vartheta_2 t+\vartheta_2^2)-t^2+2t\vartheta_1-\vartheta_1^2\right]\right\}$$

$$= \exp\left\{\frac{t}{\sigma^2}(\vartheta_2-\vartheta_1)+\dots\right\}$$ → questa funzione è crescente in t

$\Downarrow$
la gaussiana a media incognita e varianza nota
sono con LRM monotono.

$g(t,\vartheta) = h(x)\,c(\vartheta)\exp\{t\,w(\vartheta)\}$
↑ se la famiglia è esponenziale in genere basta controllare cosa succede nell'exp

TEOREMA di KARLIN-RUBIN

$\begin{cases} H_0: \vartheta \le \vartheta_0 \\ H_1: \vartheta > \vartheta_0 \end{cases}$ Sia T una statistica sufficiente per ϑ
e supponiamo che la legge di T,
$\{g(t,\vartheta)\}$ abbia MLR.

Allora, $\forall\ t_0$, il test con $RC = \{T > t_0\}$
è UMP di LIVELLO
$\alpha = P_{\vartheta_0}(T > t_0)$

Quindi il Sup è raggiunto in ϑ_0

dim $\beta(\theta) = P_\theta(RC) = P_\theta(T > t)$ (funzione potenza)

Sia $\theta' > \theta_0$

e consideriamo il test

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta' \end{cases}$$ - test dalle ipotesi semplici -

Dato che T ha MLR

voglio mostrare che $\beta(\theta)$ è non decrescente.

Consideriamo $\theta_1 < \theta_2$ $\Rightarrow \frac{d}{dt}\left[F(t,\theta_2) - F(t,\theta_1)\right] = g(t,\theta_2) - g(t,\theta_1) =$

funzione di ripartizione

$$= \underbrace{g(t,\theta_1)}_{\text{è positiva}} \left[\underbrace{\frac{g(t,\theta_2)}{g(t,\theta_1)}}_{\text{È CRESCENTE in } t} - 1\right]$$

prima < 0, poi > 0

subito > 0 e continua a crescere

↓

Quindi

$\Leftarrow$ $F(t,\theta_2) - F(t,\theta_1)$ può prima decrescere e poi crescere o crescere sempre.

Quindi il massimo di $F(t,\theta_2) - F(t,\theta_1)$ è a $+\infty$ o a $-\infty$, dove vale 0 ⇒ quindi $F(t,\theta_2) - F(t,\theta_1) \le 0$

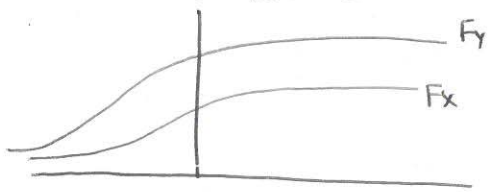
⇓

$F(t,\theta_2) \le F(t,\theta_1)$ ⊛

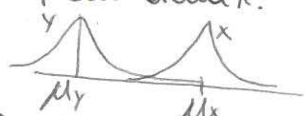
X $\ge$ Y stocasticamente, ossia non è che tutti i valori di X sono > di quelli di Y, ma tendenzialmente

Quando $F_X(t) \le F_Y(t)$

la fdr raccoglie la "massa" della v.a.



la massa della Y viene raccolta prima di quella della X.



è valori che vengono dati da X sono tendenzialmente + grandi dei valori dati da Y (... non è che proprio per forza accade!)

Quindi ho che ho funzioni ordinate stocasticamente

⇒ la mia $\beta(\theta)$ è crescente:

$\beta(\theta) = P_\theta(T > t_0)$; $\theta_2 > \theta_1$.

$\beta(\theta_1) = P_{\theta_1}(T > t_0) = 1 - F(t_0,\theta_1) \overset{⊛}{\le} 1 - F(t_0,\theta_2) = P_{\theta_2}[T > t_0] =$
$= \beta(\theta_2)$

Quindi

Avere MLR crescente

⇓

$\beta(\theta)$ crescente

(b) $H_0: \theta \leq \theta_0$ $\quad RC = \{T > t_0\}$
$H_1: \theta > \theta_0$

ho mostrato $\beta(\theta) \uparrow$.

Quindi vale

1) $\sup\limits_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = P_{\theta_0}(T > t_0) = \alpha$

2) Definiamo $K' = \inf\limits_{\tau} \frac{g(t,\theta')}{g(t,\theta_0)}$ (avevo le ipotesi semplici $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta' \end{cases}$)

con $\tau = \{T > t_0, g(t,\theta) > 0\}$

$$\Downarrow$$

$$T > t_0 \iff \frac{g(t,\theta')}{g(t,\theta_0)} \geq k'$$

$$\Updownarrow$$

$$g(t,\theta') > k' g(t,\theta_0)$$

e grazie al corollario x statistiche sufficienti del lemma NP

test $\{T > t_0\}$ è UMP di livello α. ■

Cerco stat soff con rapporto di likelihood monotono ...

Analogamente

$\begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ $\quad$ T con rapporto di likelihood crescente

$$\Downarrow$$

test con RC $\{T < t_0\}$ sono UMP di livello $\alpha = P_{\theta_0}(T < t_0)$

Spesso abbiamo

$H_0: \theta \leq \theta_0$
$H_1: \theta > \theta_0$ $\quad$ e $\quad$ T con rapporto di likelihood monotono decrescente

(x es.
$\Gamma(\alpha,\beta) \approx \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \beta^{\alpha}$

$\beta_2 > \beta_1 \quad \frac{\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta_2 x} \beta_2^{\alpha}}{\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta_1 x} \beta_1^{\alpha}} = e^{-(\beta_2-\beta_1)x}$
$\Downarrow$
è decrescente)

Allora prendo $-T$, che ha rapporto di likelihood crescente.
Allora avrò
$RC = \{-T > t_0\}$

NP ciclica che

$H_0: \theta = \theta_0$
$H_1: \theta = \theta_1$ $\quad$ $T\uparrow$ UMP è $\{T > t_0\}$

$\theta_1 > \theta_0$ $\quad$ $\alpha = P_{\theta_0}(T > t_0)$

il salto al fatto che sia = per

$H_0 \; \theta \leq \theta_0$
$H_1 \; \theta > \theta_0$ $\quad$ è giusto, xk $T > t_0$ perde memoria del θ_1

$\Downarrow$

Ho UMP del Sup,

$\{T > t_0\}$ è UMP di livello $\alpha = P_{\theta_0}(T > t_0) =$

x la monotonia della funzione potenza $\rightarrow$ $= \sup_{\theta \leq \theta_0} P_\theta(T > t_0)$

∃ SEMPRE L'UMP?

<u>ESEMPIO</u>

$X_1 \dots X_n \sim N(\mu, \sigma^2)$ $\quad$ σ^2 noto

Voglio testare l'ipotesi

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ $\quad$ $\bar{X}_n$ stat soff. abbiamo visto che ha MLR

Fissiamo $\theta_1 < \theta_0$

il test con $RC_1 = \left\{ \bar{X}_n < \frac{\sigma}{\sqrt{n}} z_\alpha + \theta_0 \right\}$ è UMP $\quad$ $\boxed{\text{TEST 1}}$

con $P_{\theta_0}(R_1) = \alpha$ (quantile di ordine α, ho standardizzato la gaussiana)

x NP
Ogni altro test di livello α con potenza in θ_1 pari a quella del test 1 deve avere la stessa R_1 (a meno di un insieme di misura nulla)

$\boxed{\text{TEST 2}}$

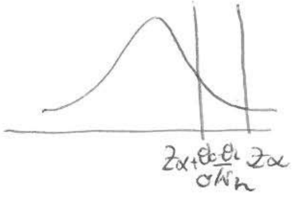
$RC_2 = \left\{ \bar{X}_n > \theta_0 - \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$ $\quad$ $P_\theta(R_2) = \alpha$

$\beta_1(\theta)$ funzione potenza test 1
$\beta_2(\theta)$ " " test 2.

prendiamo $\theta_2 > \theta_0$

devo standardizzare giusto, la media sotto θ_2 è θ_2

$$\beta_2(\theta_2) = P_{\theta_2}\left(\bar{X}_n > \theta_0 - \frac{\sigma}{\sqrt{n}} z_\alpha\right) = P_{\theta_2}\left(\frac{\bar{X}_n - \theta_2}{\sigma/\sqrt{n}} > \underbrace{\frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}}_{<0} - z_\alpha \right) >$$



$$> P(Z > -z_\alpha) = 1 - P(Z \leq -z_\alpha) = 1 + P(Z < z_\alpha) >$$

Z: gaussiana standard

$$> P_{\theta_2}\left(\frac{\bar{X}_n - \theta_2}{\sigma/\sqrt{n}} < z_\alpha + \underbrace{\frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}}_{<0} \right)$$

$$= P_{\theta_2}\left(\bar{X} < \theta_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \right) =$$

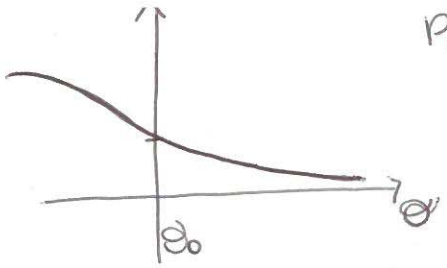
$$= \beta_1(\theta_2)$$

Quindi ho

$\beta_2(\theta_2) > \beta_1(\theta_2)$. Ma il nostro test era UMP $\Rightarrow$ Assurdo

$\Rightarrow \nexists$ UMP.

(68) IL TEST 1

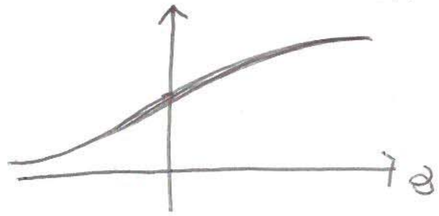


TEST 1

$$P_\theta\left(\bar X < \theta_0 + \frac{\sigma}{\sqrt n} z_\alpha\right) = \phi\left(z_\alpha + \frac{\theta_0-\theta}{\sigma/\sqrt n}\right)$$

↑ in θ_0 vale α

TEST 2 $\quad P_\theta\left(\bar X_n > \theta_0 - \frac{\sigma}{\sqrt n} z_\alpha\right) = 1 - \phi\left(-z_\alpha + \frac{\theta_0-\theta}{\sigma/\sqrt n}\right)$



TEST 2

Per $\theta = \theta_0$ vs $\theta \neq \theta_0$
dobbiamo cambiare la classe per cercare i test,
xk questi sono entrambi distorti per $\theta = \theta_0$ vs $\theta \neq \theta_0$

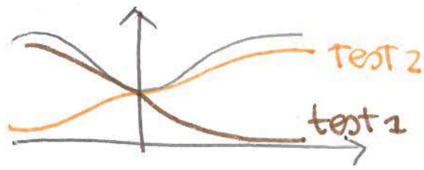
NON DISTORTO:

$$\beta(\theta'') \geq \beta(\theta')$$
$$\forall \theta'' \in \Theta_0^c$$
$$\forall \theta' \in \Theta_0$$

Qui non è vero!
Questi 2 test che sono UMP per $\theta >$ o $\theta <$
non possono esserlo x il caso bilatero, xk
qui sono distorti
⇓
ci restringiamo al caso dei test distorti
e abbiamo

$\boxed{\text{TEST 3}}$

$$R_3 = \left\{\bar X_n > \theta_0 - \frac{\sigma}{\sqrt n} z_{\frac{\alpha}{2}}\right\} \cup \left\{\bar X_n < \theta_0 + \frac{\sigma}{\sqrt n} z_{\frac{\alpha}{2}}\right\}$$



è non distorto,
il suo minimo
è proprio in θ_0

↑ non riesce a vincere
gli altri 2 test nei due rami,
ma vince nella parte non distorta.

NP si può usare anche in un contesto del tipo

$$\begin{cases} H_0 : f \sim N(\mu_0, \sigma_0^2) \\ H_1 : f \sim \text{Cauchy} \end{cases}$$

faccio il rapporto e rifiuto quando il rapporto è $> K$.

TEST UNIONE INTERSEZIONE
o
INTERSEZIONE UNIONE

UI

L'IPOTESI NULLA $H_0 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma$

Γ può essere finito o numerabile

$H_1 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma^c$

supponiamo di conoscere

$H_{0\gamma} : \theta \in \Theta_\gamma$

$H_{1\gamma} : \theta \in \Theta_\gamma^c$

abbia $RC = R_\gamma$ $\Rightarrow$ RC per test UI

$RC = \bigcup_\gamma R_\gamma$

ESEMPIO

$X_1 \dots X_n \sim N(\theta, \sigma^2)$

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

posso leggerlo come test unione intersezione:

$\theta = \theta_0 \qquad \Theta_1 \cap \Theta_2$

$\Theta_1 = \{\theta \leq \theta_0\} \qquad RC = \{\bar{X}_n < \dots\}$

$\Theta_2 = \{\theta \geq \theta_0\} \qquad RC = \{\bar{X}_n > \quad\}$

Quindi

RC per $\theta = \theta_0$: $RC = \{\bar{X}_n < t_1\} \cup \{\bar{X}_n > t_2\}$

e t_1 e t_2 le posso scegliere come voglio per avere livello α,

se prendo le due RC a $\frac{\alpha}{2}$ e $\frac{\alpha}{2}$, simmetrico, ho UMP, ma potrei anche decidere di non prenderlo simmetrico

IU

$H_0 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma$

$H_1 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma^c$

$H_{0\gamma} : \theta \in \Theta_\gamma$
$H_{1\gamma} : \theta \in \Theta_\gamma^c$ hanno regione critica R_γ

$\Rightarrow RC = \bigcap_{\gamma \in \Gamma} R_\gamma$ sono false quando sono false tutte, ossia sto nell'intersezione delle R_γ

21 aprile 2015

> TEST UI

$H_0 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma = \Theta_0$

$H_1 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma^c = \Theta_0^c$

$RC = \bigcup_{\gamma \in \Gamma} RC_\gamma$

regioni critiche a γ fissato.

Sia $\lambda_\gamma(\vec{x})$ la LRT statistics per $H_0 : \theta \in \Theta_\gamma$ vs $H_1 : \theta \in \Theta_\gamma^c$

$\lambda(\vec{x})$ la LRT statistics per $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$

TEOREMA

- $T(\vec{x}) = \inf_{\gamma \in \Gamma} \lambda_\gamma(\vec{x})$

Sia $R_T = \{ \vec{x} : \lambda_\gamma(\vec{x}) < c$ per almeno un $\gamma \in \Gamma \} =$
$= \{ \vec{x} : T(\vec{x}) < c \}$

$R_\lambda = \{ \vec{x} : \lambda(\vec{x}) < c \}$ (LRT)

Allora

a) $T(\vec{x}) \geq \lambda(\vec{x}) \quad \forall \vec{x}$

b) $\beta_T(\theta) \leq \beta_\lambda(\theta) \quad \forall \theta$

test unione intersezione

c) Se il LRT (λ) ha livello $\alpha \Rightarrow$ anche il U.IT (T) ha livello α

<u>dim</u>

a) $\Theta_0 = \bigcap_{\gamma \in \Gamma} \Theta \subset \Theta_\gamma \quad \forall \gamma \in \Gamma$

$$\underbrace{\lambda_\gamma(\vec{x})}_{LRT} = \sup_{\theta \in \Theta_\gamma} L(\theta, \vec{x}) \frac{1}{\sup_{\theta \in \Theta} L(\theta, \vec{x})} \geq \frac{\sup_{\theta \in \Theta_0} L(\theta, \vec{x})}{\sup_{\theta \in \Theta} L(\theta, \vec{x})} = \lambda(\vec{x}) \quad \forall \gamma.$$

il sup su Θ_γ è $\geq$ sup su Θ_0 che è l'intersezione dei Θ_γ

Allora abbiamo che $\forall \gamma$

$$\lambda_\gamma(\vec{x}) \geq \lambda(\vec{x})$$

$$\Downarrow$$

$$T(\vec{x}) = \inf_\gamma \lambda_\gamma(\vec{x}) \geq \lambda(\vec{x})$$

---

b) $\beta_T(\theta) = P_\theta(T(\vec{x}) < c) \leq P_\theta(\lambda(\vec{x}) < c) = \beta_\lambda(\theta)$

$\{\vec{x} : T(\vec{x}) < c\} \subset \{\vec{x} : \lambda(\vec{x}) < c\}$

(la uso usando la monotonia della probabilità)

---

c) $\sup_{\theta \in \Theta_0} \beta_T(\theta) \leq \sup_{\theta \in \Theta_0} \beta_\lambda(\theta) \leq \alpha$

■

> TEST IU

$$H_0 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma = \Theta_0$$

$$H_1 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma^c = \Theta_0^c$$

$$R = \bigcap R_\gamma$$

TEOREMA

Sia α_γ il livello del test per $H_{0\gamma}$ e R_γ

Allora il test IU $R = \bigcap_{\gamma \in \Gamma} R_\gamma$ ha livello $\sup_\gamma \alpha_\gamma$

<u>dim</u>

$\theta \in \Theta_0 : \theta \in \Theta_\gamma$ per qualche γ

$P_\theta(\vec{x} \in R) \leq P_\theta(\vec{x} \in R_\gamma) \leq \alpha_\gamma \leq \alpha = \sup_\gamma \alpha_\gamma$

72

TEOREMA

$H_0 \quad \theta \in \bigcup_{j=1}^{K} \Theta_j \qquad \Gamma = \{1, \dots K\}$

$\forall j = 1 \dots K$ sia R_j regione di rifiuto per un test di livello α per H_{0j}

e supponiamo che per qualche $i = 1, \dots, K$ $\exists$ una successione di parametri $\theta_\ell \in \Theta_i$ tc

1) $\lim_{\ell \to +\infty} P_{\theta_\ell}(\vec{X} \in R_i) = \alpha$

2) $\forall j = 1 \dots K \neq i \quad \lim_{\ell \to +\infty} P_{\theta_\ell}(\vec{X} \in R_j) = 1$

$\Rightarrow$ il TEST IU $R = \bigcap_{j=1}^{K} R_j$ ha dimensione α.

dim

dal teorema precedente : R ha livello α.

$\theta_\ell \in \Theta_i \subset \Theta_0$

$$\sup_{\theta \in \Theta_0} P_\theta(\vec{X} \in R) \geq \lim_{\ell \to +\infty} P_{\theta_\ell}(\vec{X} \in R) = \lim_{\ell \to +\infty} P_{\theta_\ell}\left(\vec{X} \in \bigcap_{j=1}^{K} R_j\right) \geq$$

- disuguaglianza di Bonferroni:

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

$$\geq \lim_{\ell \to \infty} \Big[\underbrace{\sum_{j=1}^{K} P_{\theta_\ell}(\vec{X} \in R_j)} - (K-1) \Big] =$$

il limite per $j = i$ fa α.
Negli altri casi fa 1

$$= \underbrace{K - 1 + \alpha}_{\lim_{\ell \to \infty} \sum P_{\theta_\ell}(\vec{X} \in R_j)} - (K-1) = \alpha$$

$\lim_{\ell \to \infty} \sum P_{\theta_\ell}(\vec{X} \in R_j) = 1 \cdot (K-1) + \alpha \cdot 1$

$(K-1)$: numero di j per cui fa 1

$\alpha \cdot 1$: per $j = i$ fa α

■

ESEMPIO

Azienda che produce tessuti.

2 parametri importanti per il tessuto.

θ_1 : resistenza media alla trazione

θ_2 : probabilità di passare un test di infiammabilità.

Supponiamo che ci siano degli standard:

θ_1 deve essere > 50 pounds

θ_2 deve essere > 0.95

Devo dimostrare $\theta_1 > 50$

$\theta_2 > 0.95$

Se deve passare entrambi i test è esattamente un UI

$$H_0 : \{\theta_1 \leq 50 \lor \theta_2 \leq 0.95\}$$

$$H_1 : \{\theta_1 > 50 \land \theta_2 > 0.95\}$$ ← in H_1 metto quello che voglio dimostrare.

$X_1 \dots X_n$ iid $N(\theta_1, \sigma^2)$ ← trazione

LRT, test per H_{0_1}

$$R_1 = \left\{ \frac{\bar{X}_n - 50}{S/\sqrt{n}} > t \right\}$$

$Y_1 \dots Y_n$ iid $Be(\theta_2)$ ← infiammabilità

test per H_{0_2}

$$R_2 = \left\{ \sum Y_i > b \right\}$$

$$R = \left\{ (\vec{x}, \vec{y}) : \frac{\bar{x} - 50}{S/\sqrt{n}} > t \quad e \quad \sum Y_i > b \right\}$$

Fissiamo $\alpha_1 = \alpha_2 = 0.95$

$n = m = 58$

θ_e è una successione dei numeri reali che va a ∞ x es $\theta_e = \ell$

$$t = t_{0,95}(57) = 1,672$$

$$b = 57$$

Posso dire che il test ha dimensione $\alpha = 0.05$ se riesco ad applicare il teorema precedente.

$$\vec{\theta}_\ell = (\theta_{1\ell}, \theta_2) = (\ell, 0,95) \;,\; \vec{\theta}_\ell \in \Theta.$$

$$\lim_{\ell \to +\infty} P_{\theta_\ell}(\vec{x} \in R_1) = 1$$

$$\lim_{\ell \to +\infty} P_{\theta_\ell}(\vec{y} \in R_2) = 0.05$$

Quindi posso applicare il teorema e ho che il test ha dimensione 5%.

In qst modo non ho bisogno di modellizzare la dipendenza di x e y

(74)

P-VALUE

Def p-VALUE $p(\vec{x})$ È UNA STATISTICA CHE $0 \le p(\vec{x}) \le 1$

Valori piccoli di $p(\vec{x})$ danno evidenza per H_1 vero (rifiuto H_0)

Una STATISTICA p-value è valida se $\forall \theta \in \Theta_0$ e $\forall\ 0 \le \alpha \le 1$

$$P_\theta(p(\vec{x}) \le \alpha) \le \alpha$$

sotto H_0 la probabilità di vedere valori piccoli è piccola.

$\Downarrow$

se $p(\vec{x})$ è un p-value valido e uso

$$RC = \{\vec{x} : p(\vec{x}) \le \alpha\}$$

$\Downarrow$

il test ha livello α

perché $\sup_{\theta \in \Theta_0} P_\theta(RC) = \sup_{\theta \in \Theta_0} P_\theta(p(\vec{x}) \le \alpha) \le \alpha$

ESEMPIO

$X_1 \dots X_n \sim N(\mu, 1)$

$$\begin{cases} H_0 : \mu = \mu_1 \\ H_1 : \mu \ne \mu_1 \end{cases} \qquad RC_\alpha = \left\{ \frac{|\bar{X}_n - \mu_0|}{1}\sqrt{n} > z_{1-\frac{\alpha}{2}} \right\}$$

Al 1^ Anno: p-value: più piccolo α per cui rifiuto H_0

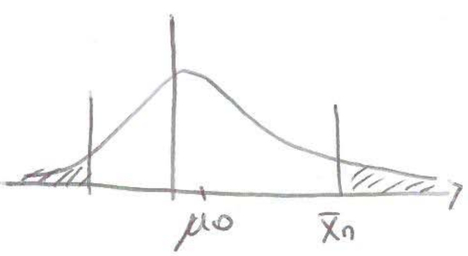
$\forall \alpha > p$ rifiuto

$\forall \alpha \le p$ accetto

È una statistica tra 0 e 1? è valida?

Possiamo riscriverlo:

$p(\vec{x}) = \inf_\alpha \{\vec{x} \in R_\alpha\}$ è una statistica. ed è un livello, quindi è tra 0 e 1.



$$p(\vec{x}) = \inf_\alpha \{\vec{x} \in R_\alpha\} =$$
$$= \inf_\alpha \{|\bar{X}_n - \mu_0|\sqrt{n} > z_{1-\frac{\alpha}{2}}\} =$$
$$= 2\left[1 - \Phi\left(\frac{|X_n| - \mu_0}{1/\sqrt{n}}\right)\right]$$

Costruiamo un test che abbia livello esattamente dove finiscono i miei dati.

Dobbiamo vedere se $\forall \theta \in \Theta_0 \quad P_\theta(p(\vec{x}) \le \alpha) \le \alpha$

Qui H_0 è $\mu = \mu_0$

Quindi devo calcolare

$$P_{\mu_0}(p(\vec{x}) \le u) = P_{\mu_0}\left[2(1-\Phi(|\bar{x}_n-\mu_0|\sqrt{n}) \le u\right] =$$

$$= P_{\mu_0}\left[\Phi(|\bar{x}_n-\mu_0|\sqrt{n}) \ge 1-\frac{u}{2}\right] =$$

$$= P_{\mu_0}\left[\underbrace{|\bar{x}_n-\mu_0|\sqrt{n}} \ge \Phi^{-1}\left(1-\frac{u}{2}\right)\right] =$$

Gaussiana standard xk la sto calcolando sotto μ_0

(è come se avessi $P(|z|>t)=2[1-\Phi(t)]$)

$$= 2\left[1-\Phi\left(\Phi^{-1}\left(1-\frac{u}{2}\right)\right)\right] = u.$$

Quindi
$P_{\mu_0}(p(\vec{x}) \le \alpha) = \alpha \Rightarrow$ il p-value del 1^ anno era valido

Il pvalue è una statistica _ lo costruiamo a partire solo dai dati

Come Costruire un p-value valido?

TEOREMA

Sia $W(\vec{X})$ una statistica test tc valori grandi di W danno evidenza a favore di H_1.

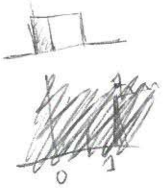
$$p(\vec{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{X}) \ge W(\vec{x}))$$ è un p-value valido

dim

$\theta \in \Theta_0$

$F_\theta(w)$ funzione di ripartizione di $-W(\vec{X})$

$$p_\theta(\vec{x}) = P_\theta(W(\vec{X}) \ge W(\vec{x})) = P_\theta(-W(\vec{X}) \le -W(\vec{x})) = F_\theta(-W(\vec{x}))$$



$Y = F_X(X)$

$\Rightarrow P(Y \le t) = P(F_X(X) \le t) = P(X \le F_X^{-1}(t)) = F_X(F_X^{-1}(t)) = t$

Quindi $Y \sim U$ nel caso continuo (nel discreto è stocasticamente $\ge$ un'uniforme)

$\forall\ 0 \le \alpha \le 1$ xk sto considerando anche casi discreti.

$$P_\theta(P_\theta(\vec{X}) \le \alpha) \le \alpha$$

$$p(\vec{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{X}) \ge W(\vec{x})) = \sup_{\theta \in \Theta_0} P_\theta(\vec{x}) \ge P_\theta(\vec{x}) \quad (*)$$

$\forall \theta \in \Theta_0$

$$\Rightarrow P_\theta(P(\vec{X}) \le \alpha) = P_\theta\left(\sup_{\theta \in \Theta_0} P_\theta(\vec{X}) \le \alpha\right) \le P_\theta(P_\theta(\vec{X}) \le \alpha) \le \alpha$$

$\theta \in \Theta_0$

uso $*$: $P(\vec{x}) \le \alpha \subset P_\theta(x) \le \alpha$

già dim. ■

 22 aprile 2015

Stime intervallari

$\theta \in \mathbb{R}$ $\qquad T(\vec{X}) : \Omega \to \mathbb{R}$

$\vec{\theta} \in \mathbb{R}^p$ $\qquad T(\vec{X})(\omega) = t$

Ora dovremmo dire

$\theta \in \Theta(\vec{X})$ regione del piano i cui limiti sono conoscibili conoscendo solo i dati
↳ STATISTICHE

Def UNA STIMA INTERVALLARE di un parametro reale θ incognito è una coppia di STATISTICHE, $L(\vec{X}), U(\vec{X})$ tc

$$L(\vec{X}) \leq U(\vec{X})$$

$$\theta \in [L(\vec{X}), U(\vec{X})]$$

Limite inferiore e limite superiore
(Potrebbero anche essere degeneri - $U(x) = +\infty$, $L(x) = -\infty$)

Esempio

$X_1 \dots X_4 \sim N(\mu, 1)$

Normalmente fino ad ora x trovare μ abbiamo usato $\bar{X}_4$
Una stima intervallare potrebbe essere

$IC = [\bar{X}_4 - 1, \bar{X}_4 + 1]$

In qst esempio $P(\bar{X}_4 = \mu) = 0$
$\Rightarrow$ Le mie stime puntuali prendono probabilità 0.

Posso calcolare invece

$$P(\mu \in IC) = P(\bar{X}_4 - 1 \leq \mu \leq \bar{X}_4 + 1) =$$
$$= P(|\bar{X}_4 - \mu| \leq 1) = P(|Z| \leq 2) = 0{,}9544$$

$\uparrow$
$\bar{X}_4 \sim N(\mu, \frac{1}{4})$

mettiamoci nel caso delle trazioni di ieri.

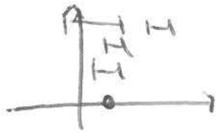
$x_1 = 50{,}1$
$x_2 = 50{,}3$ $\qquad \bar{x}_4 = 50{,}15$
$x_3 = 50{,}2$
$x_4 = 50$

Costruisco $IC = [49{,}15; 51{,}15]$

(Mappo θ su un intervallo)

Il vero μ sta lì dentro o no? Bo.
Realizzando un IC, questo o ci prende o non ci prende
Non lo posso sapere!

0.9544 mi dice $P(\mu \in [L(\vec{x}), U(\vec{x})]$
usando quelle variabili aleatorie.
Il 95% degli intervalli dirà la cosa giusta
Simulando, molti intervalli conterranno la media, ma il 5% non la terranno



Non è che $\mu \in IC$ con $P = 0.95$!!
$\nexists$ $P(\mu \in [L(\vec{x}), U(\vec{x})]) \approx 0.95$

Def. DATO UNO STIMATORE INTERVALLARE $[L(\vec{x}), U(\vec{x})]$ per θ
$\Downarrow$
$$P_\theta\left(\theta \in [L(\vec{x}), U(\vec{x})]\right)$$ È DETTA
PROBABILITÀ di COPERTURA

- Dato uno stimatore intervallare $[L(\vec{x}), U(\vec{x})]$ per θ
$\Downarrow$
$$\inf_\theta P_\theta\left(\theta \in [L(\vec{x}), U(\vec{x})]\right)$$ È DETTA
LIVELLO o COEFFICIENTE di CONFIDENZA

Non conoscendo θ, ci interessa sapere se il caso peggiore rispetto a θ (l'inf) abbia copertura limitata. Lo dobbiamo tenere sotto controllo.
La probabilità di copertura bo.
Parleremo di intervallo di confidenza di livello $1-\alpha$.

ESEMPIO

$X_1 \dots X_n \sim U_{[0,\theta]}$ non è exponenziale.

$X_{(n)} = \max(X_1 \dots X_n)$ è la statistica 40.

$IC_1 = [aX_{(n)}, bX_{(n)}]$ $\quad 1 \leq a < b$
$\uparrow$ sarebbe sciocco fare una stima sotto il massimo

$IC_2 = [X_{(n)} + c, X_{(n)} + d]$
$0 \leq c < d$.

Calcoliamo probabilità di copertura e livello di confidenza x poi scegliere quale prendere

1) $P_\theta(\theta \in I_1) = P_\theta(aX_{(n)} \leq \theta \leq bX_{(n)}) = P_\theta\left(\frac{1}{b} \leq \frac{X_{(n)}}{\theta} \leq \frac{1}{a}\right) =$
$= P_\theta\left(\frac{1}{b} \leq T \leq \frac{1}{a}\right) =$
$= F_T\left(\frac{1}{a}\right) - F_T\left(\frac{1}{b}\right) =$
$= \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n$

$$\boxed{\begin{aligned} &T = \frac{X_{(n)}}{\theta} \sim ? \\ &F_{X_{(n)}}(t) = \left(\frac{t}{\theta}\right)^n \Rightarrow F_T(t) = P(T \leq t) = \\ &= P(X_{(n)} \leq t\theta) = \\ &= F_{X_{(n)}}(t\theta) = t^n \quad \text{per } 0 \leq t \leq 1 \end{aligned}}$$

non dipende da θ!
$\Downarrow$
Facendo l'inf resta lui.
= livello di IC_1

Posso trovare infiniti a e b t.c. l'intervallo abbia $P = 95\%$

2) $P_\theta(\theta \in IC_2) = P_\theta\left(X_{(n)} + c \le \theta \le X_{(n)} + d\right) =$

$= P_\theta\left(1 - \frac{d}{\theta} \le \frac{X_{(n)}}{\theta} \le 1 - \frac{c}{\theta}\right) = P_\theta\left(1 - \frac{d}{\theta} \le T \le 1 - \frac{c}{\theta}\right) =$

$= \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n$

Questa probabilità di confidenza varia al variare di θ.
Ne calcolo l'inf.

Oss $$\lim_{\theta \to +\infty} \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n = 0$$

Allora
$\inf_\theta P_\theta(\theta \in IC_2) = 0$ ← livello di confidenza di IC_2
Bleah!

Quindi compro IC_1.
IC_1 perché memoria di θ xk è costruito su $\frac{X_{(n)}}{\theta}$,
mentre IC_2 è additivo.

la legge dipende da θ

non è una statistica. Ma la sua legge non dipende da θ

Metodi per COSTRUIRE IC

1) INVERSIONE di un TEST

Mettiamo in corrispondenza biunivoca test e IC.

Esempio
$X_1 \dots X_n \sim N(\mu, \sigma^2)$ σ^2 noto

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \ne \mu_0 \end{cases} \quad \alpha \quad R_\alpha = \left\{ \vec{x} : |\bar{X}_n - \mu_0| > z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\}$$

se σ incognita avrei $t(n-1)$ e qui S

$P_{\mu_0}(R_\alpha) = \alpha$

$P_{\mu_0}(R_\alpha^c) = 1 - \alpha$

$1 - \alpha = P_{\mu_0}(R_\alpha^c) = P_{\mu_0}\left(|\bar{X}_n - \mu_0| \le \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) =$

$= P_{\mu_0}\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \le \mu_0 \le \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$

vale per ogni μ_0

Quindi $\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} , \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]$ è IC (xk ho 2 estremi che sono 2 statistiche)

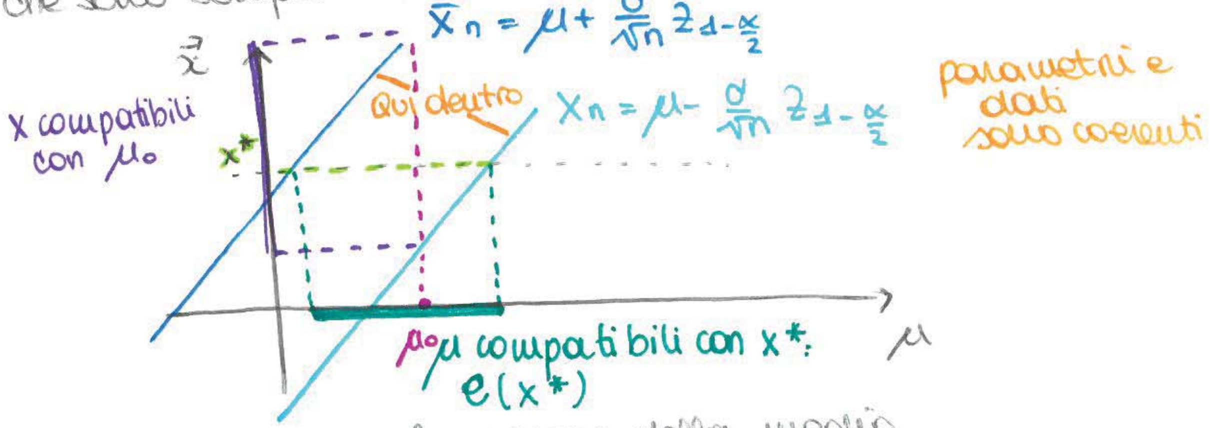
di livello $1-\alpha$ per μ ← xk di fatto posso rifare lo stesso conto $\forall \mu_0$
Dove μ è il mio parametro incognito

$$A(\mu_0) = \left\{ \vec{x} : \mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \bar{X}_n \le \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right\}$$

$$\mathcal{C}(\vec{x}) = \left\{ \mu : \bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \le \mu \le \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right\}$$

$$\boxed{(\vec{x} \in A(\mu_0)) \iff \mu_0 \in \mathcal{C}(\vec{x})}$$

Quando faccio un test fisso il parametro e accetto i dati accettabili per l'IC invece ho i miei dati e cerco i valori del parametro che sono compatibili.



plottiamo la realizzazione della media campionaria sullo spazio dei parametri.
Fossi un indovino, avrei la bisettrice.
Ma invece mi concedo una tolleranza

TEOREMA

1) $\forall\, \theta_0 \in \Theta$ sia $A(\theta_0)$ la regione di accettazione di un test di livello α per $H_0: \theta = \theta_0$

$\forall\, \vec{x}$ definiamo $\mathcal{C}(\vec{x}) = \{\theta_0 : \vec{x} \in A(\theta_0)\}$

Allora $\mathcal{C}(\vec{x})$ è una regione di confidenza di livello $1-\alpha$

2) Sia $\mathcal{C}(\vec{x})$ una regione di confidenza di livello $1-\alpha$

$\forall\, \theta_0 \in \Theta$ sia $A(\theta_0) = \{\vec{x} : \theta_0 \in \mathcal{C}(\vec{x})\}$

$\Downarrow$

$A(\theta_0)$ è la regione di accettazione di un test di livello α per $H_0: \theta \le \theta_0$

dim

1) $\underbrace{P_{\theta_0}(\vec{x} \notin A(\theta_0))}_{= P(\vec{x} \in RC)} \le \alpha \Rightarrow P_{\theta_0}(\vec{x} \in A(\theta_0)) \ge 1-\alpha$

Quindi $\forall \theta_0$ $P_{\theta_0}(\theta_0 \in \mathcal{C}(\vec{x})) = P_{\theta_0}(\vec{x} \in A(\theta_0)) \ge 1-\alpha$

$\Downarrow$ se vale $\forall \theta_0$

$\inf_{\theta} P_\theta(\theta \in \mathcal{C}(x)) \ge 1-\alpha$

$\Rightarrow$ l'IC ha livello $1-\alpha$

2) Errore di 1^ tipo per $H_0: \theta = \theta_0$

$$P_{\theta_0}\left(\vec{x} \notin A(\theta_0)\right) = P_{\theta_0}\left(\theta_0 \notin \mathcal{C}(\vec{x})\right) \le \alpha$$

Quindi il test ha livello α. ■

La forma della RC dà la forma della regione di confidenza.
Prima partivo da un intervallo e ottenevo un intervallo.
Test bilateri producono intervalli,
test unilateri producono semirette.

ESEMPIO $X_1 \dots X_n \sim N(\mu, \sigma^2)$

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \qquad \text{LRT:} \quad RC = \left\{ \bar{X}_n < \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

$$A(\mu_0) = \left\{ \vec{x} : \bar{X}_n \ge \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \right\} =$$

$$= \left\{ \vec{x} : \mu_0 \le \bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

Quindi dal teorema:

$$\mathcal{C}(\vec{x}) = \left\{ \mu : \mu \le \bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$ è regione di confidenza di livello $1-\alpha$

$$\left(-\infty, \bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\alpha\right]$$

Quindi ho una semiretta

ESEMPIO

$X_1 \dots X_n \sim \mathcal{E}\left(\frac{1}{\lambda}\right)$

Voglio costruire un intervallo di confidenza per $E[X_i] = \lambda$

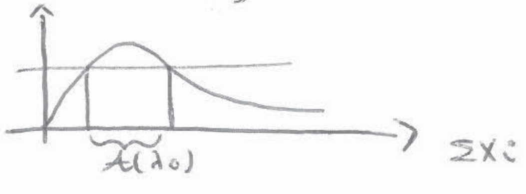
$$\begin{cases} H_0: \lambda = \lambda_0 \\ H_1: \lambda \neq \lambda_0 \end{cases} \qquad \text{LRT:} \quad \lambda(\vec{x}) = \frac{\left(\frac{1}{\lambda_0}\right)^n \exp\left(-\sum x_i \frac{1}{\lambda_0}\right)}{\left(\frac{1}{\hat{\lambda}_{MLE}}\right) \exp\left(-\sum x_i \frac{1}{\hat{\lambda}_{MLE}}\right)} = \frac{\left(\frac{1}{\lambda_0}\right)^n \exp\left(-\sum x_i \frac{1}{\lambda_0}\right)}{\left(\frac{1}{\bar{X}_n}\right)^n \exp\left(-\frac{\sum x_i}{\bar{X}_n}\right)}$$

$$\left[\hat{\lambda}_{MLE}: \left(\frac{1}{\lambda}\right)^n \exp\left(-\frac{\sum x_i}{\lambda}\right) \hookrightarrow -n \log \lambda - \frac{\sum x_i}{\lambda} \Rightarrow \text{derivo} \Rightarrow -\frac{n}{\lambda} + \frac{\sum x_i}{\lambda^2} = 0 \qquad \hat{\lambda}_{MLE} = \bar{X}_n \right]$$

Quindi

$$\boxed{\lambda(\vec{x}) = \left(\frac{\sum x_i}{n \lambda_0}\right)^n e^n e^{-\frac{\sum x_i}{\lambda_0}}}$$ statistica test

$$A(\lambda_0) = \left\{ \vec{x} : \left(\frac{\sum x_i}{\lambda_0}\right)^n e^{-\frac{\sum x_i}{\lambda_0}} > k \right\} \qquad k \text{ è } k(\alpha)$$



Quindi

$$\mathcal{C}(\vec{X}) = \left\{ \lambda : \left(\frac{\sum x_i}{\lambda} \right)^n e^{-\frac{\sum x_i}{\lambda}} \geq k \right\}$$

23 aprile 2015

2) QUANTITÀ PIVOTALE

Def QUANTITÀ PIVOTALE
una variabile aleatoria $Q(\vec{X}, \theta) = Q(X_1 \dots X_n, \theta)$ ← non è una statistica!
è detta quantità pivotale se la legge di $Q(\vec{X}, \theta)$ non dipende da θ.

- $X_1 \dots X_n \sim U[0, \theta]$
  $T = \frac{X_{(n)}}{\theta}$ $\quad F_T(t) = t^n$ → non dipende da θ ⇒ la legge non dipende da θ
  ⇒ $\frac{X_{(n)}}{\theta}$ è QUANTITÀ PIVOTALE

- $X_1 \dots X_n \sim N(\mu, \sigma^2)$
  $\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1)$ $\quad$ Quindi $\frac{\bar{X}_n - \mu}{S/\sqrt{n}}$ è quantità pivotale.

- $X_1 \dots X_n \sim \mathcal{E}\left(\frac{1}{\lambda}\right)$
  $\sum X_i \sim \Gamma\left(n, \frac{1}{\lambda}\right)$ $\quad$ Possiamo costruire, a partire da qst, un Q?
  Sì:
  $\frac{\sum X_i}{\lambda} \sim \Gamma(n, 1)$ $\quad \left(\begin{array}{l} \text{xk} \quad X \sim \Gamma(\alpha, \beta) \\ Y = kX \sim \Gamma\left(\alpha, \frac{\beta}{k}\right) \\ k > 0 \end{array} \right)$

Per motivi di comodità ci si riporta spesso a una χ
$$\chi^2(m) \sim \Gamma\left(\frac{m}{2}, \frac{1}{2}\right)$$
Quindi è utile costruire
$$\frac{2\sum X_i}{\lambda} \sim \Gamma\left(\frac{2n}{2}, \frac{1}{2}\right) \sim \chi^2(2n)$$
↑ Non è una statistica, ma è una quantità pivot.

Perché sono utili x costruzione IC?
Prendiamo $Q(\vec{X}, \theta)$ quantità pivot e
coppia (a, b) tc $\quad P(a \leq Q(\vec{X}, \theta) \leq b) = 1 - \alpha$

a e b non dipendono da θ

$$\mathcal{C}(\vec{X}) = \{ \theta : a \leq Q(\vec{X}, \theta) \leq b \} = \{ L(\vec{X}, a) \leq \theta \leq U(\vec{X}, b) \}$$

↑ SE sono in grado di invertire la quantità pivotale

 ESEMPIO

1) $X_1 \dots X_n \sim \mathcal{E}\left(\frac{1}{\lambda}\right)$

voglio un IC per λ

So che $\frac{2\sum X_i}{\lambda} \sim \chi^2(2n)$ • scegliamo l'IC in modo simmetrico

$$1-\alpha = P\left(\chi^2_{\frac{\alpha}{2}}(2n) \le \frac{2\sum X_i}{\lambda} \le \chi^2_{1-\frac{\alpha}{2}}(2n)\right)$$

$$\Updownarrow$$

$$1-\alpha = P\left(\frac{2\sum X_i}{\chi^2_{1-\frac{\alpha}{2}}(2n)} \le \lambda \le \frac{2\sum X_i}{\chi^2_{\frac{\alpha}{2}}(2n)}\right)$$

2) $X_1 \dots X_n \sim N(\mu, \sigma^2)$

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Io so

$$1-\alpha = P\left(t_{\frac{\alpha}{2}}(n-1) \le \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \le t_{1-\frac{\alpha}{2}}(n-1)\right)$$

*

$$\Updownarrow$$

$$1-\alpha = P\left(\bar{X}_n - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) \le \mu \le \bar{X}_n - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\right)$$

È UGUALE A quello che avevamo scritto

$$= \left[\bar{X}_n \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}^{(n-1)}\right]$$

xk questo è = poi.

Ma ho scelto di farlo simmetrico. Nessuno mi vieta di non farlo simmetrico.

Perché ho fatto questa scelta?

* se avessi scelto $\frac{\alpha}{3}$ avrei avuto un intervallo + lungo

Metodi per VALUTARE IC

A parità di confidenza $(1-\alpha)$ cerco l'intervallo di lunghezza minima

xk se minimizzo la lunghezza, quando prendo un intervallo che è anche preciso

C'è però ovviamente un trade-off tra confidenza e lunghezza degli intervalli

(Non posso avere confidenza 100% e piccolo IC, ma dovrei avere tutta la retta)

$X_1 \dots X_n \sim N(\mu, \sigma^2)$ σ^2 noto

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

$$P(a \le Z \le b) = 1-\alpha$$

$$\left[a \le \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le b\right]$$

$$\Updownarrow$$

$$\left[\bar{X}_n - \frac{b\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n - \frac{a\sigma}{\sqrt{n}}\right]$$

la lunghezza dell'IC è

$$L(IC) = \frac{\sigma}{\sqrt{n}}(b-a) \Rightarrow$$ dipende proprio dai quantili.

Data la simmetria della legge di z, si ha che $b-a$ è minimizzato quando $a = -b$.

Questa proprietà vale x .:.

Def si dice che $f_x(x)$ è UNIMODALE se $\exists\, x^*$ tc $f_x(x)$ è non decrescente $\forall x \leq x^*$ e non crescente $\forall x \geq x^*$

(ha un solo massimo, può essere non simmetrica)

TEOREMA

Sia $f_x(x)$ una densità unimodale.

Se l'intervallo $[a,b]$ soddisfa

a) $\int_a^b f_x(x)\,dx = 1-\alpha$

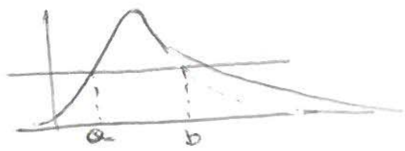
b) $f_x(a) = f_x(b) > 0$ e

c) $a \leq x^* \leq b$ dove x^* è la moda di f

$\Downarrow$

$[a,b]$ è l'intervallo di lunghezza minima tra quelli che soddisfano a)

Se avessi



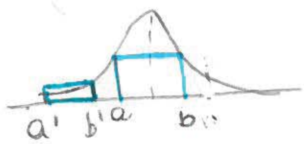
non avrei simmetria rispetto all'asse della moda

Dim

Sia $[a',b']$ un intervallo tc $b'-a' < b-a$ (*)

Non perdiamo generalità prendendo $a' < a$

i) Sia $b' \leq a \Rightarrow a' \leq b' \leq a \leq x^*$



$$\int_{a'}^{b'} f_x(x)\,dx \leq f_x(b')(b'-a') \leq f_x(a)(b'-a') < f_x(a)(b-a)$$

xk la funzione sta crescendo, sono prima del massimo

x l'assunzione (*)

$$\leq \int_a^b f_x(x)\,dx = 1-\alpha$$

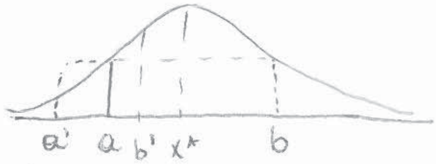
Quindi se prendo un intervallo + corto, tutto a sx di a mi dà $\int_{a'}^{b'} f_x(x)\,dx < 1-\alpha$

Quindi non sono nella classe di intervalli che soddisfano a)

(84) ii) $(b'-a') < b-a$

$b' > a$

Quindi $a' \leq a < b' < b$



$$\int_{a'}^{b'} f_X(x)dx = \int_a^b f_X(x)dx + \int_{a'}^{a} f_X(x)dx - \int_{b'}^{b} f_X(x)dx$$

$$= (1-\alpha) + \underbrace{\int_{a'}^{a} f_X(x)dx - \int_{b'}^{b} f_X(x)dx}_{\text{voglio mostrare che è } <0.}$$

- $\int_{a'}^{a} f_X(x)dx \leq f_X(a)[a-a']$
- $\int_{b'}^{b} f_X(x)dx \geq f_X(b)[b-b']$ (vale sia che b' sia prima che dopo la moda)

Allora

$$\int_{a'}^{a} f_X(x)dx - \int_{b'}^{b} f_X(x)dx \leq f_X(a)[a-a'] - f_X(b)[b-b']$$

ma $f_X(a) = f_X(b)$ per b)

$$= \underbrace{f_X(a)}_{>0}\left[\underbrace{(b'-a')-(b-a)}_{}\right]$$

ma io ho detto che $[a',b']$ ha lunghezza minore di $[a,b]$ quindi qst è <0 in totale.

Quindi di fatto ho preso sì un intervallo + corto ma ne ho cambiato la confidenza.

Ma noi cerchiamo l'IC di lunghezza minima.

$Z \sim \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ cerco a,b: $P(a \leq Z \leq b) = 1-\alpha$

Essendo unimodale il più corto è quello per cui $f_Z(a) = f_Z(b)$ con $a<0<b$

IC era $\left[\bar{X}_n - b\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n - a\frac{\sigma}{\sqrt{n}}\right]$

la cui lunghezza è proporzionale a $b-a$.

Allora scelgo i quantili simmetrici

Esempio

$X \sim \Gamma(n, \frac{1}{\beta})$

$Q = \frac{X}{\beta} \sim \Gamma(n, 1)$ è pivot ed è unimodale. per TCL per $n \to \infty$ va a una Gaussiana

$P(a < Q < b) = 1 - \alpha$

$\min(b - a) \Rightarrow$ poiché è unimodale: $f_X(b) = f_X(a)$ e $a < x^* < b$.

Ma il mio IC è

$[a \le Q \le b] \Leftrightarrow [\frac{Q}{b} \le \beta \le \frac{Q}{a}]$

Quindi

$L(IC) = Q(\frac{1}{a} - \frac{1}{b})$ Quindi devo minimizzare $\frac{1}{a} - \frac{1}{b}$.

Il teorema mi risolve il problema quando la lunghezza è proporzionale a $b - a$.

Quindi cosa faccio?

Se ho

$P(a \le Q \le b) = 1 - \alpha \Rightarrow b = b(a)$ (una volta scelto a, b è una sua funzione).

Minimizzare la lunghezza che è $[\frac{1}{a} - \frac{1}{b}] = [\frac{1}{a} - \frac{1}{b(a)}]$

con vincolo

$$\int_a^{b(a)} f_Q(x)dx = 1 - \alpha$$

Quindi

$$L = [\frac{1}{a} - \frac{1}{b(a)}]$$

$b' = \frac{db}{da}$

per trovare il minimo

$$\frac{\partial L}{\partial a} = 0 \Rightarrow 0 = -\frac{1}{a^2} + \frac{b'}{b^2(a)} \Rightarrow \frac{db}{da} = \frac{b^2}{a^2}$$

derivo il vincolo

$$\frac{d}{da}\left[\int_a^b f_Q(x)\,dx\right] = 0 = f_Q(b)b' - f_Q(a)$$

xk $1-\alpha$ non dipende da a.

Derivando il vincolo:

$$0 = f_Q(b)b' - f_Q(a) = f_Q(b)\underbrace{\frac{b^2}{a^2}}_{\text{da prima}} - f_Q(a)$$

$f_Q(b)b^2 = f_Q(a)a^2$ ← i punti che minimizzano sono quelli che soddisfano qst.

Nel caso la lunghezza fosse stata $b - a$ sarei tornata ad avere $f(b) = f(a)$

(86) $X_1 \dots X_n \sim U[0,\theta]$

$\frac{X_{(n)}}{\theta} \sim T$ con $F_T(t) = t^n$ per $0 \le t \le 1$.

Allona posso usare qst come quantità pivot:

$$[a \le T \le b]$$

con $b^n - a^n = 1-\alpha$

↑ l'area è fdR in b - fdR in a.

L'IC è $\left[a \le \frac{X_{(n)}}{\theta} \le b\right] \Rightarrow \frac{X_{(n)}}{b} \le \theta \le \frac{X_{(n)}}{a}$

IC tc $f_T(a)a^2 = f_T(b)b^2$

$f_T = n t^{n-1}$

Devo imporre che $n a^{n-1} a^2 = n b^{n-1} b^2$

ma non è possibile xk a deve essere < b.

$1-\alpha = b^n - a^n$ $\frac{db}{da}$

$0 = n b^{n-1} b' - n a^{n-1}$

Quindi

$$b' = \frac{a^{n-1}}{b^{n-1}}$$

Quindi

$$\frac{\partial L}{\partial a} = \left[-\frac{1}{a^2} + \frac{b'}{b^2}\right] = \left[-\frac{1}{a^2} + \left(\frac{a^{n-1}}{b^{n-1}}\right)\frac{1}{b^2}\right] = \frac{-b^{n+1}+a^{n+1}}{a^2 b^{n+1}}$$

è sempre <0 -> <0

tanto + decresce a

tanto più decresce la lunghezza

Se la lunghezza decresce

allora devo far crescere a il + possibile.

Ma allora anche b.

Ma b non può andare più su di 1.

Devo scegliere il + grande a $\Rightarrow b = 1$.

Ma allora ho $1-\alpha = b^n - a^n = 1 - a^n$

$$\Downarrow$$
$$a = \sqrt[n]{\alpha}$$

Quindi IC di

lunghezza minima

sono $\left[X_{(n)} \le \theta \le \frac{X_{(n)}}{\sqrt[n]{\alpha}}\right]$

STATISTICA ASINTOTICA

n (ampiezza o dimensione del campione) $\to +\infty$
Quando n è sufficientemente grande da dire che siamo a convergenza all'∞?

$X_n \xrightarrow{qc} X$ ↓ $\qquad\qquad X_n \xrightarrow{L^2} X$
$X_n \xrightarrow{P} X$ ↓ ↥ se cost.
$X_n \xrightarrow{\mathcal{L}} X$ (convergenza funzioni caratteristiche e anche delle funzioni generatrici dei momenti

$X_1 \dots X_n$ iid $\mathcal{L}(X_i)$
Se ho una statistica $T = T(\vec{x}) = \frac{1}{n} \sum X_i$
Per $n \to \infty$
Considereremo una successione di stimatori: la T la vedremo come T_n, xk dipende dall'ampiezza del campione

$$T_n = \left\{ X_1, \frac{X_1+X_2}{2}, \frac{X_1+X_2+X_3}{3}, \dots \right\}$$

↑ lo chiameremo stimatore, ma di fatto si intende una successione che dipende da n

Vale la
LGN: $\frac{\sum X_i}{n} \xrightarrow[P]{qc} E[X_i]$

e TCL: $\frac{\sqrt{n}\,(\bar{X}_n - E[X])}{\sqrt{Var[X]}} \xrightarrow{\mathcal{L}} N(0,1)$

Def una successione di stimatori $W_n = W_n(X_1 \dots X_n)$ è detta **SUCCESSIONE CONSISTENTE** di STIMATORI per θ SE $\forall \varepsilon > 0$

$$\lim_{n \to +\infty} P(|W_n - \theta| < \varepsilon) = 1 \quad \equiv W_n \xrightarrow{P} \theta$$

(CONSISTENZA FORTE: $W_n \xrightarrow{qc} \theta$
CONSISTENZA in L^2: $W_n \xrightarrow{L^2} \theta$
DEBOLMENTE CONSISTENTE: $W_n \xrightarrow{\mathcal{L}} \theta$)

le implicazioni sono le solite delle convergenze.

ESEMPIO
$X_1 \dots X_n$ iid $N(\theta, 1)$
$\bar{X}_n$ studiamone la consistenza.
$\bar{X}_n$ è consistente per θ?

• LGN: $\frac{\sum x_i}{n} \xrightarrow{qc} \vartheta$ Quindi abbiamo forte consistenza $\Rightarrow$ abbiamo consistenza

(e se ho convergenza qc. anche $g\left(\frac{\sum x_i}{n}\right) \xrightarrow{qc} g(\vartheta)$)

• Potevo anche osservare che

$$MSE(\bar{X}_n) = E[(\bar{X}_n-\vartheta)^2] = Var[\bar{X}_n] + \underbrace{bias^2}_{=0} =$$

$$= \frac{\sigma^2}{n} \to 0 \Rightarrow \text{Ho Anche consistenza in } L^2$$

$\Downarrow$

consistenza in probabilità

• Oppure potevo dire:

$\forall \varepsilon > 0$

$$P(|\bar{X}_n-\vartheta|<\varepsilon) = \int_{\vartheta-\varepsilon}^{\vartheta+\varepsilon} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left(-\frac{n}{2}(x-\vartheta)^2\right) dx$$

$\downarrow$

$\bar{X}_n \sim N(\vartheta, \frac{\overset{1}{\sigma^2}}{n})$

chiamiamo $Y = X - \vartheta$

$$= \int_{-\varepsilon}^{\varepsilon} \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{n}{2}y^2\right) dy = \quad t = y\sqrt{n}$$

$$= \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \frac{\cancel{\sqrt{n}}}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \frac{dt}{\cancel{\sqrt{n}}} =$$

↑ densità di una gaussiana standard

$$= \Phi(\varepsilon\sqrt{n}) - \Phi(-\varepsilon\sqrt{n})$$

Quindi

$$\lim_{n\to+\infty} P(|X_n-\vartheta|<\varepsilon) = \lim_{n\to+\infty}\left[\Phi(\varepsilon\sqrt{n}) - \Phi(-\varepsilon\sqrt{n})\right] = 1$$

$\Rightarrow$ la media campionaria è fortemente consistente, è consistente ed è consistente in L^2.

ÆEEE

Se so

$MSE_\theta(W_n) \to 0 \overset{?}{\Rightarrow} W_n$ è consistente per θ?

Convergenza in L^2 mi implica quella in probabilità?

PER LA DISUGUAGLIANZA di CHEBICHEV

$$P(|W_n - \theta| \geq \varepsilon) \leq \frac{E[|W_n - \theta|^2]}{\varepsilon^2}$$

Vale $\forall n \Rightarrow$ vale anche al limite

$$\lim P(|W_n - \theta| \geq \varepsilon) \leq \lim_{n \to +\infty} \frac{E[|W_n - \theta|^2]}{\varepsilon^2}$$

$\downarrow$ 0 $\Leftarrow$ $\downarrow$ 0 se l'MSE $\to 0$ qst

se l'MSE $\to 0$ tende a 0 alla dx e quindi alla sx.

(vale senza mettere il quadrato - disuguaglianza di Markov)

Oss

$$MSE_\theta(W_n) = \underbrace{Var[W_n]}_{\to 0} + \underbrace{(E[W_n] - \theta)^2}_{\to 0}$$

$\leftarrow$ se riesco a mostrare che entrambe la varianza $\to 0$ e media $\to 0$

allora è consistente

Def Se $E[W_n] \to \theta$ (sommomeni)

si dice che W_n è asintoticamente non distorto

$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x}_n)^2$ è tc $E[S^2] = \sigma^2$

$S_0^2 = \frac{1}{n} \sum (x_i - \bar{x}_n)^2$ è tc $E[S_0^2] = \left(\frac{n-1}{n}\right)\sigma^2$

è distorto,

ma per $n \to +\infty$ $\to \sigma^2$

è distorta ma è asintoticamente non distorta.

Crescendo con n

la distorsione si riduce.

90) MA POSSIAMO DIRE QUALCOSA SULLA LEGGE ASINTOTICA W_n?

Def W_n successione di stimatori

se

$$K_n \operatorname{Var}(W_n) \xrightarrow[n\to\infty]{} \tau^2 < +\infty$$

τ^2 viene chiamata varianza limite

Se ho $\bar{X}_n$ ho $\operatorname{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$ se scelgo $K_n = n$

ho che la varianza limite è σ^2

Def W_n successione di stimatori tali che

$$K_n(W_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

Allora σ^2 è detta VARIANZA ASINTOTICA

Tipicamente si sceglie $K_n = \sqrt{n}$ xk centra con TCL

Se ho convergenza in legge posso guardare stimatori ≠ valutando la loro varianza asintotica, voglio che la varianza asintotica sia piccola

[LGN: $\bar{X}_n \xrightarrow{q.c.} \mu$

TLC: $\sqrt{n}\,\bar{X}_n \to N(0,1)$ incredible!]

ESEMPIO (Modelli gerarchici)

$$Y_n \mid W_n = w_n \sim N(0, w_n + (1-w_n)\sigma_n^2)$$

dove

$W_n \sim Be(p_n)$ ← lancio una moneta $\forall n$ con una certa probabilità di successo

$$\operatorname{Var}[Y_n] = \mathbb{E}[\operatorname{Var}[Y_n \mid W_n]] + \operatorname{Var}[E[Y_n \mid W_n]] =$$

$$= \mathbb{E}[W_n + (1 - W_n)\sigma_n^2] + 0 = p_n + (1-p_n)\sigma_n^2$$

$P(Y_n \le a)$ (funzione di ripartizione di Y_n)

$$= P(Y_n \le a \mid W_n = 1) \cdot P(W_n = 1) + P(Y_n \le a \mid W_n = 0) P(W_n = 0)$$

$$= P(Z \le a)\, p_n + P\left(Z \le \frac{a}{\sigma_n}\right)(1-p_n)$$

$Y_n \mid W_n = 1$ diventa gaussiana standard

$Y_n \mid W_n = 0 \sim N(0, \sigma_n^2)$ quindi devo standardizzare.

Prendiamo una successione $p_n \to 1$

e $\sigma_n^2 \to +\infty$

ma tc $(1-p_n)\sigma^2 \to +\infty$

(per es: $p_n = 1 - \frac{1}{\sqrt{n}}$, $\sigma_n = n$)

Quanto vale quindi la varianza limite?

$$\lim (p_n + (1-p_n)\sigma_n^2) = +\infty$$

Quanto vale la varianza asintotica?

$$P(Y_n \leq a) \to \phi(a) \Rightarrow Y_n \xrightarrow{\mathcal{L}} N(0,1)$$

Non ci sono a priori regole per cui la varianza asintotica sia = alla varianza limite.

MA QUANTO VALE LA VARIANZA ASINTOTICA?

Def una SUCCESSIONE di STIMATORI W_n È ASINTOTICAMENTE EFFICIENTE per $\tau(\theta)$ SE

$$\sqrt{n}\,[W_n - \tau(\theta)] \xrightarrow{\mathcal{L}} N(0, v(\theta))$$

$$\text{con } v(\theta) = \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right]} = \frac{[\tau'(\theta)]^2}{I_1(\theta)}$$

La varianza asintotica raggiunge il limite di CR

informazione di Fisher

$W_n - \theta$ è $AN(0,\sigma^2)$

se

$\sqrt{n}(W_n - \theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2)$ $\Delta =$ ASINTOTICAMENTE NORMALE.

Per n grande

$W_n - \theta \simeq N\left(0, \frac{\sigma^2}{n}\right)$ non posso dire che converge xk dipenderebbe da n.

$\Rightarrow$ Per n grande

$$W_n - \theta \simeq N\Big(0, \underbrace{\frac{v(\theta)}{n}}\Big)$$

$$\frac{v(\theta)}{n} = \frac{[\tau'(\theta)]^2}{n I_1(\theta)} = \frac{[\tau'(\theta)]^2}{I_n(\theta)}$$

Uno stimatore è asintoticamente efficiente quando

- è ASINTOTICAMENTE normale
- varianza asintotica è il limite di CR

TEOREMA: EFFICIENZA ASINTOTICA degli MLE

$X_1 \dots X_n \sim f(x,\theta)$

$\hat{\theta}_{MLE}$ è lo stimatore ML di θ

$$\Rightarrow \sqrt{n}\,[\tau(\hat{\theta}_{MLE}) - \tau(\theta)] \xrightarrow{\mathcal{L}} N(0, v(\theta))$$

la varianza asintotica è la + piccola possibile xk è limite CR.

Quindi gli stimatori di massima verosimiglianza sono asintoticamente efficienti

92 Nascosta lì dentro c'è anche la consistenza:

OSS L'asintotica normalità ⇒ consistenza:

dim

$$\frac{\sqrt{n}\,(W_n - \tau(\theta))}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

σ ← VARIANZA ASINTOTICA

Voglio vedere se $W_n \xrightarrow{P} \tau(\theta)$

$$W_n - \tau(\theta) = \underbrace{\frac{\sigma}{\sqrt{n}}}_{\downarrow\, 0} \Bigg(\underbrace{\frac{\sqrt{n}}{\sigma}(W_n - \tau(\theta))}_{\downarrow \mathcal{L}\ N(0,1)} \Bigg) \xrightarrow[\mathcal{L}]{\text{per teo di Slutzky}} 0$$

converge in legge al prodotto.

Dato che il limite è costante la convergenza in legge a una cost. implica la convergenza in probabilità

⇓

$$W_n - \tau(\theta) \xrightarrow{P} 0$$

■

A.N: $\sqrt{n}(W_n - \theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2)$

⇓

Consistenza

A. Efficienza = A.N + σ^2 → limite CR.

Stimatori MLE sono AE

Def ASINTOTICA EFFICIENZA RELATIVA

W_n e V_n tc

$$\sqrt{n}\,[W_n - \tau(\theta)] \xrightarrow{\mathcal{L}} N(0, \sigma^2_w)$$

e

$$\sqrt{n}\,[V_n - \tau(\theta)] \xrightarrow{\mathcal{L}} N(0, \sigma^2_v)$$

ARE = ~~Asymptotic~~ Relative Efficiency ASYMPTOTIC

di V_n rispetto a W_n è

$$ARE(V_n, W_n) = \frac{\sigma^2_w}{\sigma^2_v}$$

se è > 1 V_n è + efficiente di W_n

Scelgo quello con varianza asintotica minima

ESEMPIO

$X_1 \dots X_n \sim P(\lambda)$ $\qquad P(X=0) = e^{-\lambda} = \tau(\lambda)$

1) Possibile stimatore è

$$\hat{\tau} = \frac{1}{n} \Sigma Y_i = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(X_i=0)} \qquad Y_i \sim Be(\tau(\lambda))$$

2) Altro possibile stimatore

$$\hat{\tau}_1 = e^{-\bar{X}_n} \quad \text{è MLE}$$

x il principio di varianza, xk $\hat{\lambda}_{MLE} = \bar{X}_n$

Mi aspetto vinca il secondo

$$I_1(x) = \mathbb{E}\left[\left(\frac{\partial}{\partial\lambda} \log f(x,\lambda)\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial\lambda}\left(-\lambda + x \log\lambda\right)\right)^2\right] = \mathbb{E}\left[\frac{1}{\lambda^2}[x-\lambda]^2\right] = \frac{1}{\lambda}$$

Stimatore 1) è una media campionaria ⇒ posso usare TCL.

$$\sqrt{n}\,(\hat{\tau} - e^{-\lambda}) \xrightarrow{\mathcal{L}} N(0, e^{-\lambda}(1-e^{-\lambda}))$$

varianza della singola bernoulli

Stimatore 2) uso teo sugli MLE:

$$\sqrt{n}(\hat{\tau}_1 - e^{-\lambda}) \xrightarrow{\mathcal{L}} N(0, \underbrace{e^{-2\lambda}\cdot\lambda}_{\text{gli calcolato con } \frac{(\tau'(\theta))^2}{I_1(\theta)}}) \qquad xk\ \hat{\lambda}_{MLE} = \bar{X}_n$$

$$ARE[\hat{\tau}, \hat{\tau}_1] = \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1-e^{-\lambda})} = \frac{\lambda}{e^{\lambda}-1} \quad \text{è decrescente}$$

13 maggio 2015

Successione di stimatori X_n AN e vogliamo capire se l'asintotica normalità si trasmette se prendo $g(X_n)$

METODO DELTA

$$Y_n : \quad \sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

Prendiamo g con $g'(\theta) \neq 0$. g differenziabile con $g'(\theta) \neq 0$.

$$\Rightarrow \sqrt{n}\,[g(Y_n) - g(\theta)] \xrightarrow{\mathcal{L}} N(0, \sigma^2 g'(\theta)^2)$$

 Si basa su uno sviluppo di Taylor di g

$$g(x) = g(\theta) + g'(\theta)(x-\theta) + o(x-\theta)$$

$$\sqrt{n}\,(Y_n - \theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2)$$

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n-\theta) + o(Y_n - \theta)$$

Asintotica normalità
⇓
consistenza
⇓
$\mathbb{E}[Y_n-\theta] = 0$

$$\mathbb{E}[g(Y_n)] \simeq g(\theta) \leftarrow$$

$$Var[g(Y_n)] = \mathbb{E}[(g(Y_n) - g(\theta))^2]$$
$$\simeq \mathbb{E}[(g'(\theta)(Y_n-\theta))^2] = g'^2(\theta) \cdot \sigma^2$$

↑ Varianza asintotica

Totto questo discorso cade se $g'(\theta)=0$.

Metodo Delta del Secondo Ordine

$$Y_n : \sqrt{n}\,(Y_n-\theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2)$$

$$g \in C^2$$
$$g'(\theta)=0$$
$$g''(\theta) \neq 0$$

$$\Rightarrow n\,[g(Y_n)-g(\theta)] \xrightarrow{\mathcal{L}} \sigma^2 \frac{g''(\theta)}{2} \chi^2_1$$

Ancora si basa sullo sviluppo di Taylor.

$$g(x) = g(\theta) + \cancel{g'(\theta)(x-\theta)} + \frac{1}{2} g''(\theta)(x-\theta)^2 + o(x-\theta)^2$$

$$n\,[g(Y_n)-g(\theta)] = \frac{1}{2} g''(\theta)\, n\,(Y_n-\theta)^2 + \dots$$

io so $\sqrt{n}\,\frac{Y_n-\theta}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$

si può dimostrare

$$n\,\frac{[Y_n-\theta]^2}{\sigma^2} \xrightarrow{\mathcal{L}} \chi^2_1$$ (xké è il quadrato di una normale standard)

$$n\,[g(Y_n) - g(\theta)] = \frac{\sigma^2}{2} g''(\theta)\, \underbrace{n\,\frac{(Y_n-\theta)^2}{\sigma^2}}_{\chi^2_1}$$

ESEMPI

- $X_1 \dots X_n \sim Be(p)$ $\quad \bar{X}_n$

Via TLC sappiamo $\sqrt{n}(\bar{X}_n - p) \xrightarrow{\mathcal{L}} N(0, p(1-p))$

un parametro importante per la Bernoulli era l'ODDS: $\frac{p}{1-p}$
(vale 1 se la moneta è equilibrata)

Come valutare l'odds?
La prima cosa che mi viene in mente è usare $\frac{\bar{X}_n}{1-\bar{X}_n}$!
xk $\bar{X}_n$ è stimatore di massima verosimiglianza e vale il principio di invarianza.

$g(x) = \frac{x}{1-x}$

$g'(x) = \frac{1-x+x}{(1-x)^2} = \frac{1}{(1-x)^2} \neq 0$

Possiamo usare il metodo Δ primo.

$g(\bar{X}_n) = \frac{\bar{X}_n}{1-\bar{X}_n}$

<u>Delta 1</u>

$$\sqrt{n}\left[\frac{\bar{X}_n}{1-\bar{X}_n} - \frac{p}{1-p}\right] \xrightarrow{\mathcal{L}} N\left(0, p(1-p)\frac{1}{(1-p)^4}\right) = N\left(0, \frac{p}{(1-p)^3}\right)$$

Se non avessimo usato il metodo delta:

$\frac{\bar{X}_n}{|1-\bar{X}_n|}$ è MLE di $\frac{p}{1-p} = \tau(p)$

$$\sqrt{n}\left[\frac{\bar{X}_n}{1-\bar{X}_n} - \frac{p}{1-p}\right] \xrightarrow{\mathcal{L}} N(0, v(p)) = N\left(0, \frac{\tau'^2(p)}{I(p)}\right) =$$

limite di Cramer Rao $\frac{\tau'(p)}{I_1(p)}$ $\quad = N\left(0, \frac{1}{(1-p)^4} \cdot p(1-p)\right)$ ok.

facendo i calcoli che seguono:

Allora devo calcolare l'informazione di Fisher x il campione di ampiezza 1.

$$I(p) = \mathbb{E}\left[\left(\frac{\partial}{\partial p}\log f(x,p)\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial p}\left(x\log p + (1-x)\log(1-p)\right)\right)^2\right] =$$

$f(x,p) = p^x(1-p)^{1-x}\mathbb{1}_{\{0,1\}}(x)$

$$= \mathbb{E}\left[\left(\frac{x}{p} - \frac{1-x}{1-p}\right)^2\right] = \mathbb{E}\left[\left(\frac{x - \cancel{px} - p + \cancel{px}}{p(1-p)}\right)^2\right] = \frac{1}{p^2(1-p)^2}\mathbb{E}\left[(x-p)^2\right]$$

$$= \frac{Var[x]}{p^2(1-p)^2} = \frac{1}{p(1-p)}$$

96 Quindi ho

$$\sqrt{n}\left[\frac{\bar{X}_n}{1-\bar{X}_n} - \frac{p}{1-p}\right] \xrightarrow{\mathcal{L}} N\left(0; \frac{p}{(1-p)^3}\right)$$

OSS

Noi abbiamo

$$\underbrace{\frac{\left[\frac{\bar{X}_n}{1-\bar{X}_n} - \frac{p}{1-p}\right]\sqrt{n}}{\sqrt{\frac{p}{(1-p)^3}}}}_{\downarrow \mathcal{L} \; N(0,1)} \; \underbrace{\frac{\sqrt{\frac{p}{(1-p)^3}}}{\sqrt{\frac{\bar{X}_n}{(1-\bar{X}_n)^3}}}}_{qc \downarrow P \; 1} \xrightarrow[\text{SLUTZKY}]{\mathcal{L}} N(0,1)$$

dalla legge forte dei Grandi Numeri $\bar{X}_n \xrightarrow{qc} p$ quindi $f(\bar{X}_n) \xrightarrow{qc} f(p)$

converge qc quindi in probabilità

Quindi di fatto a qst punto butto via i due $\sqrt{\frac{p}{(1-p)^3}}$ e ho

$$\frac{\sqrt{n}\left[\frac{\bar{X}_n}{1-\bar{X}_n} - \frac{p}{1-p}\right]}{\sqrt{\frac{\bar{X}_n}{(1-\bar{X}_n)^3}}} \xrightarrow{\mathcal{L}} N(0,1)$$

Costruire un IC asintotico per $\frac{p}{1-p}$ mi permette allora di avere una gaussiana

$\Downarrow$

$$IC(ODDS) = \frac{\bar{X}_n}{1-\bar{X}_n} \pm \frac{\sqrt{\frac{\bar{X}_n}{1-\bar{X}_n}}}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$$

posso farlo xk adesso al limite qst è una gaussiana standard.

- $X_1 \dots X_n \sim Be(p)$ Ora voglio stimare la varianza $p(1-p)$

$g(x) = x(1-x)$ si annulla la derivata prima in $\frac{1}{2}$.

$g'(x) = 1-2x$

- $\forall p \neq \frac{1}{2}$ posso dire

$$\sqrt{n}\left[\bar{X}_n(1-\bar{X}_n) - p(1-p)\right] \xrightarrow{\mathcal{L}} N\Big(0, p(1-p)\underbrace{(1-2p)^2}_{g'(p)^2}\Big)$$

- $p = \frac{1}{2}$

mi serve $g''(x) = -2$

$$n\left[\underbrace{\bar{X}_n(1-\bar{X}_n)}\ - \overbrace{p(1-p)}^{1/4}\right] \xrightarrow{\mathcal{L}} \overset{g''}{\overset{\downarrow}{-2}} \cdot \frac{1}{2} \cdot \overset{\sigma^2}{\overset{\downarrow}{\frac{1}{4}}} \chi^2_1$$

$$= -\frac{1}{4}\chi^2_1$$

Guardo se asintoticamente
mi va come una normale o
come una χ^2
e capisco in quale dei
due casi

torna che sia <0
xk $\bar{X}_n$ è compreso tra 0 e 1.
A sx ho una quantità
$< \frac{1}{4}$ $\Rightarrow$ converge a qlcs
di <0.

$IC(p)$

$$\left[\bar{X}_n \pm \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]$$

noi sappiamo

$$\sqrt{n}\,\frac{(\bar{X}-p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} N(0,1)$$

diciamo

$$\sqrt{n}\,\frac{(\bar{X}_n-p)}{\sqrt{p(1-p)}}\,\frac{\sqrt{p(1-p)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathcal{L}} N(0,1)$$

ASINTOTICAMENTE posso quindi avere quell'IC,
ma solo ASINTOTICAMENTE.

UBRIACHELLA ♡

(98)

TEOREMA

$H_0 : \theta = \theta_0$ vs $H_1\ \theta \neq \theta_0$

$X_1 \dots X_n \sim f(x, \theta)$

$\hat{\theta}$ l'MLE (lo assumiamo interno)

Sotto H_0 $\quad -2 \log \underbrace{\lambda(\vec{x})}_{} \xrightarrow{\mathcal{L}} \chi^2_1$

la statistica della LRT è sempre tra 0 e 1. Quindi i segni sono giusti.

dim

$$-2\log\lambda(\vec{x}) = -2\log\left(\frac{L(\theta_0,\vec{x})}{L(\hat{\theta},\vec{x})}\right) = -2\ell(\theta_0,\vec{x}) + 2\ell(\hat{\theta},\vec{x})$$

il sup su tutto è raggiunto in $\hat{\theta}$

OSS

$$\ell(\theta_0,\vec{x}) = \ell(\hat{\theta},\vec{x}) + \cancel{\ell'(\hat{\theta},\vec{x})(\hat{\theta}-\theta_0)} + \ell''(\hat{\theta},\vec{x})\frac{(\hat{\theta}-\theta_0)^2}{2} + \dots$$

ma $\hat{\theta}$ è l'MLE, quindi $\ell'(\hat{\theta},\vec{x}) = 0$

$$\Rightarrow -2\log\lambda(\vec{x}) = -2\ell(\theta_0,\vec{x}) + 2\ell(\theta_0,\vec{x}) - \underbrace{\ell''(\hat{\theta},\vec{x})(\theta_0-\hat{\theta})^2}_{\downarrow\ I(\theta_0)}$$

Sotto opportune hp di regolarità:

$$\boxed{\frac{1}{n}\left(-\ell''(\hat{\theta},\vec{x})\right) \xrightarrow{P} I(\theta_0)}$$

ma noi balziamo la dim.

$$-2\log\lambda(\vec{x}) = -\ell''(\hat{\theta},\vec{x})(\hat{\theta}-\theta_0)^2 = \underbrace{\frac{-\ell''(\hat{\theta},\vec{x})}{n I(\theta_0)}}_{\downarrow P\ +1} \underbrace{\left(\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sqrt{1/I(\theta_0)}}\right)^2}_{\downarrow\ (N(0,1))^2}$$

Asintotica Normalità stimatori MLE

per Slutzky $\xrightarrow{\mathcal{L}} + \chi^2_1$

■

ROBUSTEZZA

$\bar{X}_n$ per il modello gaussiano è molto bello.

$X_1 \dots X_n \sim N(\mu, \sigma^2)$ modello parametrico gaussiano.

Se il modello è contaminato, ossia ho una piccola parte di dati che non è gaussiano non ho più veroi ottimalità di $\bar{X}_n$

Ma vorrei che $\bar{X}_n$ restasse bello anche se ho piccole perturbazioni o comunque non voglio catastrofi.

La media campionaria è robusta per un cavolo!! Basta un dato per farla sbarellare.

Vorrei misurare la robustezza.

$X_1 \dots X_n \sim N(\mu, \sigma^2)$
$\bar{X}_n$ best

$X_i = (1-Y)W + YV_i$ "Lancio una Bernoulli" cit.

$W \sim N(\mu, \sigma^2)$

$V_i \sim f$ $\quad E[V] = \theta,$ $\quad Var[V] = \tau^2$

$Y \sim Be(\delta)$ con δ piccolo

$$X_i = \begin{cases} N(\mu, \sigma^2) & \text{prob } 1-\delta \\ f & \text{prob } \delta \end{cases}$$

$$Var[X_i] = Var[E[X_i|Y]] + E[Var[X_i|Y]]$$

$$E[Var[X_i|Y]] = \sigma^2 \underbrace{(1-\delta)}_{Var(1-Y)} + \tau^2 \cdot \underbrace{\delta}_{Var(YV_i)}$$

$$Var[E[X_i|Y]] = Var[(1-Y)\mu + Y\theta] = Var[Y(\theta-\mu)] = \delta(1-\delta)(\theta-\mu)^2$$

$$\Rightarrow Var[X_i] = (1-\delta)\sigma^2 + \delta\tau^2 + \delta(1-\delta)(\theta-\mu)^2$$

Quindi per il modello contaminato

$$Var[\bar{X}_n] = (1-\delta)\frac{\sigma^2}{2} + \delta\frac{\tau^2}{n} + \delta(1-\delta)(\theta-\mu)^2$$

tiene conto della differenza tra le 2 medie

Per piccole deviazioni ($\tau^2 \simeq \sigma^2$, $\delta \simeq \mu$) ok.
Ma se ho una Cauchy, che non ha varianza finita, questa cosa mi esplode.

Brutta $\bar{X}_n$, no robusta!

(100) Per misurare la robustezza

Def BREAKDOWN POINT

$X_{(1)} \dots X_{(n)}$ T_n: statistica

T_n ha valore di breakdown b $0 \le b \le 1$ se $\forall \varepsilon > 0$

$$\lim_{X_{(\{1-b\}n)} \to \infty} T_n < +\infty \qquad \lim_{X_{(\{1-(b+\varepsilon)\}n)} \to \infty} T_n \to \infty$$

$\bar{X}_n$ ha $b = 0$, xk basta che un punto vada a ∞ perché tutto vada a ∞

Mediana ha $b = \frac{1}{2}$, x avere che sia ∞ devo mandare poco più della metà dei miei dati a ∞

[concettualmente by leti:
b mi dice la "frazione di dati" che devono andare all'∞ perché T_n vada all'∞]

19 maggio

MODELLI LINEARI

Y = risposta (v.a).

vogliamo modellizzare la ~~risposta~~ dipendenza di Y dai valori di r quantità reale $(z_1 \dots z_r)$ predittori

Si immagina che siano aleatorie le risposte e i predittori.
Ma in realtà noi penseremo che i predittori siano noti, quantità osservabili (metratura casa, distanza dal centro...) e vogliamo vedere come cambia la risposta al variare dei predittori.

<u>Modello</u>

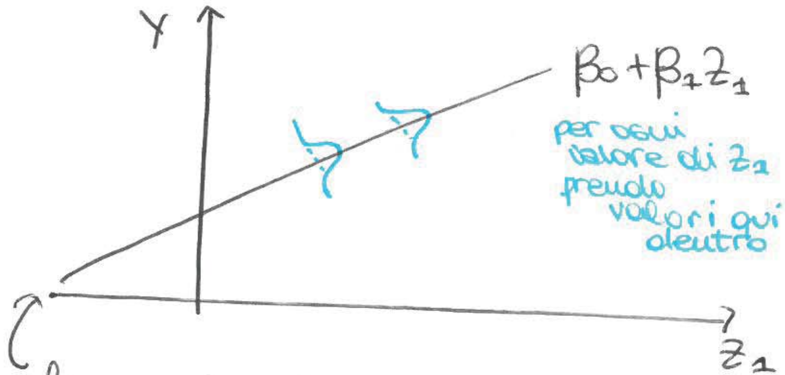
$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_r z_r + \varepsilon$$

$\vec{\beta} = (\beta_0, \beta_1 \dots \beta_r)^T$ è un vettore di parametri incogniti

ε è una v.a., $\mathbb{E}[\varepsilon] = 0$

$Var[\varepsilon] = \sigma^2$ costante qualunque sia il predittore

<u>Penso al caso $r = 1$</u>, un solo predittore.



immagino che valga un modello di qst genere: in realtà devo fare la diagnostica del modello, capire se ha senso che i dati vengano da un modello così

la media sta su qst retta, e poi ho una variabilità che ha media 0

Se leggiamo

$$\mathbb{E}[Y | z_1 \dots z_r] = \beta_0 + \beta_1 z_1 + \dots + \beta_r z_r$$

è comunque il valor medio della Y.

La parola lineare è riferita a "lineare in β",
io potrei pensare invece di avere

$$Y = \beta_0 + \beta_1 \sqrt{z_1} + \beta_2 \log z_2 + \varepsilon$$

ed è ancora lineare xk la linearità è nei β.

Non è lineare invece

$$Y = \beta_0 \sin(\beta_1 z_1)$$

02) Vogliamo stimare i parametri incogniti e capire se l'adattamento del modello è buono e poi fare previsioni.

Noi abbiamo n osservazioni delle mie variabili, abbiamo $Y_1 \dots Y_n$ che non sono più iid

MATRICE DISEGNO

$$\underset{(n \times r+1)}{Z} = \begin{bmatrix} 1 & z_{11} & \dots & z_{1r} \\ 1 & & & \\ \vdots & & & \\ 1 & z_{n1} & \dots & z_{nr} \end{bmatrix}$$

β_0 — metratura 1^ casa, 2^ casa ... n° casa — distanza dal centro 1^ casa, 2^ casa .... n° casa.

Modello che genera i dati

$$\vec{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = Z\vec{\beta} + \vec{\varepsilon} \qquad \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

v.a. a media 0 e varianza σ^2

I parametri incogniti sono 2:

$\vec{\beta} \in \mathbb{R}^{r+1}$

$\vec{\varepsilon}$, $\mathbb{E}[\vec{\varepsilon}] = \vec{0}$ $\quad Cov(\vec{\varepsilon}) = \sigma^2 \mathbb{1}$ ← n dimensionale

l'altro parametro incognito è σ^2.

$$\boxed{\text{Parametri incogniti: } \vec{\beta} \text{ e } \sigma^2}$$

Esempi: la regressione semplice

Come si stimano i parametri incogniti

$$\vec{\beta} = (\beta_0 \dots \beta_r)^T, \ \sigma^2$$

STIME AI MINIMI QUADRATI

$\hat{\beta}_{LS}$

Io ho $\vec{Y} \in \mathbb{R}^n$. Voglio trovare i β per cui la distanza tra Y e lo span lineare di Z sia minimo:

$$\vec{\beta} = \underset{\vec{b} \in \mathbb{R}^{n+1}}{\text{Arg Min}} \; (\vec{y} - Z\vec{b})^T (\vec{y} - Z\vec{b}) =$$

$$= \underset{\vec{b} \in \mathbb{R}^{n+1}}{\text{Arg Min}} \; d_e(\vec{y}, Z\vec{b})$$

distanza euclidea (somme dei quadrati)

il problema è ben posto

TEOREMA

Supponiamo che Z sia a rango pieno $(n+1) < n$

con 3 osservazioni non ha senso che voglio stimare 5 parametri

1) $\hat{\vec{\beta}} = (Z^T Z)^{-1} Z^T \vec{y}$ $\quad \hat{\vec{\beta}}$ è trasformazione lineare delle y.

2) Siano

$H = Z(Z^T Z^{-1}) Z^T$ hat matrix

È IL PROIETTORE. H mette il cappello alle y.

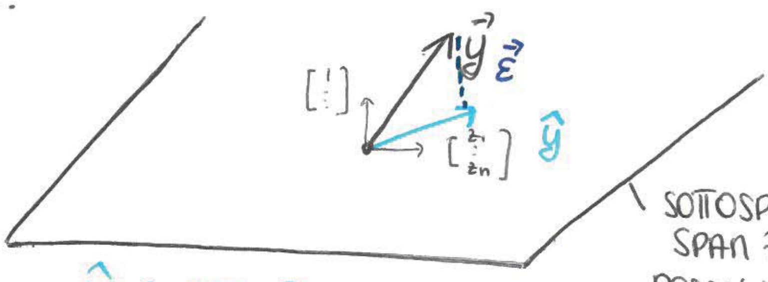
$\hat{\vec{y}} = Z\hat{\vec{\beta}} = H\vec{y}$ fittet values

mi soddisfano perfettamente il modello lineare senza $\vec{\varepsilon}$ by Mario

3) $\hat{\vec{\varepsilon}} = \vec{y} - \hat{\vec{y}} = (\mathbb{1} - H)\vec{y}$ residual

$\Rightarrow Z^T \hat{\vec{\varepsilon}} = 0$ e quindi $\hat{\vec{y}}^T \hat{\vec{\varepsilon}} = 0$ $\quad \hat{\vec{y}} \perp \hat{\vec{\varepsilon}}$

Prendiamo $n = 1$.
$\vec{y} \in \mathbb{R}^3$



SOTTOSPAZIO LINEARE, SPAN Z, $\mathcal{L}(Z)$

POSSIBILI combinazioni lineari delle colonne di Z

$\hat{\vec{y}} \in$ SPAN Z, è la proiezione di $\vec{y}$ su Span Z, che è quella che minimizza la distanza euclidea tra tutte le possibili combinazioni di Z.

(104) dim

Ho bisogno di una base ortonormale

$\mathcal{L}(Z)$ $\quad Z^T Z$ matrice $(r+1)\times(r+1)$
definita positiva
rango $(Z) = r+1$

$$Z^T Z = \sum_{i=1}^{r+1} \lambda_i \underbrace{\underset{\text{colonna}}{\vec{e}_i}\ \underset{\text{riga}}{\vec{e}_i^{\,T}}}_{(r+1)\times(r+1)}$$

λ_i sono gli autovalori
$\vec{e}_i$ sono gli autovettori di $Z^T Z$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{r+1} > 0$

Quindi posso invertire $Z^T Z$:

$$(Z^T Z)^{-1} = \sum_{i=1}^{r+1} \frac{1}{\lambda_i} \vec{e}_i \vec{e}_i^{\,T}$$ (tramite decomposizione spettrale)

Chiamo
$$\vec{q}_i = \frac{1}{\sqrt{\lambda_i}} Z \vec{e}_i \qquad i = 1 \dots r+1$$

$\vec{q}_i \in \mathcal{L}(Z)$ xk è combinazione delle colonne di Z.

Voglio sapere se i $\vec{q}_i$ sono base ortonormale

$$\rightarrow \vec{q}_i^{\,T} q_j = \frac{1}{\sqrt{\lambda_i \lambda_j}} (\vec{e}_i^{\,T} \underbrace{Z^T)(Z\vec{e}_j}_{\substack{\lambda_j \vec{e}_j \\ \text{xk } e_j \text{ era} \\ \text{autovettore}}}) = \frac{1}{\sqrt{\lambda_i \lambda_j}} \vec{e}_i^{\,T} \lambda_j \vec{e}_j =$$

$$= \sqrt{\frac{\lambda_j}{\lambda_i}} \vec{e}_i^{\,T} \vec{e}_j =$$

$$= \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Quindi i q_i sono base ortonormale per $\mathcal{L}(Z)$

Proiezione di y sullo spazio lineare di Z minimizza la distanza euclidea:

$$\underbrace{\hat{\vec{y}}}_{\text{proiezione ortogonale di } y \text{ su } \mathcal{L}(Z)} = \sum_{i=1}^{r+1} \vec{q}_i (\vec{q}_i^{\,T} \vec{y}) = \sum_{i=1}^{r+1} \frac{1}{\lambda_i} Z\vec{e}_i \left((Z\vec{e}_i)^T \vec{y}\right) =$$

$$= \sum_{i=1}^{r+1} \frac{1}{\lambda_i} Z\vec{e}_i \vec{e}_i^{\,T} Z^T \vec{y} =$$

$$= Z \left[\sum_{i=1}^{r+1} \frac{1}{\lambda_i} \vec{e}_i \vec{e}_i^{\,T}\right] Z^T \vec{y}.$$

$= Z\,(Z^T Z)^{-1} Z^T \vec{y}$ = proietta su $\mathcal{L}(Z)$

$= Z\,\hat{\vec{\beta}}$

stimatore ai minimi quadrati.

Lo stimatore dei $\hat{\vec{\beta}}$ è $(Z^T Z)^{-1} Z^T \vec{y}$

$\hat{\vec{y}} = Z\,\hat{\vec{\beta}}$

variabili aleatorie. Poi quando arrivano i dati sono uno i dati

$$\hat{\vec{\varepsilon}} = \vec{y} - \hat{\vec{y}} = [\mathbb{1} - H]\,\vec{y} = [1 - (Z)(Z^T Z)^{-1} Z^T]\,\vec{y}$$

ortogonale al sottospazio lineare generato da Z $\mathcal{L}(Z)$

$$Z^T \hat{\vec{\varepsilon}} = Z^T [\mathbb{1} - Z(Z^T Z)^{-1} Z^T]\,\vec{y} = Z^T \vec{y} - Z^T \vec{y} = 0$$

Quindi anche

$\hat{\vec{y}}\,\hat{\vec{\varepsilon}} = 0$ perché $\hat{\vec{y}} \in \mathcal{L}(Z)$, quindi se $\hat{\vec{\varepsilon}}$ è $\perp Z^T$ lo è anche a $\hat{\vec{\varepsilon}}$.

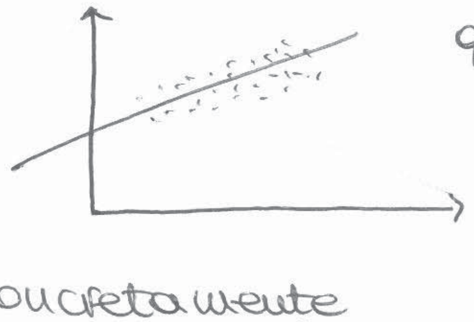
Una volta che ho gli stimatori ho anche i valori per le $\vec{y}$.

Supponiamo di avere un solo regressore, regressione lineare semplice.

$$\begin{bmatrix} 1 & z_{11} \\ \vdots & \vdots \\ 1 & z_{n1} \end{bmatrix} \qquad \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

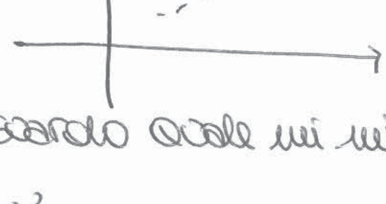
$$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$$

ho dati la cui media sta su una retta e poi ho un errore.



questo è quello che genererebbe il modello.

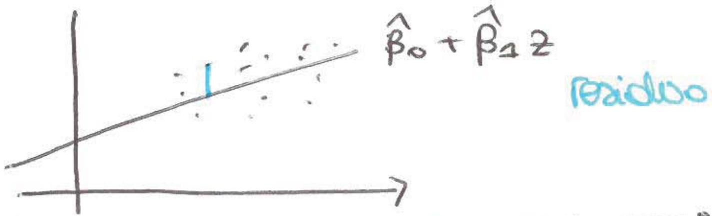
Ma β_0 e β_1 sono incogniti. Quindi ho solo i dati.



concretamente guardo tutte le possibili combinazioni lineari di Z, quindi guardo tutte le rette e guardo quale mi minimizza la distanza euclidea:

$$\sum_i (y_i - \beta_0 - \beta_1 z_{i1})^2$$

106 I valori previsti sono quelli che stanno sulla retta.
E i residui sono $y - \hat{y}$



Quando la matrice non ha rango pieno
$\mathcal{L}(Z)$ ha dim $K < r+1$.

$$Z^T Z = \sum_{i=1}^{r+1} \lambda_i \vec{e}_i \vec{e}_i^{\,T} \qquad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K > \lambda_{K+1} = 0 = \dots \lambda_{r+1}$$

Ho un po' di autovalori che sono 0, quindi non posso fare il giochino di prima.

Serve allora la Pseudoinversa

$$(Z Z^T)^{-1} = \sum_{i=1}^{K} \frac{1}{\lambda} \vec{e}_i \vec{e}_i^{\,T}$$ inversa generalizzata di $Z^T Z$

Prendo allora

$$\vec{q}_i = \frac{1}{\sqrt{\lambda_i}} Z \vec{e}_i \qquad i = 1 \dots K$$

Quindi

$$\hat{\beta} = (Z^T Z)^- Z^T \vec{y}$$

VALE ANCHE PER PREDITTORI CATEGORICI

Predittore Categorico:

y_i
z_i = categorie -

sto guardando il consumo al litro di benzina. Lo faccio x 4 categorie di motori.
Nel modello lineare modellizzo la media.
Voglio capire come cambia la media al cambiare del modello.
Per ognuno ho il consumo e la categoria - motore 1, 2, 3, 4.

Come posso costruire la mia matrice disegno?
Devo inventare un modo x dare l'"etichetta" del gruppo.

Si inseriscono variabili dummies

$$\begin{bmatrix} 1 & \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_1 & 0 & 0 & 0 \\ \vdots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & \vdots & \vdots \\ \vdots & \vdots & \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_2 & 0 & \vdots \\ \vdots & \vdots & 0 & \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_3 & 0 \\ 1 & 0 & \vdots & 0 & \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_4 \\ & & 0 & & \end{bmatrix} \qquad n = n_1 + n_2 + \dots n_j$$

Variabili dommies: sono 1 se sta in quel gruppo, 0 altrimenti

$$\begin{bmatrix} X_{11} \\ \vdots \\ X_{1n_1} \\ X_{21} \\ \vdots \\ X_{2n_2} \\ X_{31} \\ \vdots \\ X_{3n_3} \\ X_{41} \\ \vdots \\ X_{4n_4} \end{bmatrix} = \vec{y}$$

Quando scrivo $\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_g \end{bmatrix}$

$$y = \beta Z + \varepsilon$$

ottengo

$j = 1 \dots n_1$

$$X_{1j} = \beta_0 + \beta_1 + \varepsilon_j$$

$j = n_1 + 1 \dots n_2$

$$X_{2j} = \beta_0 + \beta_2 + \varepsilon_j$$

la stima sui β mi stima come cambia la media tra gruppo.

$$E[X_j] = \beta_0 + \beta_1$$
$$E[X_{2j}] = \beta_0 + \beta_2$$

potrò testare quindi se c'è differenza o meno tra i gruppi

ANOVA: matrice di segno dove il predittore è categorica

La matrice non ha però rango pieno

$$Y = Z\beta + \varepsilon$$

$$\hat{\beta}_{LS} = (Z^T Z)^{-1} Z^T \vec{y}$$

È plausibile che i nostri dati siano generati da un modello del genere?

DECOMPOSIZIONE della VARIANZA

$$\vec{y} = \hat{\vec{y}} + \hat{\vec{\varepsilon}}$$ è la nostra decomposizione ortogonale $\hat{\vec{y}} \perp \hat{\vec{\varepsilon}}$

$$\vec{y}^T \vec{y} = \hat{\vec{y}}^T \hat{\vec{y}} + \hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}$$

data l'ortogonalità tra y e ε, di fatto è il teo di Pitagora

$$\Updownarrow$$

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2$$

vorremmo capire come spezzare la varianza

$$\Updownarrow$$

$$\sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2$$

$\bar{y} = \frac{1}{n} \sum y_i$

somma degli scarti dalla media al quadrato

Qui ho la media delle y, ma non delle $\hat{y}$

MA

OSS

$$n\bar{\hat{y}} = \sum_{i=1}^n \hat{y}_i = \hat{\vec{y}}^T \vec{1} = (H\vec{y})^T \vec{1} = \left(Z(Z^TZ)^{-1}Z^T\vec{y}\right)^T \vec{1} =$$

vettore di uni

$(H^T = H)$

$$= \vec{y}^T Z (Z^TZ)^{-1} Z^T \vec{1} = \vec{y}^T H \vec{1} =$$

H è il proiettore sul sottospazio lineare generato da Z. $\vec{1} \in \mathcal{L}(Z)$ xké è la prima colonna di Z. Quindi il proiettore lo manda in se stesso

$$= \vec{y}^T \vec{1} = \sum_{i=1}^n y_i = n\bar{y}$$

$$\Downarrow$$

La media delle y è = alla media di $\hat{y}$.

$$\Rightarrow \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

Formula di decomposizione della varianza

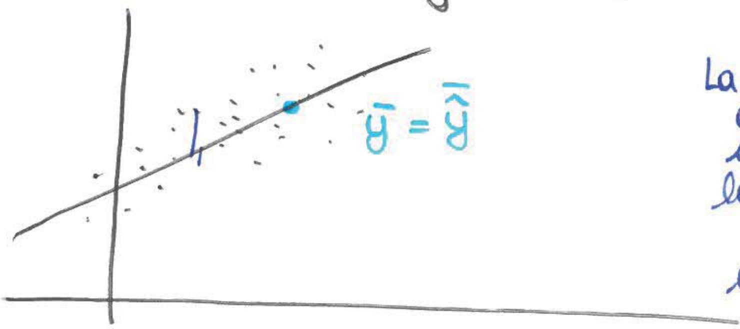
$\sum_{i=1}^{n}(y_i - \bar{y})^2$ = Sum of Square $_{TOT}$, SS_{TOT}
Somma totale delle y degli scarti dalla media al quadrato

$\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 = SS_{reg}$
Somma degli scarti al quadrato dalla media dei valori interpolati, fittati

$\sum_{i=1}^{n}\hat{\varepsilon}_i = SS_{res}$
Somma dei quadrati dei residui

La formula di decomposizione della varianza (vale quando in Z ho gli 1)

$$SS_{TOT} = SS_{reg} + SS_{res}$$



La variabilità dei dati intorno alla media la posso spezzare in variabilità della media e in variabilità dell'errore

Coefficiente di determinazione

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$0 \leq R^2 \leq 1$$

$R^2 = 1$ significa $SS_{res} = 0$, $\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = 0$
$\Updownarrow$
$\hat{\varepsilon}_i = 0 \quad \forall i$

ossia la retta interpola perfettamente i dati, i dati sono già sulla retta.

$R^2 = 0 \iff SS_{REG} = 0$

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = 0$$

la retta interpolante esce una costante

$$\Updownarrow$$

$$\hat{y}_i = \bar{y} \quad \forall i$$

la nostra previsione è tutta uguale ed è = alla media delle y.

$\downarrow$

È inutile in tal caso fare regressione, xk sto prevedendo tutto con la media, tutto costante.
~~la~~ variabilità del fenomeno è tutta scaricata nella variabilità dell'errore.

Anche $R^2 = 1$ non ci piace xk butto a 0 l'errore. i dati $\in \mathcal{L}(Z)$

Se la variabilità del modello mi coglie abbastanza la variabilità del fenomeno sono contento.
R^2 deve essere abbastanza alto (60% almeno)

(Posso togliere l'intercetta però $R^2 > 1$, xk niente garantisce più la decomposizione della varianza)

Come stimare σ^2?

TEOREMA

Rango $(Z) = r+1$.

1) $\mathbb{E}[\hat{\vec{\beta}}] = \vec{\beta}$ stimatori non distorti $\Rightarrow \hat{\vec{\beta}}$ è stimatore non distorto di $\vec{\beta}$
2) $Cov[\hat{\vec{\beta}}] = \sigma^2 (Z^T Z)^{-1}$
3) $\mathbb{E}[\hat{\vec{\varepsilon}}] = 0$
4) $Cov[\hat{\vec{\varepsilon}}] = \sigma^2 [\mathbb{1} - H]$
5) $\mathbb{E}[\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}] = \sigma^2 (n-r-1)$

$$\Downarrow$$

$$S^2 = \frac{\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}}{|n-r-1|} = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{|n-r-1|}$$

È STIMATORE non distorto per σ^2

$$Cov(\hat{\vec{\beta}}) = \begin{bmatrix} Var[\hat{\beta}_0] & & Cov(\hat{\beta}_i, \hat{\beta}_j) \\ & \ddots & \\ & & Var[\hat{\beta}_n] \end{bmatrix}$$

Gli elementi diagonali mi danno la variabilità di ogni stimatore
Ci interessa anche la covariabilità tra β_i e β_j.
Se l'inverso di $Z^T Z$ è "bello", quindi ho autovalori abb. grandi quindi l'inversa è stabile (buon numero di condizionamento), sono contento. (le colonne di $Z^T Z$ sono collineari).
Se le covariate sono molto dipendenti $Z^T Z$ diventa brutta e tutto diventa casino.
Come trattare se c'è correlazione tra le colonne degli Z?
Lo scopriremo solo vivendo ♪

dim

1) $\mathbb{E}[\hat{\vec{\beta}}] = \mathbb{E}[(Z^T Z)^{-1} Z^T \vec{Y}] = (Z^T Z)^{-1} Z^T \mathbb{E}[\vec{Y}] =$
$= (Z^T Z)^{-1} Z^T \mathbb{E}[Z\vec{\beta} + \vec{\varepsilon}] = (Z^T Z)^{-1} Z^T \underbrace{Z\vec{\beta}}_{\text{è deterministico}} = \vec{\beta}$

2) $Cov[\hat{\vec{\beta}}] = ?$

utilizziamo qst risultato:
$\vec{Y} = A\vec{X}$
$Cov(\vec{Y}) = A\, Cov(\vec{X})\, A^T$

$Cov(\vec{\varepsilon}) \times K$ la parte $Z\beta$ deterministica viene mangiata — è come una traslazione

$Cov[(Z^T Z)^{-1} Z^T \vec{Y}] = (Z^T Z)^{-1} Z^T \overbrace{Cov(\vec{Y})}^{= Cov(\vec{\varepsilon})} Z (Z^T Z)^{-1} =$
$= (Z^T Z)^{-1} Z^T (\sigma^2 \mathbb{1}) Z (Z^T Z)^{-1} =$
$= \sigma^2 (Z^T Z)^{-1}$

3) $\mathbb{E}[\hat{\vec{\varepsilon}}] = \mathbb{E}[\vec{Y} - \hat{\vec{Y}}] = \mathbb{E}[(\mathbb{1} - H)\vec{Y}] = (\mathbb{1} - H)\overbrace{\mathbb{E}[\vec{Y}]}^{= Z\beta} =$
$= (\mathbb{1} - H) Z\vec{\beta} = Z\vec{\beta} - HZ\beta = Z\vec{\beta} - Z(Z^T Z)^{-1} Z^T Z\beta = 0$

infatti H proietta sullo spazio di Z quindi mi aspetto che $HZ\beta = Z\beta$

4) $Cov(\hat{\vec{\varepsilon}}) = Cov((1 - H)\vec{Y}) = (1 - H)\, Cov(\vec{Y})\, (1 - H)^T =$
$= (1 - H)\sigma^2 (1 - H)^T = \sigma^2 (1 - H)^2 = \sigma^2 (1 - H)$

$1 - H$ proietta sull'ortogonale di $\mathcal{L}(Z)$.
Vale $(1 - H)^T = 1 - H$ e $H^T = H$ (AUTOAGGIUNTI) (simmetrici)

proprietà dei proiettori
Sono idempotenti:
$(1 - H)^2 = 1 - H$
$H^2 = H$.

La matrice $\mathbb{1}-H$ non ha rango pieno xk
proietta su un sottospazio di dimensione $n-k-1$

5) $\mathbb{E}[\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}] = \mathbb{E}\left[\sum_{i=1} \hat{\varepsilon}_i^2\right] = \mathbb{E}\left[tr\left(\hat{\vec{\varepsilon}}\, \hat{\vec{\varepsilon}}^T\right)\right] =$

ho matrice $n \times n$ con sulla diagonale $\hat{\varepsilon}_i^2$

$= \mathbb{E}\left[tr\left((\mathbb{1}-H)\vec{Y}\left((\mathbb{1}-H)\vec{Y}\right)^T\right)\right] -$

OSS $(\mathbb{1}-H)\vec{Y} = (\mathbb{1}-H)(Z\vec{\beta}+\vec{\varepsilon}) = (\mathbb{1}-H)\vec{\varepsilon}$

$\mathbb{1}-H$ proietta nell' $\perp \mathcal{L}(Z)$
$\Rightarrow (\mathbb{1}-H)Z = 0$

$= \mathbb{E}\left[tr\left((\mathbb{1}-H)\vec{\varepsilon}\,\vec{\varepsilon}^T\underbrace{(\mathbb{1}-H)}\right)\right] =$

abbiamo visto $(\mathbb{1}-H)^T = (\mathbb{1}-H)$

$tr(AB) = tr(BA)$

permuto

$= \mathbb{E}\left[tr\left((\mathbb{1}-H)\vec{\varepsilon}\,\vec{\varepsilon}^T\right)\right] = tr(\mathbb{1}-H)\,\underbrace{\mathbb{E}\left[\vec{\varepsilon}\,\vec{\varepsilon}^T\right]}=$

la media di $\vec{\varepsilon}$ è 0, quindi qst è la matrice varianza covarianze $Cov(\vec{\varepsilon}) = \sigma^2\mathbb{1}$

$= \sigma^2\, tr(\mathbb{1}-H) =$

la traccia di un proiettore è la dimensione dello spazio in cui proietto
(sarebbe $tr(\mathbb{1}) - \underbrace{tr(Z(Z^TZ)^{-1}Z^T)}_{tr(\mathbb{1}\text{ dimensione } k+1)}$)

$= \sigma^2(n-k-1)$

Quindi ottengo che

$\mathbb{E}[\hat{\varepsilon}_i^2] = \sigma^2(n-k-1)$

$\Downarrow$

$S^2 = \dfrac{\sum \hat{\varepsilon}_i^2}{n-k-1}$ è stimatore non distorto per σ^2

è l'SS res.

Æ

TEOREMA GAUSS-MARKOV

ASSUMIAMO CHE Z ABBIA RANGO $= R+1$.

$\forall \vec{c} \in \mathbb{R}^{R+1}$

$\vec{c}^T \hat{\vec{\beta}}$ è lo stimatore di varianza minima per stimare $\vec{c}^T \vec{\beta}$ tra tutti gli stimatori che soddisfano

a) non distorti per $\vec{c}^T \vec{\beta}$

b) sono della forma

$$a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n \quad (\text{lineari in } \vec{Y})$$

ovvero

$\vec{c}^T \hat{\vec{\beta}}$ è il BLUE = Best Linear Unbiased Estimator

- Quello a varianza minima nella classe degli stimatori lineari in Y non distorti

Se assumero che $\vec{\varepsilon} \sim N(\vec{0}, \sigma^2 \mathbb{1}_n)$

$$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon} \implies \vec{Y} \sim N_n(Z\vec{\beta}, \sigma^2 \mathbb{1}_n)$$

n dimensionale

26 maggio 2015

Assumiamo

$$\vec{\varepsilon} \sim N_n(\vec{0}, \sigma^2 \mathbb{1})$$

la struttura lineare del modello fa in modo che $\vec{Y}$ sia normale e quindi anche $\hat{\beta}$, xk è funzione di Y.

TEOREMA

$\vec{\varepsilon} \sim N(0, \sigma^2 \mathbb{1})$

Rango $(Z) = R+1$ (rango pieno)

1) $\hat{\vec{\beta}}$ e $\hat{\sigma}^2 = \dfrac{\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}}{n} = \dfrac{n-R-1}{n} S^2$ sono gli stimatori MLE di $\vec{\beta}$ e σ^2

2) $\hat{\vec{\beta}} \sim N_{R+1}\left(\vec{\beta}, \sigma^2 (Z^T Z)^{-1}\right)$ abbiamo in più la gaussianità

3) $\hat{\vec{\varepsilon}} \sim N_n(\vec{0}, \sigma^2(\mathbb{1} - H))$ $(\mathbb{1} - H)$ ha rango $n-R-1$, in $\mathbb{R}^n$ non ha rango pieno, si può ancora scrivere con le f. caratt.

4) $\hat{\vec{\beta}} \perp\!\!\!\perp \hat{\vec{\varepsilon}}$ indipendenza stocastica

5) $n\hat{\sigma}^2 = n\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}} = n\sum \hat{\varepsilon}_i^2 \sim \sigma^2 \cdot \chi^2(n-R-1)$

(14) Posso così fare non più solo inferenza puntuale.

dim

1) sol Jhonson Wickern

$$\vec{\varepsilon} \sim N_n(\vec{0}, \sigma^2 \mathbb{1})$$

$$\vec{Y} \sim N(\cancel{Z}\beta, \sigma^2 \mathbb{1})$$

Quindi scrivendone la verosimiglianza

$$L(\beta, \sigma^2, \vec{y}) = \dots \exp(-\Sigma(y_i - z\vec{\beta}_i)^2)$$

massimizzare la verosimiglianza in β
vuol dire minimizzare qst distanza euclidea.
La struttura di una gaussiana

$$\Downarrow$$

$$\hat{\vec{\beta}}_{ML} = \hat{\vec{\beta}}_{LS}$$

viene ausiliato, ma non ci stupisce xk nel caso gaussiano la ML non ci dava la media campionaria.

2...4)

$$\begin{bmatrix} \hat{\vec{\beta}} \\ \hat{\vec{\varepsilon}} \end{bmatrix} \begin{matrix} \}\, R+1 \\ \}\, n \end{matrix}$$

$$\Uparrow \quad \mathbb{R}^{n+R+1}$$

$$\begin{bmatrix} \hat{\vec{\beta}} \\ \hat{\vec{\varepsilon}} \end{bmatrix} = \begin{bmatrix} (Z^T Z)^{-1} Z^T \\ \mathbb{1} - Z(Z^T Z)^{-1} Z^T \end{bmatrix} \vec{Y}$$

$\vec{Y}$ è Gaussiano xk

$$\vec{\varepsilon} \sim N(0, \sigma^2 \mathbb{1})$$

$$\Downarrow$$

$$\vec{Y} \sim N_n(Z\vec{\beta}, \sigma^2 \mathbb{1})$$

quindi

$\begin{bmatrix} \hat{\vec{\beta}} \\ \hat{\vec{\varepsilon}} \end{bmatrix}$ è Gaussiano xk trasformazione lineare di $\vec{Y} \sim N$

$$\begin{pmatrix} \hat{\vec{\beta}} \\ \hat{\vec{\varepsilon}} \end{pmatrix} \sim N_{n+R+1} \left(\begin{matrix} \vec{\beta} \\ 0 \end{matrix} \; ; \; \sigma^2 \circledast \right)$$

facendo le medie dei $\hat{\vec{\beta}}$ e $\hat{\vec{\varepsilon}}$

$\circledast$ matrice varianza covarianza.

$$\begin{bmatrix} (Z^T Z)^{-1} Z^T \\ \mathbb{1} - H \end{bmatrix}$$ devo moltiplicarla x il suo trasposto

viene matrice a blocchi

$$\circledast \left[\begin{array}{c|c} (Z^T Z)^{-1} Z^T \left[(Z^T Z)^{-1} Z^T \right]^T & (Z^T Z)^{-1} Z^T (\mathbb{1} - H)^T \\ \hline (\mathbb{1} - H) \left[(Z^T Z)^{-1} Z^T \right]^T & (\mathbb{1} - H)(\mathbb{1} - H)^T \end{array} \right] =$$

$$= \begin{bmatrix} (Z^T Z)^{-1} & 0 \\ 0 & \mathbb{1} - H \end{bmatrix}$$

xk $(Z^T Z)^{-1} Z^T$ sta nel sottospazio lineare, e $\mathbb{1} - H$ proietta nell'ortogonale.

questi già li sapevamo, xk la matrice di varianza e covarianza di $\hat{\beta}$ ed $\hat{\varepsilon}$ era qst

Quindi

$$\begin{pmatrix} \hat{\vec{\beta}} \\ \hat{\vec{\varepsilon}} \end{pmatrix} \sim N_{n+R+1} \left(\begin{matrix} \vec{\beta} \\ 0 \end{matrix} \; ; \; \sigma^2 \begin{bmatrix} (Z^T Z)^{-1} & 0 \\ 0 & \mathbb{1} - H \end{bmatrix} \right)$$

dato che sono nel mondo gaussiano la scorrelazione ci dà l'indipendenza, quindi $\hat{\vec{\beta}} \perp \hat{\vec{\varepsilon}}$.

116

5) $\vec{x} \sim N(\vec{\mu}, \vec{\Sigma}) \Rightarrow$ si può dimostrare che

$$\underbrace{(\vec{x}-\mu)^T \Sigma^{-1} (\vec{x}-\vec{\mu})}_{\text{di fatto mi diventa la somma di Gaussiane al quadrato.}} \sim \chi^2(\underbrace{n}_{\text{i gradi di libertà sono il rango della matrice } \Sigma})$$

di fatto mi diventa la somma di Gaussiane al quadrato.

i gradi di libertà sono il rango della matrice Σ

$$\hat{\vec{\varepsilon}}^T \underbrace{(\mathbb{1}-H)^{-}}\frac{1}{\sigma^2} \hat{\vec{\varepsilon}} \sim \chi^2(n-(k+1))$$

metto la pseudoinversa di $\sigma^2(\mathbb{1}-H)$ xk qst non è invertibile

i gradi di libertà sono il vero rango

Io so $H\hat{\vec{\varepsilon}} = 0$ x ortogonalità.

Quindi $\hat{\vec{\varepsilon}} = (\mathbb{1}-H)\hat{\vec{\varepsilon}}$ xk $\mathbb{1}-H$ proietta dove vive $\hat{\vec{\varepsilon}}$.

Quindi posso scrivere

$$\hat{\vec{\varepsilon}}^T \cancel{(\mathbb{1}-H)^{-}} \cancel{(\mathbb{1}-H)} \hat{\vec{\varepsilon}} \frac{1}{\sigma^2} \sim \chi^2(n-k-1)$$

$$\Downarrow$$

$$\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}} \sim \sigma^2 \chi^2(n-k-1)$$

■

Corollario

$$\frac{1}{\sigma^2}(\underbrace{\hat{\vec{\beta}}-\vec{\beta}}_{})^T (Z^T Z)(\hat{\vec{\beta}}-\vec{\beta}) \sim \chi^2(k+1)$$

sto togliendo a un vettore Gaussiano la sua media

matrice varianza covarianza

quantità pivot

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k-1)$$

mi permetteranno di scrivere IC.
ma è incognito sia β che σ^2, quindi devo usare entrambi i risultati. Il vantaggio è che le due quantità sono indipendenti.

Per fare inferenza su β devo togliere il σ^2.
Potrei fare il quoziente.
Il quoziente di due χ^2 indipendenti è una Fisher.

$$\frac{\chi^2(p)/p}{\chi^2(q)/q} \sim F(p,q)$$

Allora sfruttando qst risultato:

$$\frac{(\hat{\vec{\beta}} - \vec{\beta})^T (Z^T Z)(\hat{\vec{\beta}} - \vec{\beta}) \frac{1}{R+1}}{n\hat{\sigma}^2 \cdot \frac{1}{n-R-1}} \sim F(R+1, n-R-1)$$

quantità pivotale

$$= S^2 = \frac{\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}}{n-R-1} \Rightarrow \text{stimatore non distorto di } \sigma^2$$

Per costruire ~~un~~ IC di livello $1-\alpha$ per i miei $\vec{\beta}$
Regione di confidenza

$$R(1-\alpha) = \left\{ \vec{\beta} \in \mathbb{R}^{R+1} : \underbrace{(\hat{\vec{\beta}} - \vec{\beta})^T (Z^T Z)(\hat{\vec{\beta}} - \vec{\beta})}_{} \leq (R+1)\, S^2\, F_{1-\alpha}(R+1, n-R-1) \right\}$$

ellissoide centrato in $\hat{\vec{\beta}}$ in cui ~~[illegible]~~ gli assi sono dettati dagli autovalori e autovettori $(Z^T Z)^{-1}$

stima puntuale di S^2

$\Downarrow$

più la matrice disegno è fatta male $(Z^T Z)$ mal condizionata, tanto più l'ellissoide degenera.
Ogni volta che pesco da una gaussiana un punto memoria di $(Z^T Z)^{-1}$

 Spesso si costruiscono gli intervalli di confidenza uno alla volta, "ONE AT A TIME"

$$\hat{\beta}_j \sim N\left(\beta_j ; \sigma^2 \underbrace{diag_j(z^T z)^{-1}}_{\text{elemento j-esimo sulla diagonale.}}\right)$$

Posso stimare con S^2 e avere una t.

Quindi gli IC che si creano sono marginali:

$$\hat{\beta}_j \pm \sqrt{S^2 diag_j (z^T z)^{-1}} \; t_{1-\frac{\alpha}{2}}(n-(R+1))$$

In R avremo la stima di

β_j per una statistica test $\begin{cases} \beta_j = 0 \\ \beta_j \neq 0 \end{cases}$

così decido marginalmente se quella covariata entra nel modello oppure no.

Posso fare selezione delle covariate, scelta del modello.

Però così perdo la matrice di varianza.

E prima avevo una Fisher, qui ho una t.

Cosa c'entrano?

Si ha che ~~[illegible]~~ i quantili della t al quadrato sono quelli della Fisher a 1 gdl.

Abbiamo un legame stretto tra livello, p-value e regione di accettazione.

TEST SUI PARAMETRI

C matrice $p \times (R+1)$

Posso fare test qualsiasi del tipo (mi danno sia $\beta_j = 0$ vs $\beta_j \neq 0$ che combinazioni dei β_j)

$H_0: C\vec{\beta} = 0$

$H_1: C\vec{\beta} \neq 0$

$c\hat{\vec{\beta}} \sim N_p(\vec{0}, \sigma^2 c (z^T z)^{-1} c^T)$

non posso usarlo come statistica test xk c'è σ^2.

Ma posso ricondurmi che

indip $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-R-1)$

$$\frac{(c\hat{\vec{\beta}})^T (c(z^T z)^{-1} c^T)^{-1} c\hat{\vec{\beta}}}{S^2 p} \sim F(p, n-R-1)$$

gdl del numeratore (ho rango di c pieno)

dovrei dividere per $n\hat{\sigma}^2$ ma ho messo qst

Rifiuto Ho se qst è $> f_{1-\alpha}$:

$$(c\hat{\vec{\beta}})^T (c(z^T z)^{-1} c^T)^{-1} c\hat{\vec{\beta}} \geq S^2 p\, f_{1-\alpha}(p, n-R-1)$$

con $n=1$ mi ritrovo il t-test.

Con questo posso fare tutti i controlli sui vari coefficienti

Vogliamo dire se un blocco di coefficienti è 0 o no x consideriamo blocchi di covariate.

H_1 $\beta_{R-p+2} = \dots = \beta_{R+1} = 0$ almeno uno $\neq 0$

$\rightarrow \beta_j = 0$ prendere la covariata associata. se è 0 non mi interessa se prendere o no la covariata nel modello.

Se le covariate fossero tra loro ⊥ non avrei problema di ordine, ma se non lo sono no.

$Z = [Z_1 \mid Z_2]$, con Z_1 di $R+1-p$ colonne e Z_2 di p

$Y = Z\vec{\beta} + \vec{\varepsilon}$ modello globale

$Y = Z_1\vec{\beta} + \vec{\varepsilon}$ modello ridotto.

se Z_2 mi dà le covariate $=0$, ho che i due modelli sono uguali. Quindi prenderei il secondo.

Prendo allora

$$SS_{res}(Z) = \hat{\vec{\varepsilon}}^{\,T}\hat{\vec{\varepsilon}}$$

$$SS_{res}(Z_1) = \hat{\vec{\varepsilon}}_1^{\,T}\hat{\vec{\varepsilon}}_1 \quad \left(\text{usati con } Z_1\right)$$

mi aspetto che

$$\underbrace{SS_{res}(Z_1)}_{\text{nel modello ridotto è più grande}} - \underbrace{SS_{res}(Z)}_{\text{ho arricchito il modello, quindi avrò meno residuo}}$$

RIFIUTO H_0, RIFIUTO IL FATTO CHE NON MI SERVANO I PEZZI IN Z_2 SE QUESTA QUANTITÀ È "GRANDE"

SE LA DIFFERENZA È PICCOLA INVECE AGGIUNGERE Z_2 NON MI DÀ TROPPE INFO IN PIÙ.

Si può dimostrare che

$$\frac{SS_{res}(Z_1) - SS_{res}(Z)}{S^2\,p} \sim F\left(p, n-(k+1)\right)$$

p ↑ dimensione delle colonne che stiamo tagliendo

Rifiuto se è + grande del quantile di ordine $1-\alpha$ della Fisher.

Testare l'ipotesi che i β_j legati alle covariate di Z_2 siano $=0$ o almeno una $\neq 0$.

Se $p=1$ torno al test $\beta_j = 0$

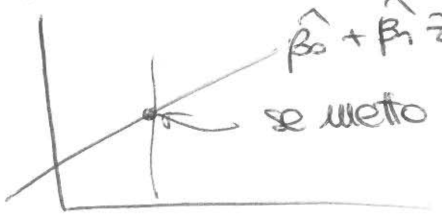
$\beta_j \neq 0$

PREVISIONE

Supponiamo di fissare $\vec{z}_0 = (1, z_{01} \dots z_{0r})^T$
generico valore delle nostre covariate.

So che
$$\mathbb{E}[Y|\vec{z}_0] = \vec{z}_0^T \vec{\beta}$$
lo so che lo stimatore blue di $\vec{\beta}$ è $\hat{\vec{\beta}}$.

la stima che faccio
$$\mathbb{E}[Y_0] = z_0^T (Z^T Z)^{-1} Z^T Y = \vec{z}_0^T \hat{\vec{\beta}}$$
mi stima la media della mia risposta in z_0



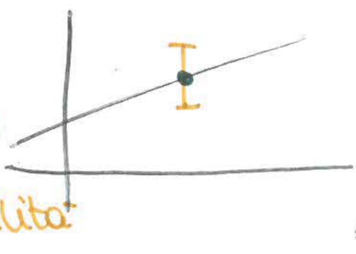
$\hat{\beta}_0 + \hat{\beta}_1 z$

se metto dentro z_0 valuto la risposta che ho in z_0, proprio quella sulla retta.

Errore
$$\vec{\varepsilon}_n \sim N_n(0, \sigma^2 \mathbb{1})$$
$$\vec{z}_0^T \hat{\vec{\beta}} \sim N_1(\vec{z}_0^T \vec{\beta}, \ \sigma^2 z_0^T (Z^T Z)^{-1} z_0)$$
è indipendente da
$$\frac{n \hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}}{\sigma^2} \sim \chi^2(n-r-1)$$
Allora è come fare il test sulla media di una Gaussiana a varianza incognita.

Quindi
$$\frac{\vec{z}_0^T (\hat{\vec{\beta}} - \vec{\beta})}{\sqrt{\vec{z}_0^T (Z^T Z)^{-1} \vec{z}_0}} \cdot \frac{1}{\sqrt{S^2}} \sim t(n-r-1)$$
Quindi IC di livello $1-\alpha$ per $\mathbb{E}[Y|\vec{z}_0] = \vec{z}_0^T \vec{\beta}$
è $\vec{z}_0^T \hat{\vec{\beta}} \pm S \sqrt{\vec{z}_0^T (Z^T Z)^{-1} \vec{z}_0} \ t_{1-\frac{\alpha}{2}}(n-r-1)$

IC per la media



stima puntuale della media
stima intervallare, tiene conto della variabilità

IC x la media reg. lineare 1^anno

 $\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$

assumiamo che i nostri dati vengano da un modello lineare.
Vogliamo costruire delle quantità x fare previsione;
fisso un vettore di covariate $\vec{z}_0^T$ e in corrispondenza di questo
ho $y_0 = \vec{z}_0^T \vec{\beta} + \varepsilon_0$, risposta che ho assumendo quel set di covariate.

(gli z_0 possono non essere presenti nel mio data set)

$$\mathbb{E}[Y_0] = \vec{z}_0^T \vec{\beta}$$

Possiamo costruire degli IC di per la media di Y_0?

$\vec{z}_0^T \hat{\vec{\beta}}$ è una v.a., posso ~~tener~~ conto della sua variabilità per costruire IC

$$\underbrace{\frac{\vec{z}_0^T (\hat{\vec{\beta}} - \vec{\beta})}{\sqrt{\vec{z}_0^T (Z^T Z)^{-1} \vec{z}_0}} \frac{1}{\sqrt{S^2}}}_{\text{quantità pivot}} \sim t(n-R-1)$$

Posso invertirne rispetto a $\vec{z}_0^T \vec{\beta}$, che è quello di cui voglio stimare

Quindi
IC di livello $1-\alpha$ per $\vec{z}_0^T \vec{\beta}$ (che è la media di Y_0)
è

$$\vec{z}_0^T \hat{\vec{\beta}} \pm S \sqrt{\vec{z}_0^T (Z^T Z)^{-1} \vec{z}_0} \; t_{1-\frac{\alpha}{2}}(n-R-1)$$

$\vec{z}_0^T \hat{\vec{\beta}}$ è la trasformazione di una gaussiana

$$\vec{z}_0^T \hat{\vec{\beta}} \sim N_1\left(\vec{z}_0^T \vec{\beta} \; ; \; \sigma^2 \vec{z}_0^T (Z^T Z)^{-1} \vec{z}_0 \right) \quad (*)$$

xk i $\hat{\vec{\beta}}$ sono non distorti

ma so anche

$$\frac{n \hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}}{\sigma^2} \sim \chi^2(n-R-1)$$

e io ho $\dfrac{z}{\sqrt{\chi^2(p)/p}} \sim t(p)$

Allora standardizzo (*) e divido per quella sotto e sono a posto

[Weisberg - Applied linear regression ha i calcoli fatti in modo esplicito]

Avevamo trovato che la varianza, in regressione semplice, era

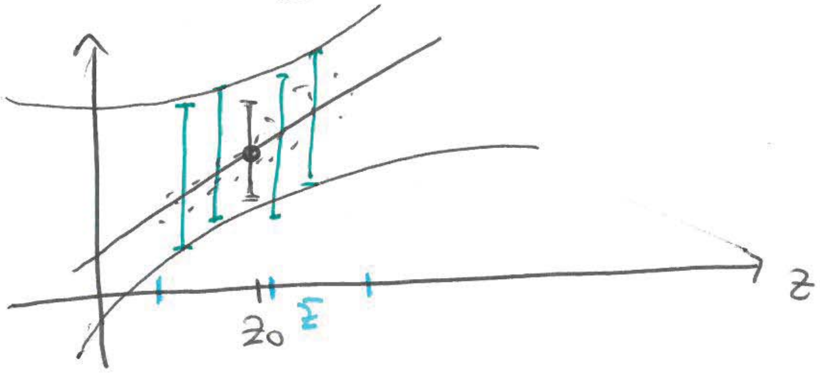
$$S^2\left(\frac{1}{n} + \frac{(z_0 - \bar{z})^2}{S_{zz}}\right)$$

QUI ALEI HA SCRITTO $\sum_{i=1}^{n}(z_i - \bar{z})^2$

con $S_{zz} = \sum (z_i - \bar{z})^2$

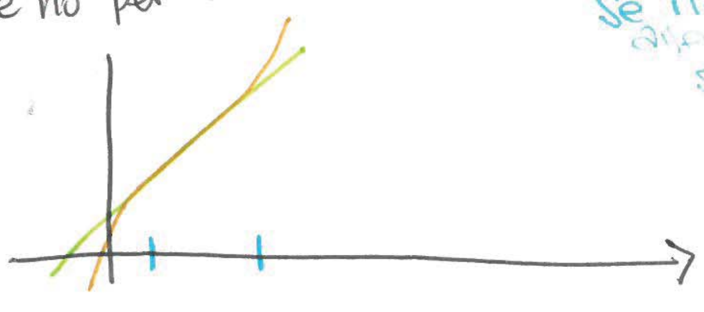
la variabilità dipende da
S^2 (variabilità errore)

e $(z_0 - \bar{z})^2$: distanza del punto rispetto al baricentro (media campionaria z)



la variabilità aumenta tanto più mi allontano dal baricentro
più mi allontano dal baricentro dai dati su cui ho costruito il modello, tanto più la mia stima è soggetta a errore:
PROBLEMA dell'ESTRAPOLAZIONE

Se ho per es



Se riesco ad avere dati solo qui
non capisco da quale dei due modelli sono generati i miei dati

(124) Problema in cui si è sbagliata la stima ha fatto esplodere uno space shuttle

Vorrei tener conto della variabilità x poter fare la previsione.

$Y_0 = \vec{z_0}^T \vec{\beta} + \varepsilon_0 \qquad \varepsilon_0 \sim N(0, \sigma^2)$

e ipotizziamo $\varepsilon_0 \perp\!\!\!\perp \vec{\varepsilon}$

$Y_0 \perp\!\!\!\perp \hat{\vec{\beta}}$, xk sono indip. da $\vec{\varepsilon}$

Allora si può dimostrare che

$$Y_0 - \vec{z_0}^T \hat{\vec{\beta}} \sim N\left(0, \sigma^2\left(1 + \vec{z_0}^T (Z^T Z)^{-1} \vec{z_0}\right)\right)$$

ho anche variabilità di $\hat{\vec{\beta}}$.

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-R-1)$$

Quindi

$$P\left[Y_0 \in \left(z_0^T \hat{\vec{\beta}} \pm t_{1-\frac{\alpha}{2}}(n-R-1)\sqrt{S^2(1 + \vec{z_0}^T(z_0^T z)^{-1}\vec{z_0})}\right)\right] = 1-\alpha$$

in che intervallo la massa di $Y_0 \in (1-\alpha)$

Intervallo di Previsione.

HA un S^2 in più sotto radice rispetto a prima.

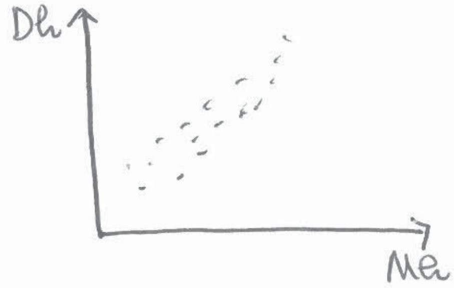
È quindi un pochino più ampio di quello di prima xk contiene un errore in più, ho anche un pezzo di variabilità dovuta all'errore.

In R: confidence / prevision a seconda di quale dei due voglio costruire.

Se pongo $z_0 = z_i$ che è uno dei miei dati vedo la bontà del mio modello, vedo $\hat{y} - y$, quindi di fatto guardo i residui.

Le Z sono deterministiche.
Ci sono due tipi di matrici disegno:
MATRICE SPERIMENTALE e OSSERVAZIONALE

Data set in cui studio l'altezza delle figlie rispetto a qll delle madri



studio osservazionale

io sperimentatore statistico non ho controllo su qll che succede

Voglio capire se la produzione del terreno aumenta all'aumentare del concime.
Sono io che sperimentalmente posso "disegnare" il mio modello

Se osservo correlazione tra X e Y, al crescere di X cresce Y,
nell'osservazionale non posso arrivare a una conclusione di causalità,
mentre nello sperimentale, a parità di tutte le cose tranne il concimante, posso avere un rapporto di causalità.

Fattore di confondimento: le cicogne che portano i bambini
Si era dimostrato che più cicogne ci sono in un paese e più c'erano cicogne. Il fattore di confondimento era che qnd nasceva il bambino si scaldava l'acqua nel camino e le cicogne facevano il nido sui comignoli

(126) Supponiamo di avere

$$\vec{Y} = Z\vec{\beta} + \varepsilon$$

cos'è
$\hat{\beta}_j$?

Fissate tutte le altre covariate $\vec{z}_1 \dots \vec{z}_{j-1}, z_{j+1}, \dots \vec{z}_n$
β_j rappresenta l'incremento nella media della Risposta all'aumentare di una unità di z_j.

$$\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3$$

se ho

$$\beta_0 + \beta_1 z_1 + \beta_2 (z_2 + 1) + \beta_3 z_3$$

all'aumentare di 1 elemento della covariata z_2 mi dice di quanto aumenta la media all'aumentare unitario della covariata.

Dobbiamo sempre stare attenti alle unità di misura e gli ordini di grandezza.

Posso avere matrici disegno che corrispondono a covariate categoriche.
Come si scrive la matrice disegno in un caso del genere?

MODELLO ANOVA

$$\vec{Y} = \begin{bmatrix} x_{11} \\ \vdots \\ x_{1n_1} \\ x_{21} \\ \vdots \\ x_{2n_2} \\ \vdots \\ x_{jn_g} \end{bmatrix}$$

$\boxed{x_{ij} \text{ risposta } i = 1 \dots n_j \text{ nel gruppo } j = 1 \dots g}$

Prendiamo lo stesso numero di osservazioni uguali x risparmiare gesso

$$n_1 = n_2 = \dots = n_g = m.$$

$$Z \quad \begin{bmatrix} x_{11} \\ \vdots \\ x_{1n_1} \\ x_{21} \\ \vdots \\ x_{2n_2} \\ x_{\vdots} \\ \vdots \\ x_{gn_g} \end{bmatrix}$$

ci verrebbe da scrivere

$$\begin{array}{c} n_1 \\ \\ n_2 \\ \\ n_3 \end{array} \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ \hline \vdots & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & 0 \\ \hline \vdots & \vdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Quando faccio la stima ottengo

$$\mathbb{E}[X_{ji}] = \beta_0 + \beta_j$$ tanto l'errore è a media nulla

Ma la matrice è mal condizionata.
Ne voglio scrivere una uguale ma invertibile.
Allora immagino

$$\mathbb{E}[X_{ij}] = \underset{\substack{\uparrow \\ \text{media} \\ \text{che dipende dal} \\ \text{gruppo}}}{\mu_j} = \mu + (\mu_j - \mu) = \mu + \tau_j \; ; \quad \boxed{\sum \tau_j = 0}$$

media globale + discostamento dalla media globale

L'ANOVA:
se mi metto nell'hp di gaussianità guardo se la media è la stessa nel gruppo oppure no.
un'anova con 2 gruppi è come fare un t-test di confronto.
Se $Var[\vec{\varepsilon}] = \sigma^2$ significa che la varianza deve essere la stessa.

$$X_{ij} \sim N(\mu_i, \sigma^2)$$

Posso capire quindi se il consumo medio delle macchine cambia o no al cambiare della marca.

$$Z = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots \; \}n_1 & \vdots & \vdots \\ \vdots & 1 & 0 & 0 \\ \vdots & 0 & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & 1 & 1 \\ \vdots & -1 & -1 & -1 \\ \vdots & \vdots \; \}n_g & \vdots \; \}n_g & \vdots \; \}n_g \\ 1 & -1 & -1 & -1 \end{bmatrix} \begin{array}{c} \beta_0 \\ \vdots \\ \beta_{g-1} \end{array}$$

matrice con g colonne, una in meno rispetto a prima.

$$\mathbb{E}[X_{i1}] = \beta_0 + \beta_1 \qquad \tau_1 = \beta_1$$

$$\mathbb{E}[X_{i2}] = \beta_0 + \beta_2 \qquad \tau_2 = \beta_2 \ldots$$

$$\mathbb{E}[X_{ig}] = \beta_0 - \beta_1 - \beta_2 - \beta_{g-1}$$

ottengo così

$$\tau_1 + \tau_2 + \ldots = 0$$

$$X = \begin{bmatrix} X_{i1} \} \ldots \\ X_{i2} \ldots \\ X_{ig} \ldots \end{bmatrix}$$

 ~~BMW~~ Fiat Toyota

$1 \frac{km}{l}$ $10 \frac{km}{l}$ .... passo un data set

La matrice disegno toglie una categoria, di solito la 1^ in ordine alfabetico.

Le stime prodotte diventano

$\hat{\beta}_0$ media nel primo gruppo (BMW)

$\hat{\beta}_1$ = differenza tra consumo fiat e BMW

⇒ Per ricostruire il consumo della fiat: $\hat{\beta}_1 - \hat{\beta}_0$

COSA FA L'ANOVA?

Produce un p-value

Matrice z_1 $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

$$\hat{\vec{y}} = \underset{\uparrow}{\bar{X}_{\cdot\cdot}} = \frac{1}{mg} \sum_{i,j} X_{ij}$$ osservazione = x tutto

~~media su~~ su gruppi che osservazione

Nella matrice disegno Z si ha

$$\hat{\vec{y}} = \begin{bmatrix} \bar{X}_{1\cdot} & \}m \\ \bar{X}_{2\cdot} & \}m \\ \vdots & \\ \bar{X}_{g\cdot} & \}m \end{bmatrix}$$ media sul primo gruppo

↑ prevedo in realtà x ogni gruppo la media del gruppo.

Vogliamo capire se c'è differenza tra il modello completo e gli delle singole.

TEST:

$$\frac{SS_{res}(z_1) - SS_{res}(z)}{\underbrace{\frac{SS_{res}(z)}{(mg-g)}}_{\text{gdl di } SS(z)} \cdot \underbrace{(g-1)}_{\text{rango } z, \text{ che ha } g-1 \text{ colonne}}} = \frac{\sum_{i=1}^{g} \sum_{j=1}^{m} (X_{ij} - \bar{X}_{..})^2 - \sum_{i=1}^{g} \sum_{j=1}^{m} (X_{ij} - \bar{X}_{i\cdot})^2}{(g-1) \frac{\sum_{i=1}^{g} \sum_{j=1}^{m} (X_{ij} - \bar{X}_{i\cdot})^2}{mg-g}}$$

Se io prendo

$$\sum_{i=1}^{g} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{..})^2$$

e faccio $\pm \bar{x}_{i.}$

i doppi prodotti si annullano.
Quindi ho

$$\underbrace{\sum_{i=1}^{g} \sum_{j=1}^{\bar{\bar{m}}} (x_{ij} - \bar{x}_{i.})^2}_{\text{variabilità nei gruppi: } SSw} + \underbrace{m \sum_{i=1}^{g} (\bar{x}_{i.} - \bar{x}_{..})^2}_{SSb}$$

per la linearità della media.

variabilità nei gruppi:

SSw
within

Per ogni gruppo guardo lo scarto con la media.

SSb between
guarda la variabilità tra i gruppi.
Guarda la variabilità tra la media del gruppo e la media totale

$\Rightarrow$ gli dell'altra pag $= \boxed{\dfrac{SSb/(g-1)}{SSw/(g(m-1))} \leftarrow = \dfrac{MSb}{MSw}}$

mean square

Rifiuterò H_0 (medie tutte uguali)
vs H_1 (c'è una media $\neq$)

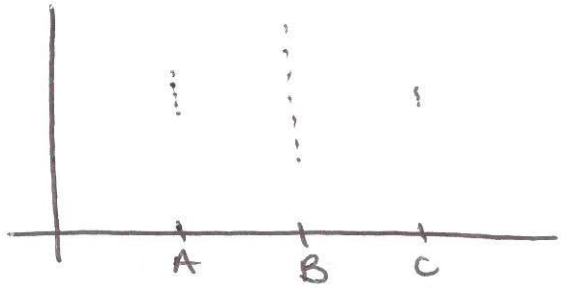
H_0 : $\tau_1 = .. \tau_g = 0$
H_1 : almeno un $\tau_i \neq 0$ (ed essendo a $\Sigma = 0$ in realtà saranno almeno due $\neq 0$)

Rifiuto se

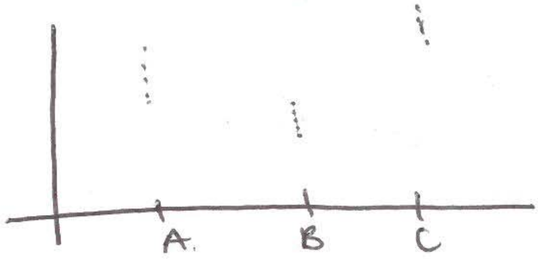
$$\boxed{\frac{SSb/(g-1)}{SSw/(m-1)g} > F_{1-\alpha}(g-1;\ (m-1)g)}$$

cioè se è troppo grande.

Perché la variabilità ~~delle me~~ mi porta a parlare delle medie?



supponiamo abbiano tutti la stessa media



media diversa

SS between: distanza della media del gruppo rispetto a qll totale
Nel primo caso qst deve essere confrontabile con la variabilità nel gruppo.

Nel secondo caso invece la variabilità delle medie è molto + grande della variabilità nel gruppo singolo.

ANALISI della VARIANZA ↗
caso particolare di modello lineare
La decomposizione della varianza vale ancora e dà risposta al fatto che le medie sono uguali o ≠.

Pvalue piccolo ⇒ i modelli sono ≠,
l'∃ del gruppo è importante

Nel caso generale
SSb è SSreg

28 maggio 2015

$$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$$

$$\vec{\varepsilon} \sim N_n(\vec{0}, \sigma^2 \mathbb{I})$$

Diagnostica del Modello

Noi assumiamo che i dati siano generati da $\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$.
Ma è vero che i dati sono generati da questo modello?
E se il modello non va bene devo buttare via tutto?

DIAGNOSTICA: controllo delle assunzioni fatte sul modello che genera i dati.

Modello Teorico

$\mathbb{E}[\vec{\varepsilon}] = 0$

$Cov(\vec{\varepsilon}) = \sigma^2 \mathbb{1}$

$\vec{\varepsilon} \sim N_n(\vec{0}, \sigma^2 \mathbb{1})$

Modello Adattato

$\mathbb{E}[\hat{\vec{\varepsilon}}] = 0$

$Cov(\hat{\vec{\varepsilon}}) = \sigma^2 (\mathbb{1} - H)$

$\hat{\vec{\varepsilon}} \sim N_n(\vec{0}, \sigma^2(\mathbb{1} - H))$

Osservazione

La varianze di ε_i $\quad Var[\varepsilon_i] = \sigma^2$

$$Var[\hat{\varepsilon}_i] = \sigma^2 (1 - h_{ii})$$

h_{ii} = leverage dell'i-esimo dato
se è troppo alto ($h_{ii} = 1$)
avremmo varianza del residuo nulla
→ il modello viene fatto passare da lì,
il punto sta facendo leva x spingere il modello adattato a passare x quel punto, e non va bene.

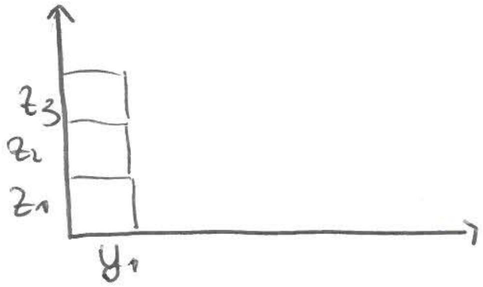
Residui Studentizzati

$$\hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i}{\sqrt{s^2(1 - h_{ii})}}$$ ← stima variabilità $\hat{\varepsilon}_i$

Grafici o plot di diagnostica

Lo STEP 1 è fare un plot dei dati: se non guardi i dati non puoi sapere se stai x andare sulla strada giusta.

plot(z,Y)
pers ← se hai tanti predittori fa gli scatterplot di tutte le combinazioni delle varie variabili



Quando le variabili che spiegano in modo lineare la risposta

STEP 2: GRAFICI dei RESIDUI

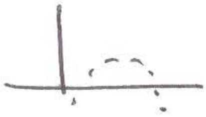
scatterplot $(\hat{\varepsilon}_i , \hat{z}_{ik})$
↑ predittore

scatterplot $(\hat{\varepsilon}_i , y_i)$
↑ previsti

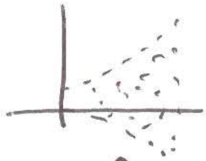
scatterplot $(\hat{\varepsilon}_i , i)$

Quale che sia il grafico dei residui,
vogliamo vedere una nuvola casuale rispetto all'asse 0

non vogliamo vedere



né



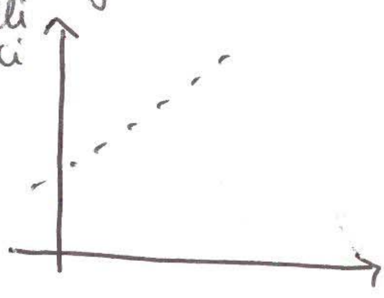
xk la varianza nel nostro modello era = x tutte

↑ eteroschedasticità

se usiamo residui studentizzati li abbiamo già riscalati
e ~~non~~ sono ≃ dello stesso ordine di grandezza

STEP 3: QQ PLOT

modo grafico x capire se i residui sono normali.



se è Gaussiano viene una retta e sono ☺

$X \sim N(\mu, \sigma^2)$

$x_\alpha = \mu + \sigma z_\alpha$
↑ quantili empirici

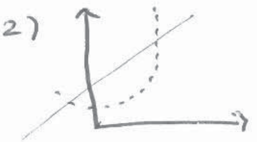
TEST di SHAPIRO:

H_0: dati Gaussiani
H_1: dati non gaussiani

Vorremmo avere pvalue alto,
non aver evidenze x rifiutare H_0

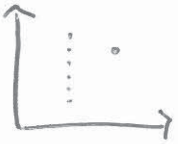
VALUTAZIONE R^2

look for Anscombe quartet in Internet
4 dataset x far vedere che guardare 1 indice da solo non serve a niente.
le stime di $\hat{\beta}_0$ e $\hat{\beta}_1$ viene sempre uguale, e anche $R^2 \simeq 65\%$.
Quindi direi che mi vanno bene tutti.
In realtà vado a vedere i dati e ho

1) 
2) 

-> il plot dei residui avrebbe visto e avrei capito che non andava bene 

3) 
dati perfettamente allineati tranne uno, che fa effetto leva e quindi non va bene

4) 
Anche qst dà un effetto per cui il modello dà la stessa retta, ma ovviamente non va bene.

PUNTI ANOMALI RISPETTO alla nostra descrizione dei dati: outlier

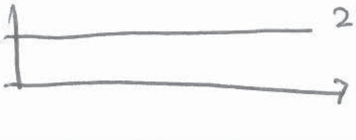
Noi sappiamo
$$\hat{\vec{y}} = Z\hat{\vec{\beta}} = H\vec{y}$$
Allora
$$\hat{y}_i = \sum_{j=1}^{n} h_{ij} y_j = h_{ii} y_i + \sum_{j \neq i} h_{ij} y_j$$

$$H\mathbb{1} = \mathbb{1}$$
$$1 = \sum_{j=1}^{n} h_{ij} \Rightarrow \text{la somma degli elementi di } h \text{ ci dà } 1.$$

$$\sum_{i=1}^{n} h_{ii} = tr(H) = r+1 \leftarrow$$ la traccia di un proiettore ci dà il rango.

Se h_{ii} è alto sto forzando il modello a passare x q.l punto, quindi lo forzo a funzionare di più il resto.

Supponiamo che tutti i punti hanno $\simeq$ lo stesso leverage.
$$\Rightarrow h_{ii} \simeq \frac{r+1}{n}$$
Regola del pollice: plotto gli h_{ii} e metto una retta $\frac{2(r+1)}{n}$ e i punti sopra qst retta sono quelli influenti


$\frac{2(r+1)}{n}$

(134) Se ho una leva molto alta,
oppure ho un outlier, cosa faccio?
Dipende da cosa sto studiando.
Devo vedere se è lampante che il dato sia un errore.
Magari è un errore di input, è il dataset che è sbagliato

$$1 - R^2 = \frac{SS_{err}}{SS_{tot}}$$

~~[illegible]~~ se stiamo confrontando un R^2 a parità di predittori,
SS_{err} diminuisce,
R^2 aumenta se continuo ad aggiungere predittori,
non centra se migliora il modello.

$R^2_{adj} \Rightarrow$ $\boxed{1 - R^2_{adj} = \frac{SS_{err} / (n-k-1)}{SS_{TOT} / (n-1)}}$

Si può dimostrare

$$R^2_{adj} = 1 - \frac{SS_{err} / (n-k-1)}{SS_{TOT} / (n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

$\downarrow$

Penalizzo l'R^2

peso il numero di predittori

Allo statistico non importa l'interpolazione,
ma interessa che i miei dati siano
generati da un modello + errore

Press: $\sum \frac{\hat{e}_i}{1-h_{ii}}$ lo voglio basso

C_p Mallows

} indici che si possono andare a vedere.

<u>Esempio</u>: dataset con k predittori

Y, Z

Quanti sottomodelli potrei costruire?

$\forall\ k = 1 \dots n$ ho $\binom{n}{k}$ possibili sottomodelli

Scelgo per ogni k un modello migliore M_k^*
e poi confronto i modelli tra i vari k,
e scelgo il modello finale.
Ma è una procedura computazionalmente
troppo onerosa.

Metodi Stepwise

1) **FORWARD SELECTION**:
aggiungo iterativamente un predittore alla volta.
(parto da modello 1)

2) **BACKWARD SELECTION**:
modello con tutti i predittori e ne tolgo uno alla volta

3) **BOTH**
fa un po' di backward e un po' di forward

$M_k(i)$, $M_{k'}(i+1)$

i: indice dell'iterazione — ho un k diverso

pensiamo $k' > k$.

Posso fare il test F: — l'SS_{res} è + grande quello del modello con meno predittori

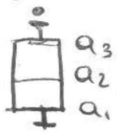
$$F = \frac{SS_{res}(M_k(i)) - SS_{res}(M_{k'}(i+1))}{(k'-k)\left(\frac{SS_{res}(M_{k'}(i+1))}{n-k'-1}\right)}$$

Mi fermo quando ho un pvalue alto. — quale modello $M_{k'}$

Spesso usando 1, 2 o 3 non ottengo lo stesso modello, non è che mi portano entrambe al modello ottimo.
Dobbiamo sempre pensare a cosa sto mettendo o togliendo dal modello

Outliers

punti strani, diversi



$a_3 - a_1 = IQR$

confini a $a_3 + 1,5\,IQR$
$a_1 - 1,5\,IQR$

e i punti fuori da qst erano outlier, di coda, non stavano vicino ai dati.

Alcuni outlier vengono generati con prob. bassa.

$Z \sim N(0,1)$

Posso calcolare: $z_{0,25}$, $z_{0,75}$, IQR — $l = z_{0,25} - 1,5\,IQR$

$u = z_{0,75} + 1,5\,IQR$

e poi $P(Z > u) + P(Z < l)$ — probabilità di generare dati che sono outlier.
$\simeq 0.007$

(136) Ho probabilità non trascurabile che il modello generi un outlier.

Oppure
$(1-\varepsilon)$ genero da $z \sim N(0,1)$

e con prob. ε perturbo $x \sim \mathcal{E}(1)$

qst dati sono outlier in un senso che sono generati da un altro modello.

y_i sia un outlier.

$$\begin{cases} y_j = \vec{z}_j^T \vec{\beta} + \varepsilon_j & j \neq i \\ y_i = \vec{z}_i^T \vec{\beta} + \delta + \varepsilon_i \end{cases}$$

modello con intercetta $\neq$.

Posso testare la probabilità che y_i sia un outlier?

$U = [u_1 \ldots u_n]^T$ con $u_j = 0$ $j \neq i$
$u_i = 1$

vettore che mi indica la riga del potenziale outlier.

$\tilde{Z} = [Z, U]$

$\vec{Y} = \tilde{Z} \vec{\beta}' + \vec{\varepsilon}$ $\quad \vec{\beta}' = (\beta_0, \beta_1, \ldots \beta_r, \delta)$

So fare test marginali x dire se un certo coeff è $=0$ o $\neq 0$.

$H_0 : \delta = 0$
$H_1 : \delta \neq 0$

$$\left(\underbrace{\hat{\beta}'_{r+1}}_{\text{stimatore del } \delta} - \underbrace{0}_{\text{vero media sotto } H_0} \right) \Big/ \underbrace{\sqrt{S^2 \, diag_{r+2} (\tilde{Z}^T \tilde{Z})^{-1}}}_{\text{variabilità}} \overset{\text{sotto } H_0}{\sim} t(n-r+2)$$

Costruisco il mio test e posso concludere se

ACCETTO $H_0 \Rightarrow y_i$ non è outlier

RIFIUTO $H_0 \Rightarrow y_i$ è un outlier.

DATI INFLUENTI

$Z_{(i)}$ matrice a cui tolgo la riga i

Quindi ho matrice $(n-1)\times(n+1)$.

Posso fittare $\to [(n+1)\times(n-1)][(n-1)\times(n+1)] \Rightarrow [(n+1)\times(n+1)]$

$$\vec{\beta}_{(i)} = \left(Z_{(i)}^T Z_{(i)}\right)^{-1} \underbrace{Z^T_{(i)}}_{(n+1)\times(n-1)} \underbrace{\vec{y}_{(i)}}_{(n-1)}$$

dimensionalmente viene ancora n+1.

Quindi ho le mie due stime.

Se il punto i è ininfluente

$\vec{\beta}$ e $\vec{\beta}_{(i)}$ devono essere vicine.

β è la stima dei parametri.

Come misuro la distanza?

DISTANZA di COOK (1977)

$$\boxed{D_i = \sqrt{\frac{\left(\hat{\vec{\beta}}_{(i)} - \hat{\vec{\beta}}\right)^T Z^T Z \left(\hat{\vec{\beta}}_{(i)} - \hat{\vec{\beta}}\right)}{(n+1)\, s^2}}}$$

↑ fatto su tutto il modello

Non misura la distanza euclidea xk deve tenere conto della struttura di varianza e cov, che di fatto è tenuta conto in $\frac{Z^T Z}{s^2}$

Si costruiscono le linee di livello di Di.

Perché è questa la distanza giusta? 29 maggio 2015

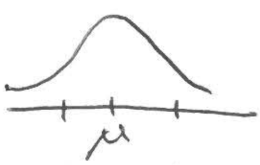
Se noi prendiamo X gaussiano

Abbiamo la distanza di Malanovis

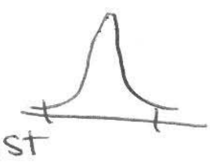
$(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})$ Σ è definita positiva

↳ la forma quadratica induce la distanza

↑ è la distanza giusta x calcolare la distanza tra la realizzazione di un dato e la media.



due punti su qst hanno distanza + da



I punti + distanti dalla media sono quelli che stanno a 4σ, non sono $+\infty$ e $-\infty$.

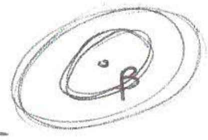
I punti a 4σ sono tutti alla stessa distanza dalla loro media.

Devo tener conto della matrice varianza-covarianza

$\vec{\hat{\beta}} \sim N(\vec{\beta}, \sigma^2 (Z^T Z)^{-1})$

L'inverso della matrice $(Z^T Z)^{-1} \sigma^2$ è $\frac{Z^T Z}{\sigma^2}$,
ma σ^2 non lo so quindi lo stimo.

Posso costruirmi le curve di livello e definire un punto non influente se è fuori da una data curva di livello



OSSERVAZIONE

$$\vec{\hat{Y}} = Z \vec{\hat{\beta}}$$

$$D_i = \frac{(\vec{Y}_{(i)} - \vec{\hat{Y}})^T (\vec{Y}_{(i)} - \vec{\hat{Y}})}{(k+1) S^2}$$

distanza quasi euclidea tra le osservazioni, ma ho S^2 a denominatore.

È di fatto la distanza di Mahalanobis sulle $Y_{(i)}$.

Si può dimostrare che

$$D_i = \frac{1}{(k+1)} r_i^2 \left(\frac{h_{ii}}{1 - h_{ii}} \right)$$

$$r_i = \frac{\hat{\varepsilon}_i}{S\sqrt{1 - h_{ii}}}$$ residuo studentizzato

Punti con alto leverage fanno esplodere questa distanza e quindi fanno esplodere la distanza di Cook ⇒ punti influenti.

E SE i TEST di DIAGNOSTICA NON FUNZIONA?

Rifittiamo il modello e riproviamo.

ETEROSCHEDASTICITÀ dei RESIDUI

Si risolvono con
- REGRESSIONE con PESI

o

- TRASFORMAZIONI di Y cambiando la varianza.

$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$

trasformo xk scopro che ε non è gli giusto.
Allora faccio

$\sqrt{\vec{Y}} = Z\vec{\beta} + \vec{\varepsilon}$, e vedo che va bene.
Però mi devo ricordare

Suggerimenti:

- SE $Var[\varepsilon_i] \propto E[Y_i]$
  prendi una trasformazione tipo $\sqrt{Y}$ o $\sqrt{Y+1}$
- SE $Var[\varepsilon_i] \propto E[Y_i]^2 \Rightarrow \log(Y)$
- SE $Var[\varepsilon_i] \propto E[Y_i]^4 \Rightarrow \frac{1}{Y}$ o $\frac{1}{Y+1}$

la trasformazione logaritmica viene spesso usata se la variabilità delle Y è molto grande

Supponiamo di avere scoperto

$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$

$Cov(\vec{\varepsilon}) = \sigma^2 \Sigma$

Supponiamo

$\exists\, C$ tc $C^T C = CC^T = \Sigma^{-1}$ (pseudo radice quadrata di Σ^{-1})

$$Var(C\vec{\varepsilon}) = C\, Var(\vec{\varepsilon}) C^T = $$
$$= C\, \sigma^2 \Sigma C^T = \sigma^2 C \underbrace{(CC^T)^{-1}}_{(\Sigma^{-1})^{-1}} C^T = \sigma^2 \mathbb{1}$$

Così riporto ad avere un comportamento omoschedastico $C\vec{\varepsilon}$.

Applico C a tutto il mio modello:

$C\vec{Y} = CZ\vec{\beta} + C\vec{\varepsilon}$

se chiamo $\vec{W} = C\vec{Y}$, $\tilde{Z} = CZ$, $\vec{\tilde{\varepsilon}} = C\varepsilon$

ho modello della forma

$\vec{W} = \tilde{Z}\vec{\beta} + \vec{\tilde{\varepsilon}}$ così $\vec{\tilde{\varepsilon}}$ rispetta l'omoschedasticità.

Se fitto il modello

$$\hat{\vec{\beta}} = (\tilde{Z}^T \tilde{Z})^{-1} \tilde{Z}^T \vec{W} = ((CZ)^T CZ)^{-1} (CZ)^T (C\vec{Y}) =$$
$$= (Z^T \underbrace{C^T C}_{\Sigma^{-1}} Z)^{-1} (Z)^T C^T C\, Y = (Z^T \Sigma^{-1} Z)^{-1} Z^T \Sigma^{-1} Y$$

$$= (Z^T \Sigma^{-1} Z)^{-1} (Z^T Z^{-1} Y)$$

regressione con peso,
peso dato da Σ^{-1}

Se Σ ha struttura diagonale, tipo

$$\Sigma = \begin{pmatrix} \frac{1}{w_1} & & \\ & \ddots & \\ & & \frac{1}{w_n} \end{pmatrix}$$

se per es. Y vengono da una media di determinate misure.

Y ~~prove~~ è media aritmetica di misure con lo stesso σ^2 dallo stesso campione.
Se le numerosità sono $\neq$, la variabilità dipenderà in modo inversamente proporzionale dalla numerosità.

$$\tilde{Z} = \begin{bmatrix} \sqrt{w_1} & \sqrt{w_1}\, z_{11} & \\ \vdots & \vdots & \cdots \\ \sqrt{w_n} & \sqrt{w_n}\, z_{n1} & \end{bmatrix} \qquad \vec{W} = \begin{bmatrix} \sqrt{w_1}\, y_1 \\ \vdots \\ \sqrt{w_n}\, y_n \end{bmatrix}$$

peso le mie osservazioni $\Rightarrow$ REGRESSIONE CON PESO.

Anche nel caso di violazione di normalità.

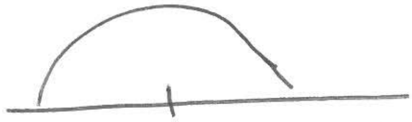
RESIDUI NON GAUSSIANI

mi saltano gli intervalli di confidenza,
i $\hat{\beta}$ non sono più GAUSSIANI

Si risolve con
- TRASFORMAZIONE BOX-COX

$$\begin{cases} y' = \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ y' = \log y & \lambda = 0 \end{cases}$$

trasformo o con polinomio o con logaritmo.
Come scelgo il λ?
λ: massima verosimiglianza gaussiana di y'



λ_{max} mi massimizza la gaussianità dei dati.

Se $\lambda = 0.47$ non farò Y a 0.47, ma la $\sqrt{Y}$, xk poi il dato lo devi raccontare,
se $\lambda \simeq 1$ significa che i dati erano già abbastanza gaussiani.

Così rischio di perdere l'omoschedasticità dei dati

Di solito se non ho residui gaussiani o non ho schedasticità trasformando Y si risolvono i problemi

Regressione Polinomiale (o con Interazione)

$$\hat{\vec{Y}} = \beta_0 + \beta_1 \vec{z}_1 + \beta_2 \vec{z}_1^{\,2} + \dots + \beta_d \vec{z}^{\,d} + \vec{\varepsilon}$$

se per es. Y è superficie e z_2 raggio.
Oppure xk vedi che può servire

È ancora un modello lineare in cui la matrice disegno sono

$$\begin{bmatrix} 1 & z & z_2^2 & \dots \\ & & & \end{bmatrix}$$

Se vado su con d posso interpolare $\simeq$ qualsiasi nuvola di punti.
I dati, elevandoli al quadrato, cubo etc. esplodono,
se già ho range $\neq$ di $z_1, z_2 \dots$

A volte si tiene conto delle iterazioni.

Supponiamo di avere

$$Y = \beta_0 + \beta_1 \vec{z}_1 + \beta_2 \vec{z}_2 + \beta_{11} \vec{z}_1^{\,2} + \beta_{22} \vec{z}_2^{\,2} + \underbrace{\beta_{12} \vec{z}_1 \cdot \vec{z}_2}_{\text{termine di interazione}} + \vec{\varepsilon}$$

Ho aggiunto un'interazione tra le due variabili

Per capirne il significato
supponiamo di guardare un modello semplificato

$$\vec{Y} = \beta_0 + \beta_1 \vec{z}_1 + \beta_2 z_2 + \beta_{12} z_1 z_2 + \varepsilon$$

i test mi dicono che i termini al quadrato posso buttarli via, ma non posso buttare via $\beta_{12} z_1 z_2$

Fissiamo z_2 e spostiamo z_1 in z_1+1.

$$Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_{12} z_1 \cdot z_2$$

ora trasformo

$$Y' = \beta_0 + \beta_1 (z_1 + 1) + \beta_2 z_2 + \beta_{12} (z_1 + 1) z_2$$

$$\Delta Y = \cancel{\beta_0} + \cancel{\beta_1 (z_1)} + \beta_1 + \cancel{\beta_2 z_2} + \beta_{12} z_1 z_2 + \cancel{\beta_{12} z_2} -$$
$$- \cancel{\beta_0} - \cancel{\beta_1 z_1} - \cancel{\beta_2 z_2} - \beta_{12} z_2 =$$

$$= \beta_1 + \beta_{12} z_2$$

$\Rightarrow$ SE C'È INTERAZIONE l'incremento che si ha nella risposta dipende da z_2.

È ANCORA + facile vedere qst interazione quando una delle due variabili è CATEGORICA.

$$Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_{12} z_1 z_2$$

sia z_2 una variabile CATEGORICA.

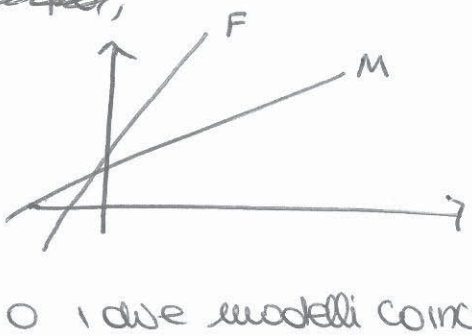
$$z_2 = \begin{cases} 0 & \text{maschio} \\ 1 & \text{femmina} \end{cases}$$

$\Downarrow$

$$y_i^M = \beta_0 + \beta_1 z_1$$

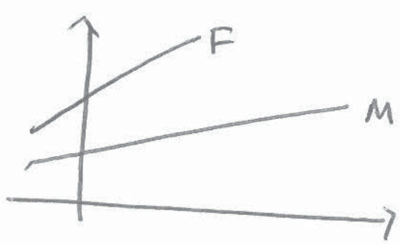
$$y_i^F = \beta_0 + \beta_1 z_1 + \beta_2 + \beta_{12} z_1 = (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) z_1$$

se $\beta_2 = 0$
$\beta_{12} \neq 0$

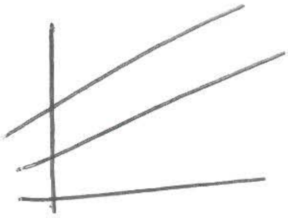


se $\beta_2 = \beta_{12} = 0$ i due modelli coincidono

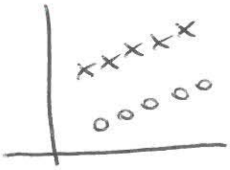
se $\beta_2 \neq 0$
$\beta_{12} \neq 0$



$\beta_2 \neq 0$
$\beta_{12} = 0$

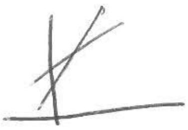


Se ho una covariata continua e una categorico
qnd faccio lo scatterplot faccio x per gruppo 1, o x gruppo 2



se i modelli sono $\simeq$ //
mi aspetto che l'interazione sia 0.

L'incremento al variare di un unità della mia risposta invece
dipende da z_2 se ho pendenza ≠,
sarà quale delle due rette che sto guardando che mi dice
di quanto sale



Se la matrice Z è altamente collineare

$(Z^T Z)^{-1}$ mi diventa un problema.
Sia xk è instabile
Sia xk la variabilità dei $\hat{\beta}$ diventa enorme.
Come mi accorgo che ho collinearità?

VIF $\quad z_1 \dots z_n$

↑
quanto il fattore influenza la varianza

faccio un modello dove

$$z_j = \gamma_0 + \gamma_1 z_1 + \dots + \gamma_{j-1} z_{j-1} + \gamma_{j+1} z_{j+1} + \dots + \gamma_r z_r$$

↓ R^2_j

$$VIF_j = \frac{1}{1 - R^2_j}$$

Andando a calcolare il VIF posso vedere quale è
più correlata alle altre.

(144) FARE RIDGE-REGRESSION

mentre nelle tecniche di Shrinkage (restringimento)

Centriamo i dati ⇒ prendiamo le colonne centrate ⇒ $\beta_0 = 0$

$$\hat{\vec{\beta}}_{LS} = \underset{\vec{b}}{\operatorname{Argmin}} \ \|\vec{y} - Z\vec{b}\|^2 = \operatorname{Argmin} (\vec{y} - Z\vec{b})^T(\vec{y} - Z\vec{b})$$

Fare Ridge Regression: mettere un vincolo, una penalizzazione xk altrimenti le mie stime possono esplodere.

$$\begin{cases} \underset{\beta}{\operatorname{Argmin}} \ \|\vec{y} - Z\vec{b}\|_2^2 \\ \|\vec{\beta}\|_2^2 \leq s \quad \Leftarrow \text{vincolo} \end{cases}$$

⇓

Ho problema di minimo vincolato

Formulazione lagrangiana

$$L(\vec{\beta}, \lambda) = (\vec{y} - Z\vec{\beta})^T(\vec{y} - Z\vec{\beta}) + \lambda(\vec{\beta}^T\vec{\beta} - s)$$

$$\begin{cases} \dfrac{\partial L}{\partial \vec{\beta}} = -2Z^T(\vec{y} - Z\vec{\beta}) + \lambda\vec{\beta} = 0 \\ \dfrac{\partial L}{\partial \lambda} = \vec{\beta}^T\vec{\beta} - s = 0 \end{cases}$$

$$\begin{cases} Z^T\vec{y} = (Z^TZ + \lambda\mathbb{1})\vec{\beta} \\ \vec{\beta}^T\vec{\beta} = s \end{cases}$$

trovo un λ per cui è soddisfatto $\beta^T\beta = s$.

fissato s

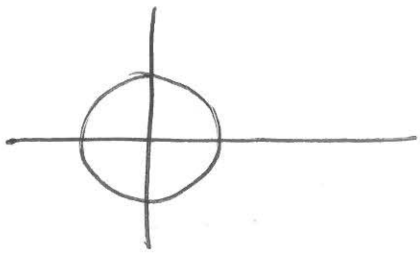
$$\begin{cases} \hat{\vec{\beta}}_{RR} = (Z^TZ + \lambda\mathbb{1})^{-1} Z^T \vec{Y} \\ \|\hat{\vec{\beta}}_{RR}\| \leq s \end{cases}$$

Così ho risolto i miei problemi,
xk $(Z^T Z + \lambda \mathbb{1})$ ora è invertibile e posso quindi
fare le inversione.

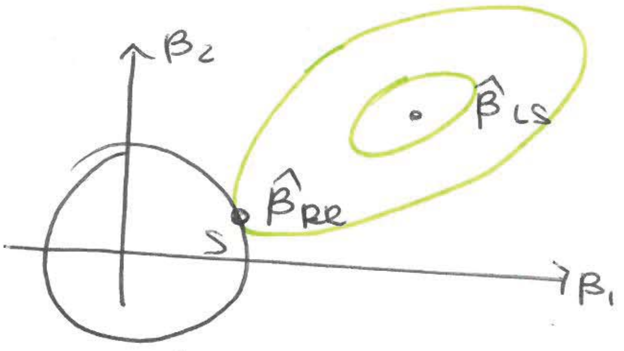
$$\vec{\beta}_{RR} = \text{Arg}\min \|\vec{y} - Z\vec{b}\|^2 + \lambda(s)\|\vec{b}\|_2^2$$

più è alta la penalità sulla norma
tanto più vincolo il mio problema a
stare in una palla piccola

$\lambda \downarrow 0 \quad s \uparrow +\infty \quad \hat{\vec{\beta}}_{RR} = \hat{\vec{\beta}}_{LS}$



se è vincolato
penalizzo molto
la norma L^2.

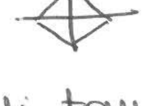


La soluzione allora del problema ridge
sarà il punto di contatto delle
regioni di confidenza e il vincolo

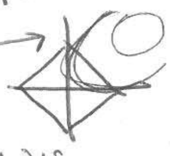
~~Pertanto~~ Però $\vec{\beta}_{RR}$ sono distorti

Mettere il vincolo sulla norma L^2 è una delle possibili.

Regressione LASSO: vincolo che $\quad$ vincolo la norma 1
$\sum |\beta_j| < s$



la ridge è + regolare.
La lasso invece è popolare xk i punti di tangenza
vanno a finire negli angoli
Quindi $\beta_1 = 0$



Quindi sto in qualche senso
facendo selezione delle variabili.

[A. Agresti, "categorical data analysis"] 4 giugno 2015

Modelli Lineari Generalizzati GLM

Modelli lineari con risposta Y non Gaussiana.

$Y_i \sim Be(p_i)$ è il caso + importante

$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}$

Vanno specificate 3 componenti

1) Random Component: Y v.a. risposta con la sua legge
2) Componente sistematica: Z matrice dei predittori deterministica
3) Link function: specifica quale funzione della $E[Y]$ vogliamo spiegare con la componente sistematica

(nel caso gaussiano la link function è l'identità: $E[Y] = Z\beta$)

Per ogni tipo di variabile c'è una link function privilegiata, ma non è l'unica.

1) Y v.a. risposta

$(y_1 \dots y_n)$ da $Y_1 \dots Y_n$ indipendenti appartenenti alla famiglia esponenziale:

$$f(y_i, \theta_i) = a(\theta_i)\, b(y_i) \exp[y_i \underbrace{Q(\theta_i)}_{\text{Parametro naturale}}]$$ ← i parametri cambiano al cambiare della y.

2) $Z\vec{\beta} = (\eta_1 \dots \eta_n)^T$ componente sistematica

η_i è il predittore lineare $\quad \eta_j = \sum_{i=0}^{n} \beta_i z_{ji} \quad j = 1 \dots n$

$\vec{\beta} \in \mathbb{R}^{n+1}$ incogniti

3) link function

$\mu_i = E[Y_i]$

$g: \quad g(E[Y_i]) = \eta_i$

↑ prendendo la media la parte aleatoria viene tolta.

$g(\cdot) = Q(\cdot)$ LINK FUNCTION CANONICA

ESEMPIO : BERNOULLI

$Y_i \sim Be(p_i)$

$\mu_i = p_i$

$$f(y, p) = p^y (1-p)^{1-y} \mathbb{1}_{\{0,1\}}(y) = (1-p)\mathbb{1}_{\{0,1\}}(y) \exp\left[y \log\left(\frac{p}{1-p}\right)\right]$$

La link function canonica è

$$\text{logit}(p_i) = \log \frac{p_i}{1-p_i}$$

È QUESTO IL PARAMETRO NATURALE della BERNOULLI.

Quindi il mio modello cercherà di spiegare

$\text{logit}(p_i) = \eta_i$

qui dentro ho i miei parametri incogniti β.

(η_i = riga i di $Z\beta$)

η_i ha prima può avere range da $-\infty$ a $+\infty$

p_i va da 0 a 1

ma la funzione $\text{logit}(p) : [0,1] \to [-\infty, +\infty]$

Quindi i valori ammissibili di logit sono compatibili con quelli del predittore

ESEMPIO : POISSON

$Y \sim P(\mu)$

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{y!} \mathbb{1}_N(y) = \frac{e^{-\mu} \mathbb{1}_N(y)}{y!} \exp[y \log \mu]$$

$\Rightarrow$ la link function è

$\log(\mu_i) = \eta_i$ — cercherò di spiegare il log della media con η_i

Ci servono degli indici per confrontare i modelli

AIC (Indice di Informazione di AKAIKE)

$-2 \ln(L) + 2K$ (L: log verosimiglianza) — K: numero dei parametri del modello

Vorrei aumentare la verosimiglianza, ma se x farlo aumento K non va benissimo

Voglio massimizzare la verosimiglianza ma avere pochi parametri.

smaller is better

DEVIANZA

$$\vec{y} = (y_1 \dots y_n)^T$$

Pensiamo alla logverosimiglianza come

$$\ell(\vec{\mu}, \vec{y})$$

Sia

$\ell(\hat{\vec{\mu}}, \vec{y})$ la log-verosimiglianza del modello

Tra tutti i possibili modelli chiamiamo

MODELLO SATURATO $\hat{\vec{\mu}} = \vec{y}$

possiamo calcolare

$\ell(\vec{y}, \vec{y})$ è la massima verosimiglianza tra tutti i modelli

stimo la media con l'osservazione, così massimizzo la logverosimiglianza

Quindi la posso vedere come Gold standard, posso paragonare quanto un modello dista da lei.

Chiamo DEVIANZA

$$-2\left[\ell(\hat{\vec{\mu}}, \vec{y}) - \ell(\vec{y}, \vec{y})\right]$$

e la uso per confrontare due modelli

È meglio il modello con devianza + bassa

$\text{logit}(p_i) = \eta_i$

Z $n \times 2$: $\beta_0 + \beta_1 z$

Voglio spiegare $\text{logit}(p(z)) = \beta_0 + \beta_1 z$

p_i sarebbe p che comunque dipende dalla covariata z.

$$\frac{p(z)}{1 - p(z)} = \exp[\beta_0 + \beta_1 z]$$

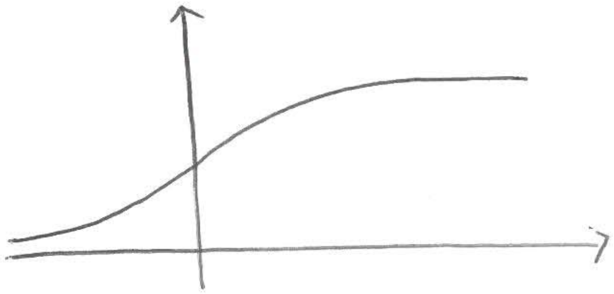
$$\Updownarrow$$

$$p(z) = \frac{\exp[\beta_0 + \beta_1 z]}{1 + \exp[\beta_0 + \beta_1 z]}$$

così è abbastanza chiaro il ruolo di β_0 e β_1.

$\beta_1 > 0 \rightarrow$ monotona crescente con z.

(anche xk è l'inversa del logit che lo era)

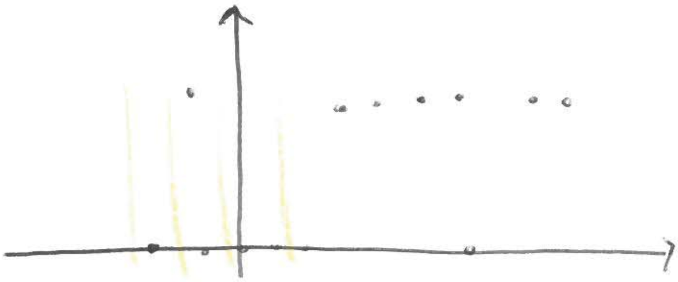


Ha l'ANDAMENTO di UNA FUNZIONE di RIPARTIZIONE!

Infatti a volte al posto del logit si usa $\phi(p_i)$

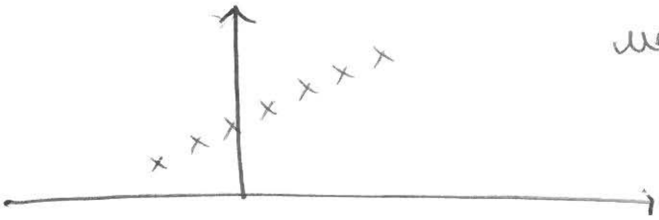
probit : logit $(p_i) \to \phi_z(p_i)$

Posso voler spiegare il mio successo con una variabile continua (età e probabilità di avere infarto)



voglio modellizzare logit(p_i)

Divido la mia variabile in fasce e calcolo la media per fasce



media x fasce

Cerco la migliore funzione di x che mi modellizza la probabilità di avere un attacco cardiaco

Metodi per Stimare i Parametri Incogniti

$$f(y_i, \theta_i, \phi) = \exp\left[(y_i \theta_i - b(\theta_i))/a(\phi) + c(y_i, \phi) \right]$$

ϕ parametro di dispersione, non dipende dal predittore lineare ma è = x tutte
(x le Gauss è la varianza)

ϕ : parametro di dispersione
θ_i : parametro naturale

se ϕ nota

$$Q(\theta) = \frac{\theta}{a(\phi)}$$

$$a(\theta_i) = \exp\left[- \frac{b(\theta_i)}{a(\phi)} \right]$$

$$b(y_i) = \exp c(y_i, \phi)$$

Ma noi ci manteniamo nella generalità con ϕ non noto

Andiamo a lavorare su f per stimare i parametri del modello.

Stimiamo la log verosimiglianza:

$$l_i = [y_i \cdot \theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)]$$

$$\frac{\partial l_i}{\partial \theta_i} = [y_i - \cancel{b'(\theta_i)}]/a(\phi)$$

$$\frac{\partial^2 l_i}{\partial \theta_i^2} = - \frac{b''(\theta_i)}{a(\phi)}$$

$$\mathbb{E}\left[\frac{\partial l_i}{\partial \theta_i} \right] = 0 \qquad \left(\text{è } \int \frac{\partial l_i}{\partial \theta} = \int \frac{\partial}{\partial \theta} \log f \cdot f = 0 \right.$$

scambiando derivata e integrale viene $\frac{\partial}{\partial \theta} 1 = 0$)

$$\mathbb{E}\left[- \frac{\partial^2 l}{\partial \theta_i^2} \right] = \mathbb{E}\left[\left(\frac{\partial l}{\partial \theta} \right)^2 \right]$$ informazione di Fisher

$$\mathbb{E}\left[\frac{\partial l_i}{\partial \theta_i}\right]=0 \iff \mathbb{E}[Y_i]=b'(\theta_i)$$

Quindi qst è la varianza

$$-\mathbb{E}\left[\frac{\partial^2 l_i}{\partial \theta_i^2}\right]=\frac{b''(\theta_i)}{a(\phi)}=\mathbb{E}\left[\left(\frac{\partial l_i}{\partial \theta_i}\right)^2\right]=\mathbb{E}\left[\left(\frac{Y_i-\underbrace{b'(\theta_i)}_{\text{qsi è } E[Y_i]}}{a(\phi)}\right)^2\right]=$$

$$=\frac{1}{a^2(\phi)}\,\text{Var}[Y_i]$$

Quindi riconosco che

$$\begin{cases}\mathbb{E}[Y_i]=b'(\theta_i)\\ \text{Var}[Y_i]=b''(\theta_i)\cdot a(\phi)\end{cases}$$

La funzione b e le sue derivate determinano media e varianza della Y.

ESEMPIO

$Y_i \sim P(\mu_i)$

$$f(y_i,\mu_i)=\exp\Big[y_i\underbrace{\log\mu_i}_{\theta_i}-\underbrace{\mu_i}_{b(\theta_i)}-\underbrace{\log(y_i)}_{c(y_i,\phi)}\Big]$$

$a(\phi)$ = identità

$\theta_i=\log(\mu_i)$

Se

$\mu_i=b(\theta_i) \Rightarrow b(\theta_i)=\exp[\log(\mu_i)]$

Quindi $b(\cdot)=\exp(\cdot)$

$\Downarrow$

$E[Y_i]=\exp(\theta_i)=\mu_i$

$\text{Var}[Y_i]=b''(\theta_i)=\exp(\theta_i)=\mu_i$

Quindi ritrovo i parametri media e varianza della Poisson.

Fare x la Bernoulli

152 La nostra Likelihood è

$$\ell(\vec{\beta}) = \sum_{i=1}^{n} \log f(y_i, \theta_i, \phi)$$

i parametri incogniti $\vec{\beta}$ sono dentro il θ_i

$$= \sum_{i=1}^{n} \left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} \right) + \sum_i c(y_i, \phi)$$

ora voglio derivare rispetto a β_i porre $=0$ e trovare i candidati di massima verosimiglianza.

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j} = 0 \qquad \forall j = 0 \dots r$$

trovo così gli MLE di $\vec{\beta}$

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$

REGOLA CHAIN RULE x la derivata

predittore lineare: $\eta_i = g(\mu_i) = \sum_j z_{ij} \beta_j$

- $\frac{\partial \ell_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)}$

  ma $b'(\theta_i) = E[Y_i]$

- $\frac{\partial \mu_i}{\partial \theta_i} = \frac{\partial b'(\theta_i)}{\partial \theta_i} = b''(\theta_i) = \frac{Var[Y_i]}{a(\phi)}$

- $\frac{\partial \mu_i}{\partial \eta_i} =$

  dipende da g xk $g(\mu_i) = \eta_i$

- $\frac{\partial \eta_i}{\partial \beta_j} = z_{ij}$

Globalmente:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a(\phi)} \frac{a(\phi)}{Var[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} z_{ij} = \frac{(y_i - \mu_i) z_{ij}}{Var[Y_i]} \frac{\partial \mu_i}{\partial \eta_i}$$

La mia stima di massima verosimiglianza,

$\hat{\beta}_{MLE}$ risolvono

$$\boxed{\sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{Var(Y_i)} \right) z_{ij} \frac{\partial \mu_i}{\partial \eta_i} = 0 \qquad \forall j = 0, 1 \dots n}$$

non sostituisco xk dipende da g

I β ci sono dentro il fatto che

$$\mu_i = g^{-1}(\eta_i) = g^{-1}\left(\sum_j z_{ij} \beta_j\right)$$

Nel caso Gaussiano, con g identità, mi torna tutto come sempre

Quasi in tutti questi problemi
dovremo usare metodi numerici x risolvere le

EQUAZIONI di VEROSIMIGLIANZA

Le Y compaiono con media e varianza.

9 giugno 2015

Y
$Z\vec{\beta}$
g link function $\quad g(\mu_i) = Z\vec{\beta}$

$$\log L(\vec{\beta}) = \sum_{i=1}^{n} \log f(y_i, \theta_i, \phi) = \sum_{i=1}^{n} \frac{y_i - b(\theta_i)}{a(\phi)} + \sum c(y_i, \phi)$$

qui ci sono i nostri β

$\sum \frac{\partial l_i}{\partial \beta_j} = 0$ ci dà gli stimatori di massima verosimiglianza

EQUAZIONI di VEROSIMIGLIANZA

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)}{Var[y_i]} z_{ij} \underbrace{\frac{\partial \mu_i}{\partial \eta_i}}_{\text{che cos'è?}} = 0 \qquad \forall j = 0, \dots, n$$

$Y_i \sim Be(p_i)$

$$\sum_{i=1}^{n} \frac{y_i - p_i}{\underbrace{p_i(1-p_i)}_{\text{Varianza della Bernoulli}}} \cdot z_{ij} \frac{\partial p_i}{\partial \eta_i} = 0$$

Il logit è la link function

$\text{logit}(p_i) = \eta_i$ ← predittore lineare

$$\log \frac{p_i}{1-p_i} = \eta_i \Rightarrow \frac{p_i}{1-p_i} = \exp \eta_i$$

$$\Updownarrow$$

$$p_i(1+\exp \eta_i) = \exp(\eta_i)$$

$$\Downarrow$$

$$p_i = \frac{\exp(\sum_{j=0}^{n} z_{ij}\beta_j)}{1+\exp(\sum_{j=0}^{n} z_{ij}\beta_j)}$$

$$\frac{\partial p_i}{\partial \eta_i} = \left(\frac{\partial}{\partial p_i}\eta_i\right)^{-1} = \left(\frac{\partial}{\partial p_i}\text{logit}\, p_i\right)^{-1} =$$

$$= \left(\frac{\partial}{\partial p_i}\log\frac{p_i}{1-p_i}\right)^{-1} = \left(\frac{1-p_i}{p_i}\,\frac{(1-p_i)+p_i}{(1-p_i)^2}\right)^{-1} = p_i(1-p_i)$$

Quindi le nostre equazioni di verosimiglianza nel caso di una Bernoulli diventano

$$\sum_{i=1}^{n} \frac{(y_i - p_i)}{p_i(1-p_i)} z_{ij}\, p_i(1-p_i) = \sum_{i=1}^{n} (\overset{\text{qui ho i }\beta}{y_i - p_i}) z_{ij} = 0 \quad \forall j = 1 \dots n$$

$$\Updownarrow$$

$$\sum \left(y_i - \frac{\exp(\sum_{j=0}^{n}\beta_j z_{ij})}{1+\exp(\sum_{j=0}^{n}\beta_j z_{ij})}\right) z_{ij} = 0$$

non riusciamo a scriverlo in forma chiusa, β_j = qualcosa

dipende solo dalle osservazioni y_i e dalla matrice disegno

E ottengo i $\vec{\beta}_{MLE}$

ma di fatto non riesco a scriverli

Però in R verrà comunque dato un risultato, questo grazie all'asintotica normalità degli ~~...~~ stimatori di verosimiglianza, così ~~...~~ posso fare test sui β. Però ho bisogno di n grande

$y_i \sim P(\mu_i)$

link function $\log(\mu_i) = \eta_i \iff \mu_i = \exp(\eta_i)$

$$\frac{\partial \mu_i}{\partial \eta_i} = \exp(\eta_i) = \mu_i$$

equazioni di verosimiglianza:

$$\sum_{i=1}^{n} \frac{y_i - \mu_i}{Var(Y_i)} \cdot z_{ij} \cdot \frac{\partial \mu_i}{\partial \eta_i} = 0$$

$$\Updownarrow$$

$$\sum_{i=1}^{n} \frac{y_i - \mu_i}{\underbrace{\mu_i}_{Var(Y_i)}} \cdot z_{ij} \cdot \mu_i = 0$$

$$\Updownarrow$$

$$\sum_{i=1}^{n} (y_i - \mu_i) \cdot z_{ij} = 0$$

↳ i β sono nascosti qui,

$$\Updownarrow$$

$$\sum_{i=1}^{n} \Big(y_i - \exp\big(\overbrace{\sum_{j=0}^{k} \beta_j z_{ij}}^{\eta_i}\big)\Big) \cdot z_{ij} = 0$$

↑

Anche qui la risoluzione di β_j è annidata nell'exp quindi questo crea problemi

Nel caso della gaussiana ritroviamo le equazioni che minimizzano la distanza col predittore lineare.

Quando metto i miei y_i e la matrice disegno Z nel modello generalizzato ottengo i miei $\hat{\vec{\beta}}$

Quindi il mio predittore lineare stimato è $\hat{\eta}_i = \sum_{j=0}^{k} \hat{\beta}_j z_{ij}$

e con il predittore lineare stimo

$$g(\hat{\mu}_i) = \hat{\eta}_i$$

Dunque $\text{logit}(\hat{p}_i) = \hat{\eta}_i$

$$\Rightarrow \hat{p}_i = \frac{\exp(\hat{\eta}_i)}{1+\exp(\hat{\eta}_i)}$$

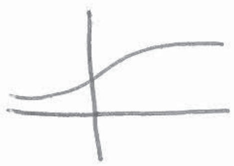
Da cui

$$\hat{y}_i = \frac{\exp\left(\sum_j z_{ij}\hat{\beta}_j\right)}{1+\exp\left(\sum z_{ij}\hat{\beta}_j\right)}$$

produciamo stime $\in [0,1]$, che sono delle probabilità.

Da $\hat{p}_i$ posso voler dire se la previsione $\hat{y}_i$ la prevederò come 0 o come 1

Io osservo degli 0 e 1.



funzione logistica della probabilità di avere effetto collaterale del farmaco in funzione dell'età.

Il modello produce una probabilità in base all'età, ci dirà la probabilità che sia 1.
Per esempio potrò dire che se la probabilità è $> \frac{1}{2}$ allora vedrò la y come 1, se $< \frac{1}{2}$ come 0.

Nel modello di Poisson
le $y_i \in \mathbb{N}$, numero di ~~telefonate con contratto~~ persone in coda tariffazione autostrada.
e lo voglio spiegare con
Avrò

$$\widehat{\log}(\mu_i) = \hat{\eta}_i = \sum_{j=0}^{n} z_{ij}\hat{\beta}_j$$

Quindi la media

$$\hat{\mu}_i = \exp(\hat{\eta}_i) = \exp\left(\sum_{j=0}^{n} z_{ij}\hat{\beta}_j\right)$$

$\hat{\mu}_i$ non ha ragione di $\in \mathbb{N}$,
quindi dovrò capire come valutare il discostamento tra l'osservazione e il μ_i previsto.

$$Cov(\hat{\vec{\beta}}) = (Z^T \hat{W} Z)^{-1}$$ si può dimostrare qst,

con W matrice diagonale con $W_i = \left(\underbrace{\frac{\partial \mu_i}{\partial \eta_i}}\right)^2 \frac{1}{Var(Y_i)}$

nella regress. lineare qste I,
Quindi mi ritrovo la covarianza nel solito modo

CASO BERNOULLI

$$\frac{\partial \mu_i}{\partial \eta_i} = p_i(1-p_i) = Var Y_i$$

$\boxed{\text{REGRESSIONE LOGISTICA}}$

$$W_i = Var(Y_i)$$

Quindi

$$Cov(\hat{\vec{\beta}}) = (Z^T \, diag(\hat{p}_i(1-\hat{p}_i)) Z)^{-1}$$

$$\hat{p}_i = \frac{\exp(\hat{\eta}_i)}{1+\exp \hat{\eta}_i}$$

Quindi quando R mi stima $\hat{\beta}_j$ ~~la~~ ~~f~~ $Se(\hat{\beta}_j)$ dato dall'elemento diagonale j-esimo.

~~Per~~ il modello è "stressato" quando ho grossa la differenza tra 0 e 1, quindi il p stimato è ≈0 o ≈1.

CASO POISSON

$$W = \frac{\mu_i^2}{\mu_i} = \mu_i$$

Quindi

$$Cov(\hat{\vec{\beta}}) = (Z^T \, diag(\hat{\mu}_i) Z)^{-1}$$

$$\exp\left(\sum_j z_{ij} \hat{\beta}_j\right)$$

Come valutare un modello

Indici di GOF - Goodness of fit

*** DEVIANZA**

saturato

$$\tilde{\theta}_i = \tilde{\mu}_i = y_i \qquad \forall i = 1 \dots n$$

i nostri parametri sono tanti quanti le osservazioni

$$\hat{\theta}_i = \hat{\mu}_i$$

sono di meno. ← del modello, stimato

$$D = -2 \left[\underbrace{\ell(\vec{\hat{\mu}}; \vec{y}) - \ell(\vec{y}, \vec{y})}_{\text{misura la discrepanza rispetto al fit migliore}} \right]$$

Nel nostro caso

$$l_i = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)}$$

$$D = 2 \sum_{i=1}^{n} \frac{y_i\tilde{\theta}_i - b(\tilde{\theta}_i)}{a(\phi)} - 2 \sum_{i=1}^{n} \frac{y_i\hat{\theta}_i - b(\hat{\theta}_i)}{a(\phi)}$$

$$= 2 \sum_{i=1}^{n} \frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a(\phi)}$$

anche chiamata devianza scalata

$$-2\log(\lambda(\vec{x})) \to \chi^2_1$$

Cosa diventa la devianza nel caso di Poisson e di Bernoulli?

Poisson

$$\tilde{\theta}_i = \log y_i$$

$$\hat{\theta}_i = \log(\hat{\mu}_i) \quad \Rightarrow \quad D(\vec{y}, \vec{\hat{\mu}}) = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - y_i + \hat{\mu}_i\right]$$

$$b(\theta_i) = \exp(\theta_i)$$

$$a(\phi) = 1$$

pesa la discrepanza tra osservazioni e media e logaritmo del quoziente

BERNOULLI

$$\hat{\theta}_i = \log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)$$

$$b(\theta) = \log(1+\exp(\theta))$$

modello saturato

$$b(\hat{\theta}_i) = \log\left(1 + \frac{\hat{p}_i}{1-\hat{p}_i}\right) = \log\frac{1}{1-\hat{p}_i} = -\log(1-\hat{p}_i)$$

$$\left.\begin{aligned}\tilde{\theta}_i &= \log\left(\frac{y_i}{1-y_i}\right)\\ b(\tilde{\theta}_i) &= -\log(1-y_i)\end{aligned}\right\} \text{modello saturato}$$

$$\Downarrow$$

$$D(\vec{y}, \vec{\mu}) = 2\sum_i \left[y_i \left(\log\left(\frac{y_i}{1-y_i}\right) - \log\frac{\hat{p}_i}{1-\hat{p}_i}\right)\right.$$
$$\left. + \log(1-y_i) - \log(1-\hat{p}_i)\right]$$
$$= 2\sum_i \left[y_i \log\left(\frac{y_i}{\hat{p}_i}\right)\right] + 2\sum_i \left[(1-y_i)\log\frac{1-y_i}{1-\hat{p}_i}\right]$$

Quindi la devianza è in un certo senso totalmente simmetrica, se fossero tutti 0 o tutti 1 una delle due parti andrebbe via

$$\boxed{\text{Prenderò il modello con devianza più bassa}}$$

Se ho due modelli annidati posso direttamente guardare tra i due modelli, senza usare il modello annidato.

M_0 sottomodello di M_1.

$$l(\vec{\hat{\mu}}_0, \vec{y}) \leq l(\vec{\hat{\mu}}_1, \vec{y})$$

$$D(\vec{y}, \vec{\hat{\mu}}_1) = -2\left[l(\vec{\hat{\mu}}_1, \vec{y}) - l(\vec{y}, \vec{y})\right] \leq -2\left[l(\vec{\hat{\mu}}_0, \vec{y}) - l(\vec{y}, \vec{y})\right]$$
$$= D(\vec{y}, \vec{\hat{\mu}}_0)$$

(se prendo un modello più semplice la devianza aumenta)

Faccio un test che dice se i due modelli sono equivalenti o meno.

R restituisce test asintotici legati al fatto che $-2\log(\lambda)$ è asintoticamente una χ^2

(160) DEVIANZE PER LA SINGOLA OSSERVAZIONE

$$D = \sum \underbrace{\left[y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right]} = \sum d_i$$

d_i devianza singola osservazione

$\sqrt{d_i}\ \text{sign}(y_i - \hat{\mu}_i)$ devianze residua

per vedere se sovrastimo o sottostimo l'osservazione

RESIDUI di PEARSON

$$\frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\text{Var}\, y_i}}}$$

Soprattutto per la Bernoulli si usa spesso BRIER'S SCORE

$$\frac{1}{n} \sum_i (y_i - \hat{p}_i)^2$$

AIC

$$-2\left(\ell(\hat{\mu}, \vec{y}) - \#\text{parametri modello}\right)$$

più parametri metto e meglio è,
ma così tengo conto che ho + #parametri

BIC

REGRESSIONE LOGISTICA Semplice

un solo parametro

$$logit(p_i) = \beta_0 + \beta_1 x$$

L'equivalente dell'incremento di un'unità mi fa vedere l'incremento del logit

$$X \to X+1$$

$$\log\left(\frac{p(x+1)}{1-p(x+1)}\right) - \log\left(\frac{p(x)}{1-p(x)}\right) = \beta_1$$

ci dà il significato di β_1:

ODDS

$$\log \frac{\frac{p(x+1)}{1-p(x+1)}}{\frac{p(x)}{1-p(x)}} = \log[\text{ODDS ratio}]$$

Quindi abbiamo che

e^{β_1} = ODDS ratio all'incremento unitario della covariata.

All'aumentare della covariata l'ODDS ratio aumenta di e^{β_1}.

I test che si fanno sono di solito

$$\text{ODDS} = 1 \quad \text{vs} \quad \text{ODDS} \lessgtr 1$$

Quindi faremo test legati a β_1.

Tra le Z possiamo avere anche una variabile categorica

y_i 0 / 1

la categoria potrebbe essere fumatori / non fumatori

$$\hat{p}_i = \hat{\beta}_0 + \hat{\beta}_1 c_i$$

e^{β_i} = ODDS ratio tra categorie fumatori e non fumatori,

xk ~~passa~~ aumentando di una unità passo da una all'altra: $\dfrac{\frac{p_S}{1-p_S}}{\frac{p_{NS}}{1-p_{NS}}}$

 Regressione Logistica

$Y_i \sim Be(p_i)$

la link function $g(\cdot) = logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$

$g(p_i) = \eta_i$ predittore lineare
può variare da $-\infty$ a $+\infty$

l'inverso di g_i, che ci spiega p_i è

$$p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} = \frac{\exp(\sum \beta_j z_{ij})}{1+\exp(\sum \beta_j z_{ij})}$$

<u>CASO di $p=1$</u>

$\eta_i = \beta_0 + \beta_1 x$ ← covariata continua

Possiamo scrivere

$logit(p(x)) = \beta_0 + \beta_1 x$

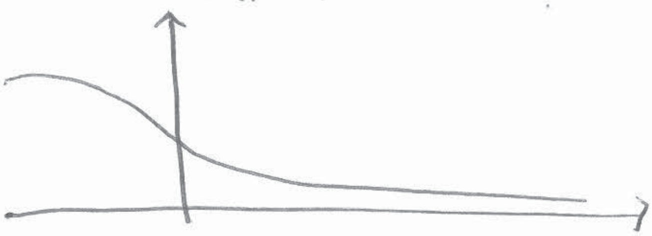
↑ funzione della covariata x

$$p(x) = \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}$$

$\beta_1 > 0$

forma delle funzioni di ripartizione.

per vedere la simmetria rispetto all'origine: $p(-x) = \frac{\exp(\beta_0-\beta_1 x)}{1+\exp(\beta_0-\beta_1 x)}$



Se incremento $x \to x+1$

$$\text{logit}(p(x+1)) - \text{logit}(p(x)) = \beta_0 + \beta_1(x+1) - \beta_0 - \beta_1 x = \beta_1$$

COSA SUCCEDE ALLA RISPOSTA, QUI LO VEDIAMO CON IL LOGIT

aumentando x di una unità

$$\beta_1 = \log\left(\frac{\frac{p(x+1)}{1-p(x+1)}}{\frac{p(x)}{1-p(x)}}\right) = \log(\text{ODDS RATIO})$$

rapporto tra due odds, ossia $\frac{\text{prob}}{1-\text{prob}}$

Se x è variabile categorica l'incremento di 1 è un passaggio di categoria.

$C_1 \dots C_K$ categorie

ottengo un coeff. relativo a ogni passaggio di categoria.

gli altri coefficienti rappresentano lo scostamento rispetto a quella categoria.

log dell'odds ratio

$$OR = \exp(\beta_1)$$

$\hat{\beta}_1$

STIME MLE ASINTOTICAMENTE GAUSSIANE

oltre la stima abbiamo uno standard error

$\hat{\beta}_1$	$Se(\hat{\beta}_1)$	Zstat	Pvalue

tabella di REGRESSIONE logistica

per $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

Possiamo costruire un IC per $\hat{\beta}_1$, xk è Gaussiano e ha standard error, quindi è

$$IC = \hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \, Se(\hat{\beta}_1)$$

$$1-\alpha = P(\beta_1 \in IC)$$

probabilità di copertura

(164) Possiamo costruire un I. Copertura per l'ODDS RATIO:

$$1-\alpha = P\left[\hat{\beta}_1 - z_{1-\frac{\alpha}{2}}\, se(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + z_{1-\frac{\alpha}{2}}\, se(\hat{\beta}_1)\right]$$

$$1-\alpha = P\left[\exp\left(\hat{\beta}_1 - z_{1-\frac{\alpha}{2}}\, se(\hat{\beta}_1)\right) \le OR \le \exp\left(\hat{\beta}_1 + z_{1-\frac{\alpha}{2}}\, se(\hat{\beta}_1)\right)\right]$$

Funzione
che ci dà la dipendenza tra p e x mi dà anche che

$$\frac{\partial p(x)}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}\right) =$$

$$= \frac{\beta_1 \exp(\beta_0+\beta_1 x)(1+\exp(\beta_0+\beta_1 x)) - \beta_1 [\exp(\beta_0+\beta_1 x)]^2}{(1+\exp(\beta_0+\beta_1 x))^2}$$

$$= \frac{\beta_1 \exp(\beta_0+\beta_1 x)}{(1+\exp(\beta_0+\beta_1 x))^2}$$

$$= \beta_1\, p(x)\,(1-p(x))$$

Quindi come cambia $p(x)$ rispetto a x?

$$\frac{\partial p(x)}{\partial x} = \beta_1\, p(x)\,(1-p(x))$$

è massima in $p(x) = \frac{1}{2}$,
dove vale

$$\frac{\beta_1}{4}$$

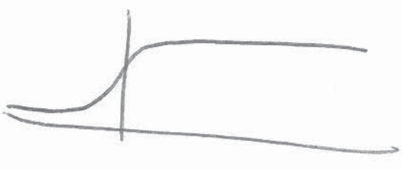
$\bar{x}$: punto di massima tangenza

$$p(\bar{x}) = \frac{1}{2} \Rightarrow \text{logit}(p(\bar{x})) = 0 = \beta_0 + \beta_1 \bar{x}$$

$$\boxed{\bar{x} = -\frac{\beta_0}{\beta_1}}$$

MEDIAN EFFECTIVE LEVEL

A seconda del β_1 ho crescita più o meno rapida della $p(x)$



È quando $\bar{y} = \frac{1}{n} \sum y_i \sim 1$ o ~ 0

Regola del pollice ci dice che dobbiamo avere abbastanza 0 o 1.

Se ho

R dice che non ho convergenza, xk mi è già chiaro che per le $x <$ tot ho 0, per $x >$ tot ho 1, non ho bisogno di regressione logistica

Regressione Logistica per Classificare

METODI di CLASSIFICAZIONE:

> CLASSIFICAZIONE SUPERVISIONATA
> CLASSIFICAZIONE NON SUPERVISIONATA

La prima è quella per cui nei miei dati so se y_i è gruppo A o gruppo B e voglio cercare modi per distinguere i due gruppi (ho maschi e femmine e voglio vedere quali sono le etichette che mi portano a classificare una persona in A o B. Lo faccio xk mi serve x fare previsione)
Impara dalle etichette che osservo e capisce quali sono le covariate che mi permettono di dare etichette.

La seconda invece non ha etichette, ma cerca di vedere se ci sono gruppi, x es. vedendo la distanza euclidea,
e quando ho nuovo dato vedrò se voglio metterlo in un gruppo o l'altro

La regressione logistica è modello di classificazione supervisionata.

Noi stimiamo

$$\hat{p} = \hat{y}_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} = \frac{\exp\left(\sum_{j=0}^{p} \hat{\beta}_j z_{ij}\right)}{1+\exp\left(\sum \hat{\beta}_j z_{ij}\right)}$$

(166) $\forall i = 1 \dots n$ produco $\hat{p}_i$ e posso dire

$$\hat{p}_i \rightsquigarrow \hat{y}_i = \begin{cases} 1 & \text{se } \hat{p}_i > p_0 \\ 0 & \hat{p}_i \leq p_0 \end{cases}$$

valore di soglia

p_0 di solito è $= \frac{1}{2}$.

Se la proporzione delle unità statistiche del dataset è molto sbilanciato (x es 75% di uni)

prendo un altro p_0

ABBIAMO QUINDI

classificazioni y_i

classificate $\hat{y}_i$

Si costruisce una

tabella di Contingenza

$\hat{y}$ \ y	0	1
0	sano e ho previsto che sn 0 n_{00}	sano 1 e ho previsto 0 n_{01}
1	n_{10}	n_{11}

← classificazione veri 0 e veri 1

qll che ho previsti

$n_{00} + n_{01} + n_{10} + n_{11} = n$ numerosità statistica dataset

Le unità classificate in modo corretto sono $\frac{n_{00} + n_{11}}{n}$

Mi piacerebbe avere una tabella diagonale!

Già qst tabella mi dice se sto classificando bene o no

1 : malato

0 : sano

I due errori fuori diagonale non sono gravi allo stesso modo,

ma pesa di più se classifico un malato come sano piuttosto che il contrario

Si introducono quindi

LA SENSIBILITÀ: probabilità di classificare un malato quando è malato

$p(\hat{y}=1 \mid y=1)$ la voglio $\simeq 1$

xk altrimenti dico che è sano ma è malato

LA SPECIFICITÀ $p(\hat{y}=0 \mid y=0)$

probabilità di classificare sano quando è sano

A ogni tabella di contingenza è associato un valore di sensibilità e di specificità.

Costruiamo la

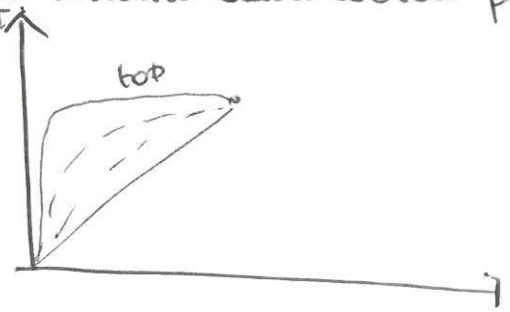
Curva ROC (curve rispost

punti del piano

Asse x : 1 - SPECIFICITÀ

Asse y : sensibilità

Ogni punto corrisponde a (x,y) di un classificatore al variare della soglia p_0



$p_0=0$
sbaglio i sani ma azzecco tutti i malati

$p_0 \in \{p_1 \dots p_n\}$ per ogni p_i ho un classificatore i, per ogni classificatore ho tabella di contingenza i e da essa calcolo sensibilità e specificità

La curva è vincolata a passare per (1,1), (0,0)

$p_0=1$
ho tutto 0 e azzecco tutti i sani

e il classificatore è migliore quanto più è sopra ~~il quadrato,~~ la diagonale, simile al quadrato

Infatti x ~~passare~~ decidere quale class è meglio si calcola l'area sotto la curva

di fatto non ha senso prendere p_0 o p_1

 Sensibilità e $\underbrace{\text{specificità}}_{\text{1-specificità bassa}}$ alta è top

1 - specificità falsi positivi (dico che sono malato ma sono sani)

$$1 - P(\hat{y}=0|y=0) = P(\hat{y}=1|y=0)$$

Sensibilità: veri positivi

Come calcolo specificità e sensibilità sulla tabella di contingenza?

$$\boxed{\text{SENSIBILITÀ} = \frac{P(\hat{Y}=1, y=1)}{\underbrace{P(Y=1)}_{\text{formula prob. condiz.}}} = \frac{n_{11}}{n_{01}+n_{11}}}$$

$$\boxed{\text{SPECIFICITÀ} = \frac{P(\hat{Y}=0, Y=0)}{P(Y=0)} = \frac{n_{00}}{n_{00}+n_{10}}}$$

Il modello sotto, stimato con la regressione logistica, non cambia

Se la tabella ha una colonna vuota viene definita degenere xk la classificazione è banale.

Di fatto in realtà la curva ROC non sarà continua ma magari è tipo a gradino, non è detto che cambino sia sens che spec.

Sensibilità e specificità sono importanti rispetto a un classificatore xk al nuovo classificato voglio sapere

$P(Y=1|\hat{Y}=1)$, il contrario, voglio sapere se, essendo classificato malato, sono malato

$$= \frac{P(Y=1, \hat{Y}=1)}{P(\hat{Y}=1)} = \frac{P(\hat{Y}=1|Y=1)\,P(Y=1)}{P(\hat{Y}=1|Y=1)P(Y=1) + P(\hat{Y}=1|Y=0)P(Y=0)}$$

$$P(M \mid test+) = \frac{\text{sensibilità} \cdot \text{Incidenza}}{\text{sensibilità} \cdot \text{Incidenza} + (1-\text{Specificità})(1-\text{Incidenza})}$$

prob essere malati quand test dice +

$\text{Incidenza} = P(Y=1)$

TABELLE di CONTINGENZA con R righe e C colonne

Quando 2 caratteristiche categoriche
(fascia reddito e marca macchina)

B \ A	A_1	A_2	
B_1			$n_{i\cdot}$
B_2			
			$n_{\cdot\cdot}$
	$n_{\cdot j}$		

Che tipo di inferenza ci può interessare?
Vedere se c'è associazione o indipendenza

Che test faccio?
Un test asintotico.

$$n_{ij} = \# \{x_i : x_i \in B_i \cap A_j\}$$

Questa tabella di contingenza viene da variabili indipendenti se la congiunta è prodotto marginali

O_{ij} = numerosità osservata

$n_{i\cdot}$ = numerosità di riga

$n_{\cdot j}$ = numerosità di colonna

$n_{\cdot\cdot}$

Le frequenze attese (numerosità attesa) in caso di indipendenza

$$E_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}}$$

xk la congiunta è prodotto marginali

170 Si può costruire una statistica test

$$\chi^2_0 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

← modo di standardizzare una Gaussiana o una Poisson

se fossero ⊥ la statistica sarebbe 0.

$$\sim \chi^2((c-1)(r-1))$$

↑ sotto H_0 (indipendenza)
[H_1: non indipendenza]

$c-1$: # colonne −1 (categorie in A −1)

$r-1$: # righe −1

Quindi

$$RC = \{ \chi^2_0 > \chi^2_{1-\alpha}((c-1)(r-1)) \}$$

torniamo a una 2x2

A \ B	1	2	
1	n_{11}	n_{12}	$n_{1\cdot}$
2	n_{21}	n_{22}	$n_{2\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$ 1

$$P^B_{A=1}\{ = \frac{n_{11}}{n_{1\cdot}} = \frac{n_{11}}{n_{11}+n_{12}} \underset{\text{se c'è indipend.}}{=} \frac{(n_{11}+n_{12})(n_{21}+n_{11})}{n_{\cdot\cdot}(n_{11}+n_{12})} = \frac{n_{11}+n_{21}}{n_{\cdot\cdot}}$$

↑ probabilità di successo rispetto a essere nella categoria 1)

$$P^B_{A=2} = \frac{n_{21}}{n_{21}+n_{22}} \underset{\perp}{=} \frac{(n_{21}+n_{22})(n_{11}+n_{21})}{(n_{21}+n_{22})n_{\cdot\cdot}} = \frac{n_{11}+n_{21}}{n_{\cdot\cdot}}$$

Indipendenza: $P^B_{A=1} = P^B_{A=2} \Rightarrow OR = 1$

Altro modo x stimare l'ODDS RATIO dalla tabella di contingenza:

$$ODDS_{A=1} = \frac{\frac{n_{11}}{n_{1\cdot}}}{\frac{n_{12}}{n_{1\cdot}}} = \frac{n_{11}}{n_{12}}$$

$$ODDS_{A=2} = \frac{\frac{n_{21}}{n_{2\cdot}}}{\frac{n_{22}}{n_{2\cdot}}} = \frac{n_{21}}{n_{22}}$$

Quindi

$$OR = \frac{n_{11}}{n_{12}} \frac{n_{22}}{n_{21}}$$

a	b
c	d

$$OR = \frac{ad}{bc}$$

stima puntuale dell'ODDS
non è detto sia = a qll di verosimiglianza con regressione logistica

Si può dimostrare che

$$SE(\log OR) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

(asintotico, per n grande)

un altro IC per log(OR)

$$\log(OR) \pm z_{1-\frac{\alpha}{2}} \, se(\log OR)$$

quindi x trovare
l'IC per OR
basta prendere l'exp.

Si può vedere quanto queste stime assomigliano a quelle della regressione.

11 giugno 2015

FACTOR ANALYSIS

Nasce in ambito psicometrico
La variabile di cui voglio cogliere il comportamento non è direttamente misurabile - fattori latenti.
Se per esempio voglio studiare la percezione del dolore, o il QI o il quoziente di apprendimento ho sì orientata di misura, ma non sono facilmente misurabili
La quantità di informazione che mi dà un questionario è limitata rispetto alla realtà, x eo se è particolarmente difficile ciao.
- Modelli di Rasch x capire come è fatto il questionario.

Ma come creare i questionari? Come disegnarli?
Vogliamo descrivere delle relazioni di covarianza tra variabili in termini di POCHI (voglio modelli semplici) FATTORI soggiacenti e NON OSSERVABILI
voglio la covarianza tra le mie variabili.
Massimizzare la sconnelatezza tra i gruppi.

Vogliamo MODELLIZZARE IL NOSTRO FENOMENO
↓
Orthogonal factor model

$\vec{X}$ aleatorio $\in \mathbb{R}^p$ — $\vec{X}$: variabili osservabili
$\vec{\mu} = \mathbb{E}[\vec{X}]$
$\sigma = Cov(\vec{X})$ $(p \times p)$

Il modello postula che $\vec{X}$ sia linearmente dipendente da POCHE variabili non osservabili
$F_1, \dots, F_m$ (fattori comuni) (come avere matrice di disegno non osservabile)
p sorgenti addizionali di variabilità $\varepsilon_1 \dots \varepsilon_p$ (ERRORI o FATTORI SPECIFICI)

$$X_1 - \mu_1 = l_{11} F_1 + l_{12} F_2 + \dots + l_{1m} F_m + \varepsilon_1$$

$$\vdots$$

$$X_p - \mu_p = l_{p1} F_1 + l_{p2} F_2 + \dots + l_{pm} F_m + \varepsilon_p$$

Un modello del genere può essere pensato come modello che genera i dati?

$$\underbrace{\vec{X} - \vec{\mu}}_{\text{vettore } (p \times 1)} = \underset{(p \times m)}{\mathbb{L}} \; \underset{(m \times 1)}{\vec{F}} + \underset{(p \times 1)}{\vec{\varepsilon}}$$

dimensionalmente torna

$l_{ij} = [L]_{ij}$: LOADING della variabile $i = 1 \dots p$ sul fattore j, $j = 1 \dots m$

Ma devo fare assunzioni sulla struttura di F, che sono di ortogonalità:

Assunzioni

- $\mathbb{E}[\vec{F}] = 0$

  $Cov[\vec{F}] = \mathbb{1}_{m \times m}$
- $\mathbb{E}[\vec{\varepsilon}] = 0$

  $Cov[\vec{\varepsilon}] = \begin{bmatrix} \varphi_1 & & \\ & \ddots & \\ & & \varphi_p \end{bmatrix}$
- $\vec{F} \perp\!\!\!\perp \vec{\varepsilon}$ fattori comuni indipendenti da quelli specifici

Se assumiamo $\vec{X} - \mu = L\vec{F} + \vec{\varepsilon}$ e qst, avremo vincoli sulla matrice di varianza covarianza di $\vec{X}$

Studiare la struttura che viene implicata nelle assunzioni su $cov(\vec{X})$

$$\underbrace{(\vec{X} - \vec{\mu})(\vec{X} - \mu)^T}_{\substack{\text{la sua media è la matrice var-cov} \\ (E(\;)=0)}} = (L\vec{F} + \vec{\varepsilon})(L\vec{F} + \vec{\varepsilon})^T =$$

$$= (L\vec{F} + \vec{\varepsilon})((L\vec{F})^T + \vec{\varepsilon}^T) =$$

$$= (L\vec{F})(L\vec{F})^T + \vec{\varepsilon}(L\vec{F})^T + L\vec{F}\vec{\varepsilon}^T + \vec{\varepsilon}\vec{\varepsilon}^T$$

$$\star\ \Sigma = \mathrm{Cov}(\vec{x}) = \mathbb{E}\left[(\vec{x}-\mu)(\vec{x}-\mu)^T\right] =$$

$$= L\,\mathbb{E}[\vec{F}\vec{F}^T]L^T + \mathbb{E}[\vec{\varepsilon}\vec{F}^T]L^T + L\,\mathbb{E}[\vec{F}\vec{\varepsilon}^T] + \mathbb{E}[\vec{\varepsilon}\vec{\varepsilon}^T]$$

$\vec{\varepsilon} \perp\!\!\!\perp \vec{F} \Rightarrow$ posso fattorizzare.

$\mathbb{E}[\vec{\varepsilon}] = 0$

$$= \underbrace{L\,\mathbb{E}[\vec{F}\vec{F}^T]}_{\text{matrice varianza cov di F}} L^T + \mathbb{E}[\varepsilon\vec{\varepsilon}^T] = LL^T + \psi$$

Allora $\mathrm{Cov}(\vec{x})$:

$$\boxed{\Sigma = LL^T + \psi}$$

ψ diagonale.

$$\star\ \mathrm{Cov}(\vec{x}, \vec{F}) = \mathbb{E}\left[(\vec{x}-\vec{\mu})\vec{F}^T\right] =$$

$$= \mathbb{E}\left[LFF^T + \vec{\varepsilon}\vec{F}^T\right] = L\,\mathrm{Cov}(\vec{F}) + \mathrm{Cov}(\vec{\varepsilon}, \vec{F}) =$$

$$= L$$

Quindi
se voglio spiegare la mia risposta $\vec{x}-\vec{\mu}$ come $L\vec{F}+\vec{\varepsilon}$
con le assunzioni di orthogonal factor model

1. $\Sigma = LL^T + \psi$ $\quad \begin{bmatrix} l_{11} & & l_{p1} \\ l_{1m} & & l_{pm} \end{bmatrix} \begin{bmatrix} l_{11} & & l_{1m} \\ l_{p1} & & l_{pm} \end{bmatrix}$ mat realmente

$$\mathrm{Var}(x_i) = \underbrace{l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2}_{\text{somma dei quadrati dei loadings}} + \psi_i \rightarrow \text{Varianza specifica}$$

h_i^2: communality o varianza comune $\quad \sum_{j=1}^{m} l_{ij}^2 = h_i^2$

ψ_i: elemento diagonale matrice i

$\mathrm{Var}(x_i) = h_i^2 + \psi_i$

$$\mathrm{Cov}(x_i, x_k) = l_{i1} \cdot l_{k1} + l_{i2} \cdot l_{k2} + \dots + l_{im} l_{km}$$

2. $\mathrm{Cov}(\vec{x}, \vec{F}) = L$

$\mathrm{Cov}(x_i, F_j) = l_{ij} \qquad i = 1 \dots p \ \text{(variabili)}$
$\qquad\qquad\qquad\qquad\quad j = 1 \dots m \ \text{(fattori)}$

Il modello assume che
p varianze (delle X_i) + $\frac{p(p-1)}{2}$ covarianze

$\Rightarrow \frac{p(p+1)}{2}$ parametri

spiegabili da
pm loading + p varianze specifiche ψ_i

$= p(m+1)$ oggetti

Se $m << p$ ho vantaggio: se mi servono 100 fattori per spiegare 100 variabili mica ci guadagno

Il problema diventa:
Quando la struttura di Σ (osservabile o meglio, stimabile)
può essere decomposta
in $LL^T + \Psi$ $m << p$?

La matrice Σ è la matrice del modello probabilistico che c'è sotto, non la osservo, posso stimarla.

$$S = \frac{\Sigma (x_i - \bar{x})(y - \bar{y})}{n-1}$$ stimatore della varianza a partire dai dati

Il problema è ben posto? No.

Esempio

$p = 3$
$m = 1 \leftarrow$ voglio trovare un fattore comune

$$\Sigma = \begin{bmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.4 \\ 0.7 & 0.4 & 1 \end{bmatrix}$$

dai nostri vincoli:
$1 = l_{11}^2 + \psi_1$
$0.9 = l_{11} \cdot l_{21}$
$0.7 = l_{11} \cdot l_{31}$
$1 = l_{22}^2 + \psi_2$
$0.4 = l_{22} \cdot l_{31}$
$1 = l_{33}^2 + \psi_3$

$$\begin{cases} 0.7 = l_{11} l_{31} \\ 0.4 = l_{21} l_{31} \end{cases} \Rightarrow l_{21} = \frac{0.4}{l_{31}} = \frac{0.4}{0.7} l_{11}$$ vincolo

lo butto in

$$0,9 = l_{11} l_{21}$$

$$\Rightarrow 0.9 = l_{11}^2 \cdot \frac{0.4}{0.7} \Rightarrow l_{11}^2 = 1.575$$

$$\Downarrow$$

$$\boxed{l_{11} = \pm 1.255}$$

ma $l_{11} = Cov(\vec{X}_1, \vec{F}_1) = Cov(X_1, F_1) / [\underbrace{Var(F_1)}_{1} \cdot \underbrace{Var(X_1)}_{1}]$

tanto ho assunto che sono 1

correlazione

ma

$$\boxed{|l_{11}| \leq 1}$$

il modulo della correlazione deve essere ≤ 1
(qui correlazione = Cov)

$$\Downarrow$$

Contraddizione!

oppure posso vederlo come

$$1 = l_{11}^2 + \varphi_1 \Rightarrow \underset{\substack{\parallel \\ Var(\varepsilon_1)}}{\varphi_1} = 1 - l_{11}^2 < 0$$

$Var(\varepsilon_1) < 0$ impossibile!

Quindi con questa matrice ciao
Non troviamo LL^T

→ dovrei aumentare m
le cose così migliorano xk ho meno vincoli
ma

$m > 1$

T matrice m×m ortogonale (rotazione) $TT^T = T^TT = 1$
succede che

$$\vec{X} - \vec{\mu} = L\vec{F} + \vec{\varepsilon} = \underbrace{LT}_{L^*}\underbrace{T^T\vec{F}}_{F^*} + \vec{\varepsilon} = L^*\vec{F}^* + \vec{\varepsilon}$$

soddisfa ancora le proprietà di ortogonalità,
xk

$\mathbb{E}[\vec{F}^*] = T^T \mathbb{E}[\vec{F}] = 0$

$Cov(\vec{F}^*) = T^* Cov(\vec{F}) T = T^T T = \mathbb{1}$

Quindi per $m > 1$ se ho soluzione non ho unicità, xk posso fare una trasformazione ortogonale
Quindi quando trovo soluzione devo ricordarmi che anche ruotando lo è.

Ok.
Supponiamo che riesco a trasporre $\Sigma = LL^T + \mathcal{E}$.
Come posso stimare?
Se ho già struttura diagonale le variabili sono già scorrelate quindi è inutile che voglio costruire un modello che mi dà cov.
Se nella stima di Σ ho elementi molto piccoli fuori diagonale ha senso cercare il modello.

Metodi di Stima

se la struttura di fattorabilità è forte dovrei ottenere le stesse risposte dai due metodi.

1) PRINCIPAL ~~FA~~ FACTOR Model

Noi abbiamo Σ e (λ_i, e_i) coppie di autovalori e autovettori

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$

Vale la decomp. spettrale

$$\Sigma = \sum_{j=1}^{p} \lambda_j \vec{e}_j \vec{e}_j^{\,T}$$

$$= \left[\sqrt{\lambda_1}\, \vec{e}_1 \,\middle|\, \sqrt{\lambda_2}\, \vec{e}_2 \,\middle|\, \dots \,\middle|\, \sqrt{\lambda_p}\, \vec{e}_p \right] \begin{bmatrix} \sqrt{\lambda_1}\, \vec{e}_1^{\,T} \\ \sqrt{\lambda_2}\, \vec{e}_2^{\,T} \\ \vdots \\ \sqrt{\lambda_p}\, \vec{e}_p^{\,T} \end{bmatrix}$$

factor decomposition
con $\varphi\Sigma = 0$ e $m = p$

infatti
$\Sigma = LL^T + 0$ con L $(m \times p)$

Sarebbe una decomposizione legittima
MA $p = m$!!

certo m. ... componiamo da quella struttura e troncare a un

posso decidere di non considerare il contributo da un certo m in poi:

$\lambda_{m+1} \vec{e}_{m+1}\vec{e}_{m+1}^T + \dots + \lambda_p \vec{e}_p\vec{e}_p^T$ li butto via:

dato che sono ordinati i λ che tolgo sono piccoli

$$\Sigma \simeq \sum_{j=1}^{m} \lambda_j \vec{e}_j\vec{e}_j^T$$

Quindi

$$L = \left[\sqrt{\lambda_1}\,\vec{e_1} \mid \dots \sqrt{\lambda_m}\,\vec{e_m}\right]$$

$$\Sigma = LL^T + \psi$$

$$\psi_i = \Sigma_{ii} - \sum_{j=1}^{m} l_{ij}^2 \quad \forall i = 1 \dots p$$

nella parte diagonale faccio la cosa giusta ma nel fuori diagonale no.

Così ho la mia decomposizione ortogonale che mi serviva:

$$\Sigma = \left(\begin{matrix}\text{troncamento a m}\\ \text{della decomp.}\\ \text{spettrale}\end{matrix}\right) + \underset{\text{residui delle varianze}}{\Psi}$$

butto via quando λ_m piccolo

(si guarda $\dfrac{\sum^{m} \lambda_i^2}{\sum^{p} \lambda_i^2}$ guardo la variabilità delle componenti scelte su quella totale)

si può dimostrare che se prendo la stima della Σ, S

$$\text{sed}\Big(\underset{\in p\times p}{S} - (\underbrace{\hat{L}_m\hat{L}_m^T}_{\in p\times p} + \hat{\varphi})\Big) \le \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2$$

stime della decomposizione troncata

$\Rightarrow$ l'errore che commetto lo posso quantificare come somma del quadrato degli autovettori che butto via.