Quantili: q_α è tc $\mathbb{P}(X \leq q_\alpha) = \alpha$, analogamente $q_{1-\alpha}$ è tc $\mathbb{P}(X > q_{1-\alpha}) = \alpha$

Stat s-m-c

Stat sufficienti

Def di stat: è una funz del campione, cioè $T(\mathbf{X})$

Def di suff: $T(\mathbf{X})$ è stat suff se la legge di $\mathbf{X}|T = t$ non dipende da ϑ per ogni t

Principio di stat suff: $T(\mathbf{X})$ è suff per ϑ quando ogni inferenza su ϑ dip da $\mathbf{X}$ solo tramite $T(\mathbf{X})$

Criterio di fatt: una stat $T(\mathbf{X})$ è suff per ϑ *sse* $\exists g, h$ tc $f(\mathbf{x}, \vartheta) = h(\mathbf{x}) \cdot g(T(\mathbf{x}), \vartheta)$

Proof. (caso discreto, senza perdita di generalità) Per ipotesi, $T(\mathbf{X})$ è stat suff. Preso $T(\mathbf{x}) = t$:

$$\begin{aligned} f(\mathbf{x}, \vartheta) &= \mathbb{P}_\vartheta(\mathbf{X} = \mathbf{x}) \\ &= \mathbb{P}_{\cancel{\vartheta}}(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t) \cdot \mathbb{P}_\vartheta(T(\mathbf{X}) = t) \\ &= h(\mathbf{x}) \cdot g(t, \vartheta) \;\Rightarrow\; \text{vale fatt} \end{aligned}$$

Ora per ipotesi vale la fatt. Prese $q(t, \vartheta)$ la densità di T e $A_T = \{\mathbf{y} \text{ tc } T(\mathbf{y}) = T(\mathbf{x})\}$, la legge di $\mathbf{Y}|T(\mathbf{x})$ è

$$\frac{f(\mathbf{x}, \vartheta)}{q(T(\mathbf{x}), \vartheta)} = \frac{g(T(\mathbf{x}), \vartheta) \cdot h(\mathbf{x})}{\sum_{\mathbf{y} \in A} g(T(\mathbf{y}), \vartheta) \cdot h(\mathbf{y})} =$$

$$= \frac{\cancel{g(T(\mathbf{x}), \vartheta)} \cdot h(\mathbf{x})}{\cancel{g(T(\mathbf{y}), \vartheta)} \sum_{\mathbf{y} \in A} h(\mathbf{y})} = \frac{h(\mathbf{x})}{\sum_{\mathbf{y} \in A} h(\mathbf{y})}$$

che non dip da $\vartheta \;\Rightarrow\;$ stat suff ■

Cor: date $T(\mathbf{X})$ stat suff e $r(\cdot)$ funz biunivoca, $T^*(\mathbf{X}) = r(T(\mathbf{X}))$ è stat suff

Proof. Dato che $T(\mathbf{X})$ è stat suff, per il criterio di fatt posso scrivere:

$$\begin{aligned} f(\mathbf{x}, \vartheta) &= h(\mathbf{x}) \cdot g(T(\mathbf{x}), \vartheta) \\ &= h(\mathbf{x}) \cdot g(r^{-1}(T(\mathbf{x})), \vartheta) \\ &= h(\mathbf{x}) \cdot g^*(T^*(\mathbf{x}), \vartheta) \;\Rightarrow\; T^* \text{ stat suff} \end{aligned}$$ ■

Stat (sufficienti e) minimali

Def di stat suff e min: $T(\mathbf{X})$ è s-m se è funzione di ogni altra stat suff

Teo di LS1: $T(\mathbf{X})$ è stat s-m per ϑ se

$$\frac{f(\mathbf{x}, \vartheta)}{f(\mathbf{y}, \vartheta)} \text{ non dip da } \vartheta \;\Leftrightarrow\; T(\mathbf{x}) = T(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y}$$

Stat complete

Def: detta $h(t, \vartheta)$ la famiglia di densità per la stat $T(\mathbf{X})$, diciamo che h è completa (cioè T stat completa) se $\forall g$ misurabile e $\forall \vartheta$ vale che $\mathbb{E}_\vartheta[g(T)] = 0 \Rightarrow g(T) = 0$ qo

Teo di Bahadur: suff compl $\Rightarrow$ min

Exponential Family

$f(x, \underline{\vartheta}) = h(x) \cdot c(\underline{\vartheta}) \cdot \exp\left\{\sum_{i=1}^k w_i(\underline{\vartheta})\, t_i(x)\right\}$

Trovare stat s-m-c nella EF: $T(\mathbf{x}) = \left(\sum_{i=1}^n t_1(x_i), \ldots, \sum_{i=1}^n t_k(x_i)\right)$ è suff, è anche compl se $(w_1(\underline{\vartheta}), \ldots, w_k(\underline{\vartheta}))$ mappa Θ in un insieme che contiene almeno un aperto di $\mathbb{R}^k$, è anche min per Bahadur

Stat ancillari

Def: stat S la cui f non dip da ϑ (S $\perp\!\!\!\perp$ s-m-c)

Stimatori

Def di stimatore puntuale per ϑ: è una qualunque funz solo di $\mathbf{X}$

Metodo dei momenti

Uguaglio momenti teorici $\mathbb{E}\left[X_i^k\right]$ a momenti empirici $\frac{1}{n}\sum_{i=1}^n X_i^k$ (stime fornite non sempre ammissibili, cioè $\notin \Theta$)

Stimatori di max verosimiglianza

Def: $\widehat{\vartheta}_{\mathrm{MLE}}(\mathbf{x}) = \mathrm{ArgSup}_{\vartheta \in \Theta} L(\mathbf{x}, \vartheta)$ (stime sempre ammissibili)

Principio di invarianza per MLE: lo stimatore MLE di $\tau(\vartheta)$ è $\tau\left(\widehat{\vartheta}_{\mathrm{MLE}}\right)$

MSE

$\mathrm{MSE}(T) = \mathbb{E}[(T - \vartheta)^2] = \mathrm{Var}(T) + (\mathbb{E}[T] - \vartheta)^2$

Stimatore corretto/non distorto: se $\mathbb{E}_\vartheta[T] = 0$

UMVUE e Cramer-Rao

Def: $\widetilde{T}$ è UMVUE per ϑ se $\widetilde{T}$ è lo stimatore non distorto di ϑ a varianza/MSE minima/o

Dis. di Cramer-Rao: siano $X_i \overset{iid}{\sim} f(x, \vartheta)$ e sia $T(\mathbf{X})$ stimatore di ϑ; se
(1) $\mathrm{supp}(X_i)$ non dip da ϑ
(2) $\mathbb{E}_\vartheta\left[T^2\right] < \infty$
(3) $\frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathbb{E}_\vartheta[T] = \int_{\mathbb{R}^n} T(\mathbf{x})\, \frac{\partial}{\partial\vartheta} f(\mathbf{x}, \vartheta)\, \mathrm{d}\mathbf{x}$
allora $\mathrm{Var}_\vartheta(T) \geq l_{CR} = (\frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathbb{E}_\vartheta[T])^2 / I_n(\vartheta)$, dove $I_n(\vartheta) = \mathbb{E}_\vartheta\left[\left(\frac{\partial}{\partial\vartheta}\, \log\, f(\mathbf{x}, \vartheta)\right)^2\right]$ è l'info di Fisher

Proof. Siano $X = T$ e $Y = \frac{\partial}{\partial\vartheta}\, \log\, f(\mathbf{x}, \vartheta)$.
Passo I $\to$ scrivo l'ip (3) per $T(\mathbf{x}) = 1$:
$0 = \frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathbb{E}[1] = \int_{\mathbb{R}^n} \frac{\partial}{\partial\vartheta}\, [f(\mathbf{x}, \vartheta)]\, \mathrm{d}\mathbf{x}$
$= \left\{\circledast : \frac{\partial}{\partial\vartheta}\, (\log\, f) = \frac{1}{f} \cdot f'\right\}$
$= \int_{\mathbb{R}^n} \frac{\partial}{\partial\vartheta}\, [\log f(\mathbf{x}, \vartheta)]\, f(\mathbf{x}, \vartheta)\, \mathrm{d}\mathbf{x} = \mathbb{E}[Y]$
Passo II $\to$ calcolo il numeratore:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathbb{E}[T] &\overset{(3)}{=} \int_{\mathbb{R}^n} T(\mathbf{x})\, \frac{\partial}{\partial\vartheta} f(\mathbf{x}, \vartheta)\, \mathrm{d}\mathbf{x} \\ &\overset{\circledast}{=} \int_{\mathbb{R}^n} T(\mathbf{x}) \frac{\partial}{\partial\vartheta}\, [\log f(\mathbf{x}, \vartheta)]\, f(\mathbf{x}, \vartheta)\, \mathrm{d}\mathbf{x} \\ &= \mathbb{E}[T \cdot Y] \overset{\mathrm{I}}{=} \mathrm{Cov}(X, Y) \end{aligned}$$

Passo III $\to$ calcolo il denominatore:
$I_n(\vartheta) = \left[\left(\frac{\partial}{\partial\vartheta}\, \log\, f(\mathbf{x}, \vartheta)\right)^2\right] = \mathbb{E}\left[Y^2\right] \overset{\mathrm{I}}{=} \mathrm{Var}(Y)$
Sostituendo II e III nella disuguaglianza di CS: $\mathrm{Var}(X) \geq \frac{|\mathrm{Cov}(X,Y)|^2}{\mathrm{Var}(Y)} \rightsquigarrow$ tesi ■

Parentesi sull'info di Fisher

Lemma 1: sotto le hp della dis. di CR (che valgono sempre se $\in$EF) vale che $I_n(\vartheta) = n I_1(\vartheta)$ (densità unimodale)

Proof.
$I_n(\vartheta) = \mathbb{E}_\vartheta\left[\left(\frac{\partial}{\partial\vartheta}\, \log\, f(\mathbf{x}, \vartheta)\right)^2\right]$
$= \mathbb{E}_\vartheta\left[\left(\frac{\partial}{\partial\vartheta}\, \log\, \prod_i f(x_i, \vartheta)\right)^2\right]$
$= \mathbb{E}_\vartheta\left[\left(\sum_i \frac{\partial}{\partial\vartheta}\, \log\, f(x_i, \vartheta)\right)^2\right]$
$\overset{\mathrm{I}}{=} \mathrm{Var}\left(\sum_i \frac{\partial}{\partial\vartheta}\, \log\, f(x_i, \vartheta)\right)$
$= \sum_i \mathrm{Var}(\ldots) = n\, \mathrm{Var}(\ldots) = n I_1(\vartheta)$ ■

Lemma 2: se in aggiunta alle hp della dis. di CR $\frac{\mathrm{d}}{\mathrm{d}\vartheta}\left[\int_{\mathbb{R}^n} \frac{\partial}{\partial\vartheta} f(\mathbf{x}, \vartheta)\, \mathrm{d}\mathbf{x}\right] = \int_{\mathbb{R}^n} \frac{\partial^2}{\partial\vartheta^2} f(\mathbf{x})\, \mathrm{d}\mathbf{x}$ allora $I_n(\vartheta) = -\mathbb{E}_\vartheta\left[\frac{\partial^2}{\partial\vartheta^2} \log f(\mathbf{x}, \vartheta)\right]$

Proof. $\mathbb{E}\left[\partial_{\vartheta\vartheta} \log f\right] := \int \partial_{\vartheta\vartheta} \log f \cdot f =$
$\overset{\circledast}{=} \int \partial_\vartheta \left(\frac{\partial_\vartheta f}{f}\right) \cdot f = \int \partial_{\vartheta\vartheta} f - \int \frac{(\partial_\vartheta f)^2}{f^2} \cdot f$
Ora analizzo i due pezzi separatamente:
$\int \partial_{\vartheta\vartheta} f\, \mathrm{d}\mathbf{x} \overset{Hp}{=} \frac{\mathrm{d}}{\mathrm{d}\vartheta}\left[\int \partial_\vartheta f\, \mathrm{d}\mathbf{x}\right] =$
$= \frac{\mathrm{d}}{\mathrm{d}\vartheta}\left[\int \frac{\partial_\vartheta f}{f} \cdot f\, \mathrm{d}\mathbf{x}\right] = \mathbb{E}[Y]' \overset{\mathrm{I}}{=} 0$
$\int \frac{(\partial_\vartheta f)^2}{f^2} \cdot f\, \mathrm{d}\mathbf{x} = \int (\partial_\vartheta \log f)^2 f\, \mathrm{d}\mathbf{x} = I_n$ ■

Continuo su UMVUE

T unbiased per $\vartheta \Rightarrow$ il numeratore di CR è 1

Se T è non distorto ed efficiente (cioè la sua var raggiunge l_{CR}) allora T è UMVUE (però non è che l'UMVUE raggiunge sempre tale limite)

C'è sempre al masismo una funzione di ϑ che è stimata in maniera efficiente

Teo di Rao-Blackwell: presi T stimatore non distorto per $\tau(\vartheta)$ e W stat suff per ϑ, allora la va $\mathbb{E}[T|W]$ è stimatore non distorto per $\tau(\vartheta)$ e $\mathrm{Var}(\mathbb{E}[T|W]) \leq \mathrm{Var}(T)\ \forall \vartheta$

Proof. (solo per $\tau(\vartheta) = \vartheta$)
$\mathbb{E}[T|W] = \int_{\mathbb{R}^n} T(\mathbf{x}) f(\mathbf{x}|w)\, \mathrm{d}\mathbf{x}$ non dip da ϑ perché W è stat suff $\Rightarrow$ è stimatore
$\mathbb{E}[\mathbb{E}[T|W]] = \mathbb{E}[T] = \vartheta \Rightarrow$ è unbiased
$\mathrm{Var}(T) = \mathrm{Var}(\mathbb{E}[T|W]) + \underbrace{\mathbb{E}[\mathrm{Var}(T|W)]}_{\geq 0}$ ■

Teo di unicità: $T(\mathbf{X})$ UMVUE per ϑ è unico

Proof. Per assurdo $\exists S \neq T$ UMVUE per ϑ. Allora $W = (T + S)/2$ è unbiased e per CR: $\mathrm{Var}(W) = \frac{1}{4}(\mathrm{Var}(T) + \mathrm{Var}(S) + 2\,\mathrm{Cov}(T, S)) \leq \frac{1}{4}(\mathrm{Var}(T) + \mathrm{Var}(S) + 2\sqrt{\mathrm{Var}(T)\,\mathrm{Var}(S)}) = \mathrm{Var}(T)$ xk $\mathrm{Var}(S) = \mathrm{Var}(T)$ dato che sono UMVUE. Ma dato che W è uno stim unbiased con var $\leq$ alla var dell'UMVUE, allora deve valere l'=, quindi S è una trasf. lin. di T: $S = aT + b$. Tuttavia, $\mathrm{Var}(T) = \mathrm{Cov}(T, S) = \mathrm{Cov}(T, aT + b) = a\,\mathrm{Var}(T) \;\Rightarrow\; a = 1$ e $\mathbb{E}[T] = \mathbb{E}[S] = \mathbb{E}[aT + b] \;\Rightarrow\; b = 0$, cioè $S = T$ assurdo ■

Teo di LS2: T stim unbiased e W stat s-c(-m) per ϑ (o $\tau(\vartheta)$) $\Rightarrow M = \mathbb{E}[T|W]$ è UMVUE per ϑ

Proof. Per RB M è stim unbiased, per assurdo non è UMVUE $\Rightarrow \exists T'$ stim unbiased tc $\mathrm{Var}(T') < \mathrm{Var}(M)$. Allora per RB $M' = \mathbb{E}[T'|W]$ è stim unbiased con $\mathrm{Var}(M') \leq \mathrm{Var}(T') < \mathrm{Var}(M)$. Osservo che la funz $g(W) = M - M'$ ha media nulla: $\mathbb{E}[M - M'] = \vartheta - \vartheta = 0$, perciò per la completezza di W deve essere $M - M' = 0$ qc, che contraddice M non UMVUE ■

Test d'ipotesi

Def d'ipotesi statistica: è un'affermazione sui par incogniti della legge del campione

Def di test d'ipotesi: è una regola che specifica per quali $\mathbf{x}$ accetto/rifiuto H_0, cioè specifica la regione critica $\mathcal{R} = \{\mathbf{x} \in \mathbb{R}^n : \text{rifiuto } H_0\}$

Test LRT: $\mathcal{R} = \{\lambda(\mathbf{x}) \leq c\}$, $c \in (0, 1)$

	accetto H_0	rifiuto H_0
H_0 vera	OK	$\mathcal{E}1$ condanno innoc
H_1 vera	$\mathcal{E}2$ assolvo colp	OK

OSS: $\mathcal{E}1$ più grave

Def di potenza di un test d'ip con regione $\mathcal{R}$: $\beta(\vartheta) : \Theta \to [0, 1]$ tc

$$\beta(\vartheta) = \mathbb{P}_\vartheta(\mathbf{x} \in \mathcal{R}) = \begin{cases} \mathbb{P}_{\vartheta \in \Theta_0}(\mathcal{E}1) \\ 1 - \mathbb{P}_{\vartheta \in \Theta_0^c}(\mathcal{E}2) \end{cases}$$

Scopo (graficam.): $\beta \to 0$ in Θ_0, $\beta \to 1$ in Θ_0^c
Dimensione (livello): test è di dim (liv) α $(0 \leq \alpha \leq 1)$ se $\sup_{\vartheta \in \Theta_0} \beta(\vartheta) = (\leq)\, \alpha$
Test non dist: $\beta(\vartheta') \geq \beta(\vartheta^*) \forall \vartheta' \in \Theta_0^c, \vartheta^* \in \Theta_0$

Def di UMP: sia $\mathcal{C}$ una classe di test per verificare $H_0 : \vartheta \in \Theta_0$ vs $H_1 : \vartheta \in \Theta_0^c$. Allora un test in $\mathcal{C}$ è UMP se $\beta(\vartheta) \geq \beta'(\vartheta)\ \forall \vartheta \in \Theta_0^c$ e per ogni β' funz potenza di un test in $\mathcal{C}$

Teo di NP: considero la classe di test con ipotesi $H_0 : \vartheta = \vartheta_0$ vs $H_1 : \vartheta = \vartheta_1$ in cui la densità di prob associata è $f(\mathbf{x}, \vartheta_i)$, $u \in \{0, 1\}$. Se la $\mathcal{R}$ è tc
1)$\mathbf{x} \in \mathcal{R}$ se $f(\mathbf{x}, \vartheta_1) > k f(\mathbf{x}, \vartheta_0)$ $(k \geq 0)$
$\mathbf{x} \in \mathcal{R}^c$ se $f(\mathbf{x}, \vartheta_1) < k f(\mathbf{x}, \vartheta_0)$
2)$\alpha = \mathbb{P}_{\vartheta_0}(\mathbf{x} \in \mathcal{R})$, allora:
(suff) ogni test che sodd. 1-2 è test UMP di liv α
(nec) se $\exists$ test che sodd. 1-2 con $k > 0$, allora ogni test UMP di liv α ha anche dim α, sodd. 2 e sodd. 1 tranne che su un insieme a mis 0.

MLR: una fam di densità $\{g(t, \vartheta) : \vartheta \in \Theta \subset \mathbb{R}\}$ ha MLR (cr. o decr.) se $\forall \vartheta_2 > \vartheta_1$ la funz $g(t, \vartheta_2)/g(t, \vartheta_1)$ è monotona (cr. o decr.) in t

Teo di KR: considero la classe di test con ipotesi $H_0 : \vartheta \leq \vartheta_0$ vs $H_1 : \vartheta > \vartheta_0$ e sia T stat suff per ϑ la cui legge ha MLR cr. (decr.). Allora $\forall t_0$ il test con $\mathcal{R} = \{T > (<)\, t_0\}$ è UMP di livello $\alpha = \mathbb{P}_{H_0}(T > (<)\, t_0)$

P-value

Def: $p(\mathbf{x})$ è una qlnq stat tc $0 \leq p \leq 1$
P-value valido: se $\mathbb{P}_\vartheta(p(\mathbf{X}) \leq \alpha) \leq \alpha\ \forall \vartheta \in \Theta_0$
Teo: p valido $\Rightarrow$ test con $\mathcal{R} = \{p \leq \alpha\}$ ha liv α
CS per p-value validi: sia $W(\mathbf{X})$ una stat tc valori grandi di W danno evidenza contro H_0. Allora $p(\mathbf{x}) = \sup_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta(W(\mathbf{X}) \geq W(\mathbf{x}))$ è p-value valido

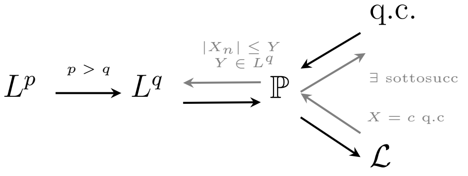
Intervalli di confidenza

Def: stimatore intervallare di ϑ è coppia di stat $(L(\mathbf{X}), U(\mathbf{X}))$ tc $L(\mathbf{X}) \leq U(\mathbf{X})$, e l'int. aleatorio $[L, U]$ è la stima intervallare
Def: la prb di copertura di una stima int. $[L, U]$ per ϑ è $\mathbb{P}_\vartheta(\vartheta \in [L, U])$
Def: il liv. di conf. di una stima int. $[L, U]$ per ϑ è $\inf_\vartheta \mathbb{P}_\vartheta(\vartheta \in [L, U])$

Teoria asintotica

Convergenze

▷ $X_n \xrightarrow{\text{qc}} X$ se $\mathbb{P}(X_n \xrightarrow{n} X) = 1$
▷ $X_n \xrightarrow{\mathbb{P}} X$ se $\lim_n \mathbb{P}(|X_n - X| < \mathcal{E}) \to 1\ \forall \mathcal{E} > 0$
▷ $X_n \xrightarrow{\mathcal{L}} X$ se $F_{X_n}(t) \xrightarrow{n} F_X(t)\ \forall t$ di cont
▷ $X_n \xrightarrow{L^p} X$ se $\mathbb{E}\left[|X_n - X|^p\right] \xrightarrow{n} 0$



Altre proprietà e teorema di Slutsky

Dati $X_n \xrightarrow{\mathcal{L}} X, Y_n \xrightarrow{\mathcal{L}} c \in \mathbb{R}, a \in \mathbb{R}, h \in \mathcal{C}^0(\mathbb{R})$:
▷ $X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$
▷ $X_n \xrightarrow{\mathcal{L}} c \Rightarrow X_n \xrightarrow{\mathbb{P}} c\ (c \in \mathbb{R})$
▷ $aX_n \xrightarrow{\mathbb{P}} aX$
▷ $h(X_n) \xrightarrow{\mathbb{P}} h(X)$
▷ Slutsky $\begin{cases} X_n + Y_n \xrightarrow{\mathcal{L}} X + c \\ X_n Y_n \xrightarrow{\mathcal{L}} +cx \\ \frac{X_n}{Y_n (\neq 0)} \xrightarrow{\mathcal{L}} \frac{X}{c(\neq 0)} \end{cases}$

Proprietà generali media campionaria

Date $X_1, \ldots, X_n$ iid si ha:
$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}[X]$
$\text{Var}\left(\overline{X}_n\right) = \text{Var}(X)/n$
$\mathbb{E}\left[\overline{X}_n^2\right] = \text{Var}(X)/n - \mathbb{E}[X]^2$

LGN

Date $X_i \in L^1$ iid si ha:

$$\mathbb{E}\left[\overline{X}_n\right] = \mu \in \mathbb{R} \iff \overline{X}_n \xrightarrow{\text{qc}, \mathbb{P}, L^1} \mu$$

(è utile per verificare la proprietà di consistenza di uno stimatore tipo $\widehat{\vartheta}_{\text{MLE}} = \overline{X}_n$, senza passare per il toerema apposito)

TCL

Date $X_i \in L^2$ iid si ha:

$$\frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{\text{Var}(X)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{ovvero}$$

$$\sqrt{n}(\overline{X}_n - \mathbb{E}[X]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}(X)) \quad \text{ovvero}$$

$$\overline{X}_n \approx \mathcal{N}(\mathbb{E}[X], \text{Var}(X)/n)$$

(tale ris. è esatto se campione gaussiano)

Applicazioni:
▷ $\frac{\text{Bi}(n,p) - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$
$\Rightarrow \text{Bi}(n,p) \overset{n \text{ big}}{\simeq} \mathcal{N}(np, np(1-p))$
▷ $\frac{\chi^2(n) - n}{\sqrt{2n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$
$\Rightarrow \chi^2(n) \overset{n \text{ big}}{\simeq} \mathcal{N}(n, 2n)$
▷ $t(n) = \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2(n)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

Proprietà as. stimatori

Def as. correttezza: una succ di stimatori W_n per ϑ è as. non distorta se $\mathbb{E}[W_n] \xrightarrow{n} \vartheta$
Def consistenza: una succ di stimatori W_n per ϑ è consistente se $W_n \xrightarrow{\mathbb{P}} \vartheta$, cioè $\mathbb{P}(|W_n - \vartheta| < \mathcal{E}) \xrightarrow{n} 1$
Come verifico tale proprietà?
Se $W_n \equiv \overline{X}_n$ allora è consistente (LGN sopra)
Altrimenti verifico che $\text{MSE}(W_n) \xrightarrow{n} 0$ (xk la convergenza quadratica implica quella in probabilità)
Def as. normalità: una succ di stimatori W_n per $\tau(\vartheta)$ è as. normale se
$\sqrt{n}(W_n - \tau(\vartheta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$, cioè $W_n \approx \mathcal{N}(\tau(\vartheta), \sigma^2/n)$. La qtà σ^2 è detta varianza asintotica ($\neq \lim \text{Var}$)
NB: anche questa prop. è gratis se $W_n \equiv \overline{X}_n$ (TCL)
Teo: as. normalità $\Rightarrow$ consistenza

Proof. Per Slutsky:
$W_n - \tau(\vartheta) = \frac{\sigma}{\sqrt{n}} \left(\frac{W_n - \tau(\vartheta)}{\sigma/\sqrt{n}} \right) \xrightarrow{\mathcal{L}} 0 \cdot \mathcal{N}(0,1) = 0$
e dato che la conv. in $\mathcal{L}$ ad una cost implica quella in prob si ottiene la tesi: $W_n \xrightarrow{\mathbb{P}} \tau(\vartheta)$ ■

Def as. efficienza: una succ di stimatori W_n per $\tau(\vartheta)$ è as. efficiente se
$\sqrt{n}(W_n - \tau(\vartheta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v(\vartheta))$ con $v = (\tau')^2 / I_1$, cioè $W_n \approx \mathcal{N}(\tau(\vartheta), l_{\text{CR}})$

Confronto: più la var as. è piccola, più lo stimatore è eff.

Risultati per MLE

Sotto le ipotesi di regolarità della dis. di CR (nella pratica $\vartheta \notin \text{supp} X_i$) gli MLE sono:
1) (almeno) as. non distorti
2) consistenti
3) as. normali ed efficienti, cioè
$\sqrt{n}(\widehat{\vartheta} - \vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/I_1)$

Ciò vale anche per $\tau(\vartheta)$, e la 3 diventa:
$\sqrt{n}(\tau(\widehat{\vartheta}) - \tau(\vartheta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v)$

Metodi delta

Delta 1: siano W_n tc $\sqrt{n}(W_n - \vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ e g funz. tc $\exists g'(\vartheta) \neq 0$, allora:
$\sqrt{n}(g(W_n) - g(\vartheta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \cdot (g'(\vartheta))^2)$

Applicato a $\overline{X}_n$ $(\mu = \mathbb{E}[X]), \sigma^2 = \text{Var}(X)$:
$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2\ (g'(\mu))^2)$

Delta 2: siano W_n tc $\sqrt{n}(W_n - \vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ e g funz. tc $\exists g'(\vartheta), \exists g''(\vartheta) \neq 0$, allora:
$n(g(W_n) - g(\vartheta)) \xrightarrow{\mathcal{L}} \frac{\sigma^2}{2} g''(\vartheta)\, \chi^2_1$

Intervalli di confidenza asintotici

Capiamo tutto da un esempio: $X_i \overset{\text{iid}}{\sim} f(x, \vartheta)$ (ϑ non nel supporto), trovo $\widehat{\vartheta}_{\text{L}}$. Allora so che $\widehat{\vartheta}_{\text{L}}$ è as. non distorto, consistente, as. normale. Per capire con che legge calcolo I_1 e risulta pari a ϑ^{-2}. Allora
$\sqrt{n}\left(\widehat{\vartheta}_{\text{L}} - \vartheta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \vartheta^2)$
cioè $\widehat{\vartheta}_{\text{L}} \approx \mathcal{N}(\vartheta, \vartheta^2/n)$.

Ora mi chiedo: che legge asintotica ha $\tau(\vartheta) = e^{-\vartheta}$? Con il metodo delta 1 posso subito dire che
$\widehat{\tau}_{\text{L}} \approx \mathcal{N}\left(e^{-\vartheta}, \frac{\vartheta^2}{n} \cdot e^{-2\vartheta}\right)$

Allora, se il campione è numeroso:
$\tau(\vartheta) = e^{-\vartheta} \in \left[e^{-\widehat{\vartheta}_{\text{L}}} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{\vartheta}_{\text{L}}^2}{n}\, e^{-2\widehat{\vartheta}_{\text{L}}}} \right]$

Teo: considero test $H_0: \vartheta = \vartheta_0$ vs $H_1: \vartheta \neq \vartheta_0$ e $X_1, \ldots, X_n \sim f(x, \vartheta) \in \text{EF}$. Allora sotto H_0 vale che $W = -2\log(\lambda(\mathbf{x})) \xrightarrow{\mathcal{L}} \chi^2_1$

Proof. Per le prop dei logaritmi $W = 2l(\mathbf{x}, \widehat{\vartheta}_L) - 2l(\mathbf{x}, \vartheta_0)$. Inoltre, per Taylor al second'ordine $l(\mathbf{x}, \vartheta_0) = l(\mathbf{x}, \widehat{\vartheta}_L) + \cancel{l'(\mathbf{x}, \widehat{\vartheta}_L)}(\widehat{\vartheta}_L - \vartheta_0) + \frac{1}{2} l''(\mathbf{x}, \widehat{\vartheta}_L)(\widehat{\vartheta}_L - \vartheta_0)^2$. Dunque $W = -l''(\mathbf{x}, \widehat{\vartheta}_L)(\widehat{\vartheta}_L - \vartheta_0)^2$. Sfruttando che $I_n(\widehat{\vartheta}_L) = -l''(\mathbf{x}, \widehat{\vartheta}_L)$ e $\frac{I_n(\widehat{\vartheta}_L)}{n} \xrightarrow{\text{qc}, \mathbb{P}} I(\vartheta)$, sotto H_0 vale

$$W = \underbrace{-l''(\mathbf{x}, \widehat{\vartheta}_L) \cdot \frac{1}{nI(\vartheta_0)}}_{\xrightarrow{\mathcal{L}} 1} \underbrace{\left(\frac{\sqrt{n}(\widehat{\vartheta}_L - \vartheta_0)}{1/\sqrt{I(\vartheta_0)}} \right)^2}_{\xrightarrow{\mathcal{L}} \chi^2_1} \quad ■$$

Risultati notevoli

Varianza campionaria

$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \overline{X}_n)^2$
$= \frac{n}{n-1} \left[\frac{\sum X_i^2}{n} - \left(\frac{\sum X_i}{n} \right)^2 \right]$
$\mathbb{E}\left[S^2\right] = \sigma^2 \qquad S^2 \perp\!\!\!\perp \overline{X}_n$
$S^2 \xrightarrow{\text{qc}} \sigma^2$ analogo alla LGN per $\overline{X}_n$

Quantili di una gamma

$Y \sim \Gamma(n, \lambda) \Rightarrow 2\lambda Y \sim \Gamma(n, \frac{n}{2}) = \chi^2(2n)$
$\Rightarrow \gamma_{1-\alpha}(n, \lambda) = x^2_{1-\alpha}(2n)/2\lambda$

Regressione

Modello e GOF

Forma compatta: $\overrightarrow{y} = \mathbb{Z}\overrightarrow{\beta} + \overrightarrow{\varepsilon}$
R^2, R^2_{adj}, metodi stepwise (con test F) per selezionare covariate, $VIF_j = \frac{1}{1-R_j^2}$ (per verificare la colinearità tra le covariate), plot dei leverages, plot dei residui, influential plot ...

Stimatori OLS

Def: $\widehat{\overrightarrow{\beta}}_{\text{LS}} = \underset{\overrightarrow{b} \in \mathbb{R}^{r+1}}{\text{ArgMin}} \left(\overrightarrow{y} - \mathbb{Z}\overrightarrow{b} \right)^T \left(\overrightarrow{y} - \mathbb{Z}\overrightarrow{b} \right)$

Teo 1: $\widehat{\overrightarrow{\beta}}_{\text{LS}} = (\mathbb{Z}^T\mathbb{Z})^{-1}\mathbb{Z}^T\overrightarrow{y}$

Teo 2: $\mathbb{E}\left[\widehat{\overrightarrow{\beta}}_{\text{LS}}\right] = \overrightarrow{\beta}$ (stim. non distorti)

Proof. $\mathbb{E}\left[\widehat{\overrightarrow{\beta}}_{\text{LS}}\right] = \mathbb{E}\left[(\mathbb{Z}^T\mathbb{Z})^{-1}\mathbb{Z}^T\overrightarrow{y}\right]$
$= (\mathbb{Z}^T\mathbb{Z})^{-1}\mathbb{Z}^T\mathbb{E}\left[\overrightarrow{y}\right]$
$= (\mathbb{Z}^T\mathbb{Z})^{-1}\mathbb{Z}^T\mathbb{E}\left[\mathbb{Z}\overrightarrow{\beta} + \overrightarrow{\varepsilon}\right]$
$= (\mathbb{Z}^T\mathbb{Z})^{-1}\mathbb{Z}^T\mathbb{Z}\overrightarrow{\beta} = \overrightarrow{\beta}$ ■

Teo 3: $\text{Cov}\left(\widehat{\overrightarrow{\beta}}_{\text{LS}}\right) = \sigma^2 \left(\mathbb{Z}^T\mathbb{Z}\right)^{-1}$

Decomposizione della var

▷ $\text{SS}_{\text{tot}} = \sum_{i=1}^n (y_i - \bar{y})^2$ variabilità tot
▷ $\text{SS}_{\text{reg}} = \sum_{i=1}^n (\widehat{y}_i - \bar{\widehat{y}})^2$ var dei fittati
▷ $\text{SS}_{\text{res}} = \sum_{i=1}^n (\widehat{\varepsilon}_i)^2$ var dei residui
$\implies \text{SS}_{\text{tot}} = \text{SS}_{\text{reg}} + \text{SS}_{\text{res}}$

Coeff di determinazione

Qntfca la % di var dei dati spiegata dal modello:
$R^2 = \frac{\text{SS}_{\text{reg}}}{\text{SS}_{\text{tot}}} = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}} \in [0,1]$
Può essere aggiustato rispetto al no. di covariate:
$R^2_{\text{adj}} = 1 - \frac{\text{SS}_{\text{reg}}/(n-r-1)}{\text{SS}_{\text{tot}}/(n-1)} = 1 - (1-R^2)\frac{n-1}{n-r-1}$

Stimatore non distorto della varianza

Assumendo errore gaussiano (omoschedasticità), lo stimatore non distorto della varianza dell'errore è $S^2 = \frac{n\widehat{\sigma}^2}{n-r-1}$ perché
$n\widehat{\sigma}^2 = \widehat{\overrightarrow{\varepsilon}}^T\widehat{\overrightarrow{\varepsilon}} \sim \sigma^2\chi^2(n-r-1) \Rightarrow \mathbb{E}[S^2] = \sigma^2$

Proof. $\widehat{\overrightarrow{\varepsilon}} \sim \mathcal{N}(0, \sigma^2(I-H))$ e quindi $\widehat{\overrightarrow{\varepsilon}}^T \frac{(I-H)^-}{\sigma^2} \widehat{\overrightarrow{\varepsilon}} \sim \chi^2(n-r-1)$. Inoltre $H\overrightarrow{\varepsilon} = 0$ e quindi $\widehat{\overrightarrow{\varepsilon}} = (I-H)\widehat{\overrightarrow{\varepsilon}}$. Allora:
$\chi^2_{n-r-1} \sim \widehat{\overrightarrow{\varepsilon}}^T \frac{(I-H)^-}{\sigma^2}(I-H)\widehat{\overrightarrow{\varepsilon}} = \frac{\widehat{\overrightarrow{\varepsilon}}^T\widehat{\overrightarrow{\varepsilon}}}{\sigma^2}$ ■

IC per i beta_j di un modello con err gauss

$\text{IC}_{1-\alpha} = \left[\widehat{\beta}_j \pm \sqrt{S^2 \text{diag}_j(\mathbb{Z}^T\mathbb{Z}^{-1})}\, t^{n-r-1}_{1-\frac{\alpha}{2}} \right]$

Test per i parametri beta

Costruisco il test $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$. Sotto H_0 vale:
$T_j = \frac{\widehat{\beta}_j}{\sqrt{S^2 \text{diag}_j(\mathbb{Z}^T\mathbb{Z}^{-1})}} \sim t^{n-r-1}$
$\Rightarrow \mathcal{R} = \left\{ |T_j| > t^{n-r-1}_{1-\frac{\alpha}{2}} \right\}$

ANOVA È un test che serve per confrontare le medie dei gruppi all'interno di un modello lineare in cui le covariate sono categoriche. Richiede che i residui siano normali e omoschedastici.

Modelli lineari generalizzati Servono per estendere i modelli lineari ai casi in gui la variabile indipendente Y non è necessariamente gaussiana, ma appartiene alla EF. Bisogna specificare: la legge di Y, la matrice dei predittori $\mathbb{Z}$ e la link function $g(\cdot)$ ovvero quale funzione di $\mathbb{E}[Y]$ si vuole modellare.

Modello di regressione logistica La random component è $Y \sim \text{Be}(p_i)$ e la link function è quella canonica: $\text{logit}(p) = \log\left(\frac{p_i}{1-p_i}\right) = \nu_i = \beta_0 + \beta_1 z_{ij} + .. + \beta_r z_{ir}$. Dunque $p_i = \frac{e^{\nu_i}}{1+e^{\nu_i}}$ e le stime dei coeff non sono altro che l'odds-ratio relativo alla cov j-esima: $\text{OR}_j = e^{\beta_j}$.

Curva ROC per classificatori binari Grafico in cui x è 1-specificità e y è sensibilità