Modello statistico (parametrico) : $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$ ①

probabilità su Ω

Obiettivo : trovare la P^θ "migliore" ! ossia θ :
(dato un esperimento)

→ STIMA (θ) , → ESAUSTIONE (intervallo per θ) , → TEST (sotto o sopra una certa soglia)

(Es. $\Theta \subseteq \mathbb{N}$, $\Theta \subseteq \mathbb{R}$, $\Theta \subseteq \mathbb{R}^2$, ...)

Lavoriamo (quasi) sempre con modelli stat. DOMINATI :

$\exists$ misura σ-finita μ su $(\Omega, \mathcal{A})$ tale che $P^\theta \ll \mu$ per ogni $\theta \in \Theta$.

NOTE (1) μ σ-finita su $(\Omega, \mathcal{A})$ significa : esistono $\Omega_1, \Omega_2, \ldots, \Omega_n, \ldots$ in $\mathcal{A}$ che sono disgiunti, con $\mu(\Omega_n) < \infty$ $\forall n \geq 1$ e tali che $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$.

(2) $P^\theta \ll \mu$ ("assolutamente continua" per misure σ-finite) significa: per ogni $A \in \mathcal{F}$, se $\mu(A) = 0$ allora $P^\theta(A) = 0$. Ovvero, un evento trascurabile per μ è anche trascurabile per P^θ.

$\boxed{\text{Teorema}}$ (Radòn-Nikodym): per ν, μ σ-finite su $(\Omega, \mathcal{F})$, $\nu \ll \mu \iff \forall A \in \mathcal{F}$, $\boxed{\nu(A) = \int_A f \, d\mu}$ per una ed una sola applicazione $f \in L^1(\mu)$, $f \geq 0$ μ-q.o.

$\boxed{\text{NOTAZIONE}}$ $f =: \dfrac{d\nu}{d\mu}$ (i.e. "$\nu = f \cdot \mu$")

$\boxed{\text{OSS}}$ ν e μ sono "equivalenti" quando $\nu \ll \mu$ e anche $\mu \ll \nu$, ossia quando ν, μ hanno esattamente gli stessi trascurabili. In tal caso, $\exists!\, f \in L^1(\mu)$ tale che $f > 0$ μ-q.o., $\frac{1}{f} \in L^1(\nu)$, e $\nu(A) = \int_A f \, d\mu$ e $\mu(A) = \int_A \frac{1}{f} \, d\nu$ per ogni $A \in \mathcal{F}$.

(Cioè, $1/f = \dfrac{d\mu}{d\nu}$).

$\boxed{\text{Caso speciale}}$ Fra i modelli stat. $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ dominati da una μ σ-finita (su $\mathcal{F}$), quelli regolari sono tali che $P^{\theta_1} \ll P^{\theta_2}$ $\forall \theta_1, \theta_2 \in \Theta$ (cioè $P^{\theta_1} \sim P^{\theta_2}$ $\forall \theta_1, \theta_2 \in \Theta$), per cui qui P^θ è dominante. (classe di equivalenza!)

→ Il primo vantaggio (matematico) dei modelli stat. dominati è l'esistenza della VEROSIMIGLIANZA del modello stesso: precisamente, se $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ è dominato da μ, allora "la" verosimiglianza di tale modello è la funzione $L(\theta, \omega) : \Theta \times \Omega \to \mathbb{R}$,

$$L(\theta, \omega) := \frac{dP^\theta}{d\mu}(\omega)$$

(la "derivata" di Radòn-Nikodym, essendo appunto $P^\theta \ll \mu$).

(2)

$\Rightarrow \forall A \in \mathcal{F}$, e $\forall \theta \in \Theta$, $P^\theta[A] = \int_A L(\theta,\omega)\, d\mu(\omega)$. Anzi, più in generale, $(\forall \theta \in \Theta$ e$)$ $\forall$ per ogni $g \in \mathcal{L}^1(P^\theta)$ limitata, è $\int_\Omega g(\omega)\, dP^\theta(\omega) = \int_\Omega g(\omega) L(\theta,\omega)\, d\mu(\omega)$,

o anche
$$\boxed{\mathbb{E}^{P^\theta}[g] = \mathbb{E}^\mu[g\, L(\theta,\cdot)] \quad (=: \mathbb{E}^\theta[g])}.$$

Riassumendo, $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\}, \mu) \rightsquigarrow L^\theta = \dfrac{\partial P^\theta}{\partial \mu}$.

$\forall \theta \in \Theta$, P^θ (risp. $\ll \mu$ σ-finita)

OSS $\forall \theta \in \Theta$, $L(\theta,\cdot) = \dfrac{dP^\theta}{d\mu}(\cdot)$ sta in $\mathcal{L}^1(\mu)$!

OSS. Dato un mod. stat. DOMINATO $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$, con μ misura σ-finita dominante, possiamo supporre che μ sia e una volta una mis. di probabilità. (Dunque, una misura σ-finita è sempre equivalente ad una probabilità.)

Dim. La tesi è che esiste una probabilità ν su $\mathcal{F}$ dominante, cioè tale che $P^\theta \ll \nu$ per ogni $\theta \in \Theta$, sfruttando naturalmente che esiste per ipotesi una misura σ-finita μ su $\mathcal{F}$ dominante.

$\rightarrow$ Se $\Omega = \dot{\bigcup_{n \geq 1}} \Omega_n$, $\Omega_n \in \mathcal{F}$ tutti disgiunti con $\mu(\Omega_n) < \infty$, allora
$Z \doteq \sum_{n \geq 1} \dfrac{\mathbb{1}_{\Omega_n}}{2^n \mu(\Omega_n)}$ è chiaramente ben definita e $\boxed{>0}$, ed inoltre (ovviamente) sta in $\mathcal{L}^1(\mu)$ $\left(\sum_{n \geq 1} 2^{-n} = 1 < \infty\right)$, per cui ponendo
$$\boxed{\nu \doteq \left(\frac{Z}{\int_\Omega Z\, \partial\mu}\right) \cdot \mu}$$
(cioè $\nu(A) \doteq \int_A (\ldots)\, \partial\mu$, $A \in \mathcal{F}$) $(\nu \sim \mu$, ecc.$)$

abbiamo ottenuto la tesi. $\square$

Teorema (senza dim.....) Sia $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ mod. stat. avente probabilità dominante μ. Allora esiste una PROBABILITÀ DOMINANTE PRIVILEGIATA: esiste cioè un'altra probabilità dominante, che sia di una "forma speciale"; che sia precisamente una combinazione convessa numerabile delle P^θ:

esiste una succ. reale $(a_n)_{n\geq 1}$ con $a_n \geq 0$ e $\sum_{n\geq 1} a_n = 1$, ed esiste $(\theta_n)_{n\geq 1}$ in Θ, tali che la probabilità

$$\boxed{P_0 \doteq \sum_{n\geq 1} a_n P^{\theta_n}}$$

sia dominante (e, cioè, tale che $P^\theta \ll P_0$ per ogni $\theta \in \Theta$).

$\boxed{\text{OSS}}$ Dato un mod. stat. $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ dominato, e considerata quindi una probabilità dominante privilegiata $P_0 = \sum_{n\geq 1} a_n P^{\theta_n}$ (TEO.), abbiamo che: $\forall f \in \mathcal{L}^1(P_0)$, $E^{P_0}[f] = \sum_{n\geq 1} a_n E^{\theta_n}[f]$

$\boxed{\text{NOTA}}$ Non tutti i mod. stat. parametrici sono dominati .....

$\boxed{\text{RIASSUNTINO}}$:

per un mod. stat. parametrico $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$, essere che è dominato (da una misura σ-finita) equivale ad avere che esiste una probabilità dominante, e ciò equivale ad avere che esiste una probabilità dominante PRIVILEGIATA, ed in uno di questi casi (appunto equivalenti) ha senso il concetto di: VEROSIMIGLIANZA ($\leftarrow$ Radon-Nikodym)

PARENTESI PREPARATORIA.

(naturale ≥ 1 fissato!) (probabilità su $\mathcal{B}(\mathbb{R})$ fissata!)

Campione di taglia n $(\in \mathbb{N}_*)$ e di legge μ (su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) :

è un vettore aleatorio $(X_1, X_2, \dots, X_n)$ (in $\mathbb{R}^n$) su $(\Omega, \mathcal{F}, \cancel{\ldots})$ P tale che le v.a.r. $X_1, \dots, X_n$ siano mutuamente indipendenti e di legge μ : cioè $X_1(P) = \dots = X_n(P) = \mu$, e $(X_1, \dots, X_n)(P) = \mu^{\otimes n}$ (su $\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)$)).

Costruzione canonica : per mezzo dello schema delle prove indipendenti. Precisamente : $(\Omega, \mathcal{F}, P) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu^{\otimes n})$, ed ora (per ogni $i = 1, \dots, n$) $X_i(x_1, x_2, \dots, x_n) \doteq x_i$ (così $X_i = \xi_i = \pi_i$). Infatti, in questo modo, $(\Omega, \mathcal{F}, P)$ è uno spazio di probabilità ed qui

X_i è una v.a.r. su Ω. Inoltre, soprattutto, ~~ogni X_i ha legge μ (contiene della base degli insiemi cilindrici) per ogni $A_1, \dots, A_m \in B(\mathbb{R})$~~ le X_i sono tra loro indipendenti per il banale motivo che il vettore $(X_1, \dots, X_m)$ è l'identità di $(\mathbb{R}^m, B(\mathbb{R}^m))$, per cui ha ovviamente legge $\mu^{\otimes m}$. Infine ogni X_i ha legge μ: $\forall A \in B(\mathbb{R})$, $\overset{(\mu^{\otimes m})}{P}[X_i \in A] =$

$$= \overset{(\mu^{\otimes m})}{P}[X_i^{-1}(A)] \underset{(\text{Def. } X_i)}{=} \mu^{\otimes m}[\mathbb{R} \times \dots \times \mathbb{R} \times \underset{i\text{-esimo posto}}{A} \times \mathbb{R} \times \dots \times \mathbb{R}] \underset{(\text{Def. } \mu^{\otimes m})}{=} 1 \times \dots \times 1 \times \mu(A) \times 1 \times \dots \times 1 =$$

$= \mu(A)$, quindi effettivo, $\forall A \in B(\mathbb{R})$, $P[X_i^{-1}(A)] = \mu(A)$.]

Dati Θ insieme $\neq \emptyset$ (qualsiasi), $m \in \mathbb{N}_*$, e μ^ϑ legge di probabilità su $(\mathbb{R}, B(\mathbb{R}))$ al variare di $\vartheta \in \Theta$, esiste un modello statistico parametrico $(\Omega, \mathcal{A}, \{P^\vartheta \mid \vartheta \in \Theta\})$ sul quale sia definito per ogni $\vartheta \in \Theta$ un campione di taglia m e di legge μ^ϑ (se ne esiste uno più!), sia $(X_1^{(\vartheta)}, \dots, X_m^{(\vartheta)})$, ed inoltre tale modello è dominato se ~~(e solo se)~~ ogni μ^ϑ lo è: nel senso che $\exists$ misura μ σ-finita su $(\mathbb{R}, B(\mathbb{R}))$ tale che $\mu^\vartheta \ll \mu$ $\forall \vartheta \in \Theta$.

[Infatti, come già capito, basta considerare $\Omega \doteq \mathbb{R}^m$, $\mathcal{A} \doteq B(\mathbb{R}^m)$, e ora, $\forall \vartheta \in \Theta$, $P^\vartheta \doteq (\mu^\vartheta)^{\otimes m}$, quindi (per ogni $i = 1, \dots, m$) $X_i^{(\vartheta)}(\omega) = X_i^{(\vartheta)}(x_1, \dots, x_m) \doteq x_i$. Adesso, supponiamo che esiste una misura σ-finita μ su $(\mathbb{R}, B(\mathbb{R}))$ tale che $\mu^\vartheta \ll \mu$ $\forall \vartheta \in \Theta$ (ossia $\mu(A) = 0 \Rightarrow \mu^\vartheta(A) = 0$ per ogni $A \in B(\mathbb{R})$). Allora la misura $\mu^{\otimes m}$ su $(\mathbb{R}^m, B(\mathbb{R}^m))$ resta σ-finita, ed inoltre $(\mu^\vartheta)^{\otimes m} \ll \mu^{\otimes m}$.

→ In questo caso, infine, $\dfrac{d(\mu^\vartheta)^{\otimes m}}{d\mu^{\otimes m}}(x_1, \dots, x_m) = \displaystyle\prod_{i=1}^{m} \frac{d\mu^\vartheta}{d\mu}(x_i)$.]

$(= L(\vartheta; x_1, \dots, x_m))$

$\boxed{\text{RIASSUNTO}}$: esiste un modello statistico parametrico dominato $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ tale per cui, assegnati anche $n \in \mathbb{N}_*$ e μ^θ legge di probabilità su $\mathcal{B}(\mathbb{R})$ ($\forall \theta \in \Theta$) oltre che l'insieme Θ ($\neq \emptyset$), con μ^θ dominate da μ, esiste ($\forall P^\theta$ scelta) un campione di taglia n e di legge μ^θ su $(\Omega, \mathcal{F}, P^\theta)$. Se $(X_1, \dots, X_n)$ è tale campione, allora possiamo scrivere: $\omega = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$X_i(x_1, \dots, x_n) = x_i$, $\quad L(\theta, \omega) = L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{d\mu^\theta}{d\mu}(x_i)$

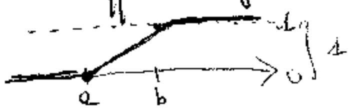
(ES. *) Controllo qualità: fissato $N \in \mathbb{N}_*$ ("non piccolo"), il mod. stat. binomiale è $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ con, naturalmente,

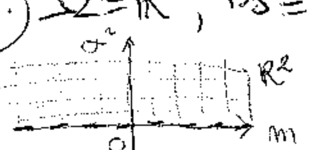
$\Omega \doteq \{0, 1, 2, 3, \dots, N-1, N\}$, $\mathcal{F} \doteq \mathcal{P}(\Omega)$, $\Theta \doteq (0,1)$ ($\subsetneq \mathbb{R}$), e per ogni $k \in \Omega$ $P^\theta[\{k\}] \doteq \binom{N}{k} \theta^k (1-\theta)^{N-k}$. ("Dunque" θ è la probabilità "di essere difettoso" in una singola estrazione.)

(?) Questo modello è dominato? Certamente, perché esso è regolare. Infatti, $\forall \theta \in \Theta$, P^θ NON ha trascurabili che siano $\neq \emptyset$, per cui ovviamente $P^{\theta_1} \sim P^{\theta_2}$ $\forall \theta_1, \theta_2 \in \Theta$. ($\boxed{\text{NOTA}}$ Una misura σ-finita dominante può essere $\mu(\{k\}) \doteq 1$ $\forall k \in \Omega$, o $\nu(\{k\}) \doteq \frac{1}{N}$ volendo una probabilità dominante.)

$\boxed{\text{EX}}$ Studiamo il precedente modello: $\Omega \doteq \mathbb{N}$, $\mathcal{F} \doteq \mathcal{P}(\Omega)$, $\Theta \doteq (0,1)_p \times \mathbb{N}_n$, $P^\theta(\{k\}) \doteq \binom{n}{k} p^k (1-p)^{n-k}$ per ogni $k \in \Omega$ ed ogni $\theta = (p, n) \in \Theta$. Allora tale modello è dominato ma NON regolare. [Infatti, semplicemente, ogni P^θ ha dei trascurabili $\neq \emptyset$ ($\forall \theta = (p,n) \in \Theta$, $P^\theta[\{k\}] = 0$ per ogni $k > n$!) tale per cui NON si ha che le P^θ siano mutuamente equivalenti, per cui NON si ha regolarità.

Tuttavia, $\mu(\{k\}) \doteq 1 \ \forall k \in \Omega$ resta una misura σ-finita (ma NON finita) dominante, non essendo trascurabili $\neq \phi$.] ④

(ES.) $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\Theta = \{(a,b) \in \mathbb{R}^2 \mid a<b\} =$ , e $\forall \theta = (a,b) \in \Theta$ $P^\theta \doteq$ «la legge uniforme $\mathcal{U}(a,b)$ su $[a,b]$» (cioè quella corrispondente alla f.d.r. ). Di nuovo, analogamente, questo modello non è regolare: qui P^θ è diffusa ma ha (ma "tanti" intervalli ($\neq \phi$)) trascurabili, in base a θ stesso. Resta evidente tuttavia che il modello è dominato dalla misura di Lebesgue su $\mathcal{B}(\mathbb{R})$.

(ES.) $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\Theta = \{(m,\sigma^2) \in \mathbb{R}^2 \mid m \in \mathbb{R}, \sigma^2 \in (0,\infty)\} = \mathbb{R}_m \times (0,\infty)_{\sigma^2}$ $=$ , $P^{\theta} \doteq$ «la legge gaussiana di media m e varianza σ^2 $N(m,\sigma^2)$», equivalente alla misura di Lebesgue su $\mathcal{B}(\mathbb{R})$, di densità $\mathcal{L}(\theta, x) \doteq$

$$\doteq f^\theta(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right) \ \forall x \in \mathbb{R} \quad (\forall (m,\sigma^2) \in \Theta)$$

(>0). Tutti e soli i trascurabili per la generica P^θ sono i singoletti, per cui il modello è dominato in questo caso è regolare.

[EX] Il modello $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\varepsilon_x \mid x \in \mathbb{R}\})$ NON è dominato

[Infatti, ricordiamo che, $\forall x \in \mathbb{R}$ e $\forall A \in \mathcal{B}(\mathbb{R})$, $\varepsilon_x(A) \doteq \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{altrimenti} \end{cases}$ (che non ha alcuna relazione di assoluta continuità, ad es., con la misura di Lebesgue su $\mathcal{B}(\mathbb{R})$). Ragionando per assurdo, se lo fosse allora (TEO.!) esisterebbero $(a_n)_{n\geq 1}$ in $[0,\infty)$ con $\sum_{n\geq 1} a_n = 1$ e $(x_n)_{n\geq 1}$ su $\mathbb{R}$ tali per cui, $\forall x \in \mathbb{R}$, $\varepsilon_x \ll \sum_{n \geq 1} a_n \varepsilon_{x_n}$ ($\leftarrow$ probabilità dominante privilegiata i cui NON trascurabili sono "generabili" a partire dall'insieme discreto $\{x_n \mid n \geq 1\} = \{x_1, x_2, \ldots, x_n, \ldots\}$): è impossibile, potendo prendere ad es. qualsiasi $\underset{(\{x\})}{x} \notin \{x_n \mid n \geq 1\}$.]

All'interno di un modello statistico param. dominato, affrontiamo il problema della STIMA di una funzione di $\theta \in \Theta$, $g(\theta)$.

Sia dunque $(\Omega, \mathcal{B}, \{P^\theta \mid \theta \in \Theta\})$ un mod. stat. parametrico (dominato...), e sia $D \subseteq \mathbb{R}$ un aperto $(\neq \emptyset)$. Sia $g(\theta): \Theta \to D$, e sia $U(\omega): \Omega \to D$ una v.a.(r.) : l'idea - molto generale - è che U (come v.a.r.) sia STIMATORE di g (come funzione), mentre di conseguenza $U(\omega)$ sia una STIMA di $g(\theta)$ quali che siano $\omega \in \Omega$ e $\theta \in \Theta$.

Raffiniamo il concetto (considerando anche ...) : avendo in mente quest'altra idea per lo stesso, fissiamo una funzione costo (/ funzione errore)

$$C : \Theta_\theta \times \mathbb{R}_a \to [0, \infty) \quad , \text{ ossia } C(\theta, a) \geq 0 ,$$

(o D)

per esempio $\boxed{C(\theta, a) \doteq (g(\theta) - a)^2}$ ← (dipendente da θ attraverso la g!),

la quale anzi è la funzione costo "standard" della teoria.

Allora vorremmo "minimizzare" $C(\theta, U(\omega))$, per ogni $\theta \in \Theta$, al variare di ~~per ogni~~ $\omega \in \Omega$ $(= (g(\theta) - U(\omega))^2)$

... ▶ Per ogni $\theta \in \Theta$, definiamo il RISCHIO di U nello stimare g (relativo alla funzione costo C) come $(\forall \theta \in \Theta)$

$$\boxed{R_U(\theta) \doteq \mathbb{E}^{P^\theta}[C(\theta, U(\cdot))]} \underset{\substack{\uparrow \\ (C \text{ canonica})}}{=} \mathbb{E}^{P^\theta}[(U - g(\theta))^2] \text{ nel caso}$$

che $C(\theta, a) = (g(\theta) - a)^2$ (ma "rischio quadratico di U").

(U normale)

OSS Più $R_U(\theta)$ è vicino allo zero, più U può esser considerato (come v.a.r.) un buon stimatore di $g(\theta)$ (per ogni $\theta \in \Theta$ fissato), almeno in P^θ - media, al caso estremo seguente :

$R_U(\vartheta) = \mathbb{E}^{P^\vartheta}[|U - g(\vartheta)|^2] = 0 \overset{(\forall \vartheta \in \Theta \text{ fissato})}{\Leftrightarrow} U \equiv g(\vartheta)$ (costante), che in effetti sarebbe proprio l'ideale, ma (costante) (U limitato...) vabbè!

OSS $R_U(\vartheta) = \mathbb{E}^{P^\vartheta}[(U - g(\vartheta))^2] \underset{(\pm)}{=} \mathbb{E}^{P^\vartheta}[\{(U - \mathbb{E}^{P^\vartheta}[U]) + (\mathbb{E}^{P^\vartheta}[U] - g(\vartheta))\}^2] =$

$\underset{(\text{ovvio})}{=} \mathbb{E}^{P^\vartheta}[\{U - \mathbb{E}^{P^\vartheta}[U]\}^2] + \{\mathbb{E}^{P^\vartheta}[U] - g(\vartheta)\}^2 =$

$= \mathrm{Var}^{P^\vartheta}[U] + \{\mathbb{E}^{P^\vartheta}[U] - g(\vartheta)\}^2$.

(! Stiamo supponendo che gli stimatori stiano in $\mathcal{L}^2(P^\vartheta)$ per ogni $\vartheta \in \Theta$! $\Rightarrow R_U(\vartheta) \geq \mathrm{Var}^{P^\vartheta}[U]$, ed ci $=$ se, e solo se, $\mathbb{E}^{P^\vartheta}[U] = g(\vartheta)$)

$\Rightarrow U : \Omega \to D$ è uno STIMATORE CORRETTO di g quando, per ogni $\vartheta \in \Theta$, $\boxed{\mathbb{E}^{P^\vartheta}[U] = g(\vartheta)}$ e cioè (appunto sopra)

$R_U(\vartheta) = \mathbb{E}^{P^\vartheta}[(U - g(\vartheta))^2] = \mathrm{Var}^{P^\vartheta}[U]$, che è il minimo possibile ($\forall \vartheta \in \Theta$) per U.

(Uno stimatore U è tanto migliore quanto più, $\forall \vartheta \in \Theta$, $\mathbb{E}^{P^\vartheta}[U]$ si avvicina a $g(\vartheta)$!)

(RELAZIONI FRA STIMATORI)

$\rightarrow$ Siano $U, V : \Omega \to D$ due stimatori di g. Allora U è PREFERIBILE a V se, $\forall \vartheta \in \Theta$, $R_U(\vartheta) \leq R_V(\vartheta)$, ed in tal caso U è STRETTAMENTE PREFERIBILE a V se esiste un $\vartheta_0 \in \Theta$ tale che $R_U(\vartheta_0) < R_V(\vartheta_0)$.

Immaginando quindi una famiglia $(U_i)_{i \in I}$ di stimatori di g, e ponendo $\mathcal{D} \doteq \{U_i \mid i \in I\}$ il loro insieme, consideriamo una $U \in \mathcal{D}$ AMMISSIBILE (relativamente a $\mathcal{D}$) nel caso che NON esistano $V \in \mathcal{D}$ strettamente preferibili rispetto alla U; consideriamo una $U \in \mathcal{D}$ OTTIMALE (relativamente a $\mathcal{D}$) se è preferibile rispetto ad ogni altro membro di $\mathcal{D}$. (In particolare, uno stimatore ottimale è, a maggior ragione, ammissibile.)

Naturalmente, due stimatori potrebbero NON esser confrontabili nel senso della preferibilmente ...! { **RIASSUNTINO**: assegnato un mod. stat. param. $(\Omega, \mathcal{F}, \{P^\theta \mid \theta\in\Theta\})$, assegnato un aperto $\neq\emptyset$ $D \subseteq \mathbb{R}$, ed assegnata una funzione $g(\theta) : \Theta \to D$ "da stimare", cerchiamo una v.a.r. $U : \Omega \to D$ "sommabile" ("REGOLARE") (per es. rispetto ad una prob. dominante; o limitata) tale che, $\forall \theta\in\Theta$, in riferimento allo spazio di probabilità $(\Omega, \mathcal{F}, P^\theta)$, risulti che $R_U(\theta) := E^{P^\theta}[|U-g(\theta)|^2]$ sia il più vicino possibile alla $Var^{P^\theta}[U]$, e cioè che $E^{P^\theta}[U]$ sia il più vicino possibile a $g(\theta)$ stesso (sempre per ogni $\theta\in\Theta$). (Il caso ottimale è avere $R_U(\theta) = Var^{P^\theta}[U] = 0$ e cioè $U \equiv g(\theta)$.) ✓

Detto questo, due stimatori "buoni" U, V si confrontano guardando proprio $R_U(\theta)$ e $R_V(\theta)$ al variare di $\theta\in\Theta$. ✓✓

(ES.) Torniamo al controllo di qualità "standard": supponiamo $n\in\mathbb{N}_*$ e mettiamoci nel mod. stat. param. (regolare) $(\Omega, \mathcal{F}, \{P^\theta \mid \theta\in\Theta\})$ dove $\Omega = \{0,1,\dots,N\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\Theta = (0,1)$, $P^\theta[\{k\}] = \binom{n}{k}\theta^k(1-\theta)^{n-k}$ per ogni $k\in\Omega$ (cioè $k = 0,1,\dots,N$). Problema di stima: $D = \Theta$ e $g : \Theta \to D = \Theta$ l'identità, e cioè $\boxed{\text{stimare } g(\theta) = \theta}$. Proviamo ad usare stimatori $U : \Omega \to D = \Theta$ della forma

$$U = \frac{\sum_{i=1}^{n} X_i + a}{n+b}$$

con $X_1, \dots, X_n$ indipendenti e di legge $B(1,\theta)$, e con $a, b \in [0,\infty)$ in modo che U sia a valori in Θ.

⚠ Detto così NON va bene ...! Il modo giusto è: consideriamo

$n \in \mathbb{N}_*$ ed un campione $(X_1, \dots, X_n)$ di taglia n e di legge $B(1,\vartheta)$, $\vartheta \in (0,1)$, relativo al mod. stat. param. (regolare) che è possibile costruire canonicamente. (Sia questo $(\Omega, \mathcal{F}, \{P^\vartheta \mid \vartheta \in \Theta\})$, $\Theta = (0,1)$.
In particolare, $X_1, \dots, X_n$ sono INDIP. ed esatti $\mathbb{E}^{P^\vartheta}[X_i] = \vartheta$ e $Var^{P^\vartheta}[X_i] = \vartheta(1-\vartheta)$ per ogni $\vartheta \in \Theta$ (ed ogni $i = 1, \dots, n$).

$\boxed{\text{NOTAZIONI}}$:

$$S_n := X_1 + \dots + X_n = \sum_{i=1}^n X_i$$

$$X := (X_1, \dots, X_n)$$

$$\overline{X} := \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{("frequenza empirica") (o "media campionaria")}$$

$\boxed{\text{OSS}}$ $\forall \vartheta \in \Theta$, $\mathbb{E}^{P^\vartheta}[S_n] = n\vartheta$ (linearità), mentre per indipendenza ($\Rightarrow$ scorrelazione) $Var^{P^\vartheta}[S_n] = n\vartheta(1-\vartheta)$. (Ovvero, $\mathbb{E}^{P^\vartheta}[\overline{X}] = \vartheta$ e $Var^{P^\vartheta}[\overline{X}] = \frac{1}{n}\vartheta(1-\vartheta)$.)

Problema di stima: stimare $g(\vartheta): \Theta \to D := \Theta = (0,1)$, $\boxed{g(\vartheta) = \vartheta}$.
(Cioè STIMARE ϑ.) Proviamo con le v.a.r. $U : \Omega \to D = \Theta = (0,1)$ della forma (vedi D...)

$$\boxed{U = \frac{S_n + a}{n+b}} = \frac{\sum_{i=1}^n X_i + a}{n+b}, \text{ dove}$$

$a, b \in [0, \infty)$ parametri fissi tali che però $U(\omega) \in (0,1)$ $\forall \omega \in \Omega$.

(1) $\forall \vartheta \in \Theta$, $\mathbb{E}^{P^\vartheta}[U] = \frac{n\vartheta}{n+b} + \frac{a}{n+b}$ $\left(U = \frac{S_n}{n+b} + \frac{a}{n+b}\right)$, per cui chiaramente

$$U = \frac{S_n}{n} = \overline{X} \text{ è uno stimatore corretto di } g(\vartheta) = \vartheta \quad (\in (0,1)!) .$$

[perché, appunto, per $a = b = 0$ avrei $\mathbb{E}^{P^\vartheta}[U] = \vartheta = g(\vartheta)$.] ($U = \overline{X}$)

(2) Calcoliamo, $\forall \vartheta \in \Theta$, il rischio (quadratico) di U in ϑ:

$$R_U(\vartheta) \overset{(\text{def.})}{=} \mathbb{E}^{P^\vartheta}\left[|U - g(\vartheta)|^2\right] \; \cancel{\mathbb{E}^{P^\vartheta}\left[\left|\frac{S_n + a}{n+b} - \vartheta\right|^2\right]} =$$

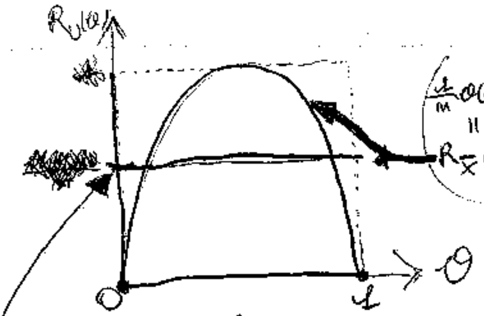
$$= \mathrm{Var}^{p_\theta}[U] + \Big[\underbrace{\mathbb{E}^{p_\theta}[U]}_{\left(\frac{m\theta+a}{m+b}\ \text{(visto ora)}\right)} - \overbrace{p(\theta)}^{(\theta!)}\Big]^2 =$$

(visto!)

$$= \underbrace{\mathrm{Var}^{p_\theta}\left[\frac{S_m+a}{m+b}\right]}_{\left(\frac{1}{(m+b)^2}\mathrm{Var}^{p_\theta}[S_m] \overset{(\text{indip.})}{=} \frac{m\theta(1-\theta)}{(m+b)^2}\right)} + \Big(\underbrace{\frac{m\theta+a}{m+b} - \theta}_{\left(\frac{a-b\theta}{m+b}\right)}\Big)^2 =$$

$$= \frac{m\theta(1-\theta) + (a-b\theta)^2}{(m+b)^2} = \frac{(b^2-m)\theta^2 + (m-2ab)\theta + a^2}{(m+b)^2} \quad \checkmark$$

Due casi particolari:



(Costante $= \dfrac{a^2}{(m+b)^2}$ per $b=\sqrt{m}$ e $a=\dfrac{\sqrt{m}}{2}=\dfrac{b}{2}$)

→ (esistono valori per a e b tali per cui i relativi stimatori $U=\dfrac{S_m+a}{m+b}$ risultano NON confrontabili nel senso della preferibilità!)

$\boxed{\text{NOTA}}$ Sia sempre $(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in(0,1)\})$ un mod. stat. param. (regolare) in corrispondenza del quale esiste un campione di taglia m ($\in\mathbb{N}^*$, fissato) e di legge $B(1,\theta)$, $\theta\in(0,1)$, sia $(X_1,\dots,X_m)$ questo campione. Abbiamo trovato che uno stimatore corretto del parametro $\theta\in(0,1)$ è la v.a.

$$\bar X = \frac{S_m}{m} = \frac{1}{m}\sum_{i=1}^m X_i \qquad \Big(\text{cioè}, \forall\theta\in(0,1),\ \mathbb{E}^{p_\theta}[\bar X]=\theta\Big).$$

(e NON serve l'indip.)

(?) Cosa dovrebbe significare nella pratica?

$\boxed{\text{Caso } m=1}$ $\bar X = X_1 \begin{cases} 1 & \text{prob. } \theta \\ 0 & \text{prob. } 1-\theta \end{cases}$ ⇒ se $\bar X(\omega)=1$, allora dovremmo concludere $\theta=1$! (Tutti difettosi)

ROZZO

$\boxed{\text{Caso } m=2}$ $\bar X = \dfrac{X_1+X_2}{2} \begin{cases} 1 & \text{prob. } \theta^2 \\ \frac{1}{2} & \text{prob. } \dots \\ 0 & \text{prob. } \dots \end{cases}$ ⇒ $\begin{cases} \theta=1 \\ \theta=\frac{1}{2} \\ \theta=0 \end{cases}$ rispettivamente!

Dunque, metodo rozzo per m bassi ... (Soltanto sostituendo le prob. con [illegible])

EX. (GENERALIZZ.) **OSS.** (preliminare) Scegliamo $m \in \mathbb{N}_*$, un insieme Θ $(\neq \emptyset)$, ed una legge di probabilità μ^θ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ per ogni $\theta \in \Theta$. Consideriamo allora un modello stat. param. $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ sul quale esiste un campione $(X_1, \dots, X_m)$ di taglia m e di legge "μ^θ" (in corrisp. di P^θ).

Supponiamo che il campione sia di quadrato integrabile, nel senso che $(\forall \theta \in \Theta)$ $\int_{-\infty}^{+\infty} x^2 \, d\mu^\theta(x) < \infty$. Poniamo $m^\theta = \int_{-\infty}^{+\infty} x \, d\mu^\theta(x)$ (che sta in $\mathbb{R}$ e meglio ancora), per cui $(\forall \theta \in \Theta)$ $\mathbb{E}^{P^\theta}[X_i] = m^\theta$ $\forall i = 1, \dots, m$, ed inoltre $\mathrm{Var}^{P^\theta}[X_i] = \int_{\mathbb{R}} x^2 d\mu^\theta(x) - (m^\theta)^2 =: \sigma_\theta^2$. ~~Supponiamo $m^\theta \neq 0$~~.

IDEA GENERALE (della teoria della stima): stimare "tutti" i momenti di un dato campione per stimarne la legge (univocamente associata) !!

NOTA Se avessimo una $(X_n)_{n \geq 1}$ di v.a.r. i.i.d. di legge μ^θ, $\forall \theta \in \Theta$, ed integrabili, allora la legge forte dei grandi numeri di Kolmogorov darebbe la convergenza, $\forall \theta \in \Theta$, $\overline{X} = \frac{S_m}{m} = \frac{1}{m}\sum_{i=1}^m X_i \xrightarrow[m \to \infty]{P^\theta\text{-q.c.}} \mathbb{E}^{P^\theta}[X_1] = m^\theta$ ...!

EX **a** Fra tutti gli stimatori corretti di m^θ $(= g(\theta)!)$ della forma $U = \sum_{i=1}^m e_i X_i$ $(e_1, \dots, e_m \in \mathbb{R})$, la media campionaria $\overline{X} = \frac{S_m}{m}$ $\left(e_1 = \dots = e_m = \frac{1}{m}\right)$ è (corretto ed) ottimale.

b La "varianza campionaria"

$$\boxed{S^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \overline{X})^2}$$

è uno stimatore corretto della varianza $\sigma_\theta^2 = \int_{\mathbb{R}} (x - m^\theta)^2 d\mu^\theta(x)$

Svolgo. **a** Anzitutto osserviamo che, $\forall \theta \in \Theta$, $\mathbb{E}^{P^\theta}\left[\sum_{i=1}^m e_i X_i\right] = m^\theta \sum_{i=1}^m e_i$, per cui $U = \sum_{i=1}^m e_i X_i$ è corretto se, e solo se, $\sum_{i=1}^m e_i = 1$ (corretto nello stimare m^θ, ovviamente), a meno che $m^\theta = 0$. Comunque è cosicché che $\mathbb{E}^{P^\theta}[\overline{X}] = m^\theta$, per cui $\overline{X}$ è stimatore corretto di m^θ.

Ora, ricordando che $R_U(\theta) = Var^{P^\theta}[U]$ per stimatori corretti; ma $U = \sum_{i=1}^{m} e_i X_i$ e X_i indip. $\Rightarrow Var^{P^\theta}[U] = \sum_{i=1}^{m} e_i^2 Var^{P^\theta}[X_i] =$

$= \sigma_\theta^2 \sum_{i=1}^{m} e_i^2 \quad \left(= \frac{\sigma_\theta^2}{m} \text{ se } U = \overline{X} \right) \ldots$

▶ $\forall e_1, \ldots, e_m \in \mathbb{R}$ con $\sum_{i=1}^{m} e_i = 1$, vale $\sum_{i=1}^{m} e_i^2 \geq \frac{1}{m}$.

[Mettiamoci in $(\mathbb{R}^m, \langle\cdot;\cdot\rangle_2)$ e consideriamo i due vettori $\begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}$ e $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$: per ipotesi, hanno prod. scalare unitario, cioè effettivamente $\sum_{i=1}^{m} e_i = 1$. Ma, adesso, grazie alla dis. di Cauchy-Schwarz, $\underset{1^2}{1} = \sum_{i=1}^{m} e_i = \left\langle \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} ; \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\rangle_2 \overset{(C\text{-}S)}{\leq} \left\| \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} \right\|_2^2 \cdot \left\| \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\|_2^2 =$

$= \left(\sum_{i=1}^{m} e_i^2 \right) \cdot m$, cioè effettivamente $\frac{1}{m} \leq \sum_{i=1}^{m} e_i^2$. ✓]

$\boxed{b}$ ($\boxed{\text{IDEA}}$ Di nuovo, $\frac{1}{m} \underbrace{\sum_{i=1}^{m} (X_i - m^\theta)^2}_{\text{(i.i.d. con speranza } \sigma_\theta^2 !)} \xrightarrow[m \to \infty]{P^\theta\text{-q.c.}} \sigma_\theta^2$, MA non avendo σ_θ^2 "probabilmente" già non avevo m^θ...! Ecco che viene da provare con $\frac{1}{m} \sum_{i=1}^{m} (X_i - \overline{X})^2$! ...) Facciamo "dei calcoli" $(\mathbb{E}^{P^\theta}[\cdot])$ per $\sum_{i=1}^{m} (X_i - \overline{X})^2$:

$\mathbb{E}^{P^\theta}\left[\sum_{i=1}^{m} (X_i - \overline{X})^2 \right] = \sum_{i=1}^{m} \mathbb{E}^{P^\theta}[(X_i - \overline{X})^2]$; adesso, $\forall i$, $X_i - \overline{X} =$

$= (X_i - m^\theta) + (m^\theta - \overline{X})$, il quale ora elevato al quadrato e calcolato $\mathbb{E}^{P^\theta}[\cdot]$ (per poi sommare il tutto) :

$\rightarrow (X_i - \overline{X})^2 = (X_i - m^\theta)^2 + 2(X_i - m^\theta)(m^\theta - \overline{X}) + (\overline{X} - m^\theta)^2$

$\rightarrow \mathbb{E}^{P^\theta}[(X_i - m^\theta)^2] = \sigma_\theta^2$

$\rightarrow \mathbb{E}^{P^\theta}[(\overline{X} - \underset{(\text{già } m^\theta)}{m^\theta})^2] \underset{(\boxed{a}!)}{=} Var^{P^\theta}[\overline{X}] \overset{(\text{indip.})}{=} \frac{\sigma_\theta^2}{m}$

$$\to \sum_{i=1}^{n} 2(X_i - m^\theta)(m^\theta - \overline{X}) = 2(m^\theta - \overline{X}) \underbrace{\sum_{i=1}^{n}(X_i - m^\theta)}_{=\ (n(\overline{X} - m^\theta))} =$$

$$= -2n(\overline{X} - m^\theta)^2 .$$ Per definitiva,

$$E^{P^\theta}\Big[\sum_{i=1}^{n}(X_i - \overline{X})^2\Big] = n\sigma_\theta^2 + \sigma_\theta^2 \underbrace{- 2n\frac{\sigma_\theta^2}{n}}_{=\ (-\sigma_\theta^2)} = (n-1)\sigma_\theta^2 ,$$

e cioè effettivamente $E^{P^\theta}[S^2] = E^{P^\theta}\Big[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\Big] = \sigma_\theta^2$ □.

$\boxed{\text{RIASSUNTINO}}$: scelti $n \in \mathbb{N}_*$, Θ insieme $\neq \emptyset$, e μ^θ legge di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ per ogni $\theta \in \Theta$ (possibilmente dominate), ed inseriti in un mod. stat. param. $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ sul quale esiste un campione $(X_1, \dots, X_n)$ di taglia n e di legge μ^θ ($\forall \theta \in \Theta$), stimiamo il meglio possibile i momenti delle X_i, supponendoli a seconda sufficientemente sommabili, e partiamo dalla media $m^\theta := \int_{\mathbb{R}} x \, d\mu^\theta(x)$ e dalla varianza $\sigma_\theta^2 := \int_{\mathbb{R}} (x - m^\theta)^2 d\mu^\theta(x)$, con la solita idea teorica di stimatore corretto: dato un aperto $\neq \emptyset$ $D \subseteq \mathbb{R}$ e data una funzione $g(\theta): \Theta \to D$ da stimare, cerchiamo una v.a. $U: \Omega \to D$ tale che, $\forall \theta \in \Theta$, $E^{P^\theta}[|U - g(\theta)|^2] \approx Var^{P^\theta}[U]$ e cioè

$\boxed{E^{P^\theta}[U] \approx g(\theta)}$ ($\to U$ stimatore corretto di g).

Ebbene, abbiamo trovato che $\overline{X} = \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ è uno stimatore corretto di m^θ (ed è evidente), e soprattutto che $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ è uno stimatore corretto di σ_θ^2.

Su un mod. stat. param. $(\Omega, \mathcal{A}, \{P^\vartheta \mid \vartheta \in \Theta\})$, per il momento, abbiamo incontrato solo v.a.r. NON-dipendenti da $\vartheta \in \Theta$, come le X_i di un campione o gli stimatori (di una qualche funzione $g(\vartheta)$).

($\boxed{\text{NOTA}}$ Anzi, negli esempi visti, gli stimatori erano funzioni delle X_i!)
(se ciò non è una coincidenza, per DOOB.....)

All'interno quindi di un tale mod. stat. param., chiamiamo STATISTICA una qualsiasi funzione misurabile su $(\Omega, \mathcal{A})$ che NON dipende da Θ: quindi $S : (\Omega, \mathcal{A}) \to \underbrace{(E, \mathcal{E})}_{\text{(altro spazio misurabile)}}$ misurabile è una statistica.

Ovviamente, esisteranno statistiche più interessanti di altre rispetto al nostro modello stat. ...

$\boxed{\text{Breve digressione}}$ (sulla speranza condizionale). Partendo sempre dal mod. stat. $(\Omega, \mathcal{A}, \{P^\vartheta \mid \vartheta \in \Theta\})$, l'"ideale-idealissimo" sarebbe quello di conoscere, $(\forall \vartheta \in \Theta,)$ $P^\vartheta[A]$ per ogni $A \in \mathcal{A}$. L'IDEA naturale è ora quella di raggiungere tale traguardo avendo le sole conoscenze di $P^\vartheta[A']$ su ogni A' in un'opportuna sotto-σ-algebra di $\mathcal{A}$! Più precisamente, se $\mathcal{B} \subseteq \mathcal{A}$ è una σ-algebra, allora: $\forall A \in \mathcal{A}$, $P^\vartheta[A] = \mathbb{E}^{P^\vartheta}[\mathbb{1}_A] = \mathbb{E}^{P^\vartheta}[\mathbb{E}^{P^\vartheta}[\mathbb{1}_A \mid \mathcal{B}]] =$

$$= \int_\Omega \mathbb{E}^{P^\vartheta}[\mathbb{1}_A \mid \mathcal{B}](\omega)\, dP^\vartheta(\omega) \underset{\substack{(\mathbb{E}^{P^\vartheta}[\mathbb{1}_A \mid \mathcal{B}] \text{ è} \\ \mathcal{B}\text{-misurabile!})}}{=} \int_\Omega \mathbb{E}^{P^\vartheta}[\mathbb{1}_A \mid \mathcal{B}](\omega)\, d(P^\vartheta|_{\mathcal{B}})(\omega).$$

Invece $\mathbb{1}_A$ NON è $\mathcal{B}$-misurabile, in generale, ovviamente ...

Però, appunto, se $\mathcal{B}$ fosse speciale a tal punto che $\mathcal{B} \subsetneq \mathcal{A}$ e $\forall Y$ statistica reale limitata (ad es.), $\mathbb{E}^{P^\vartheta}[Y \mid \mathcal{B}]$ NON dipende da $\vartheta \in \Theta$, per cui "$\mathbb{E}^{P^\vartheta}[Y \mid \mathcal{B}] = \mathbb{E}^{\cdot}[Y \mid \mathcal{B}]$",

(ad es., possiamo immaginare il modello regolare ...)

allora affermato $\sqrt{\text{la formula}}$ $P^{\theta}[A] = \int_{\Omega} \mathbb{E}^{\cdot}[\mathbb{1}_A | \mathcal{B}](\omega)\, d(P^{\theta}|_{\mathcal{B}})(\omega) \quad \forall \theta \in \Theta$ e $\forall A \in \mathcal{F}$ (9)

mostra che conosciamo ogni P^{θ} su tutta $\mathcal{F}$ se le conosciamo solo su tutta $\mathcal{B}$.

→ Una STATISTICA SUFFICIENTE, o ESAUSTIVA, oppure anche un RIASSUNTO ESAUSTIVO, su un mod. stat. $(\Omega, \mathcal{F}, \{P^{\theta} | \theta \in \Theta\})$ dominato da μ, è una statistica $T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ "esaustiva", appunto, nel senso seguente: per ogni v.a.r. Y limitata, esiste μ-q.o. una versione di $\mathbb{E}^{P^{\theta}}[Y|T]$ ("$\mathcal{B} = \sigma(T)$", nelle notazioni prec.) che ($\equiv \mathbb{E}^{P^{\theta}}(Y|\sigma(T))$) NON dipende da $\theta \in \Theta$. Brevemente, $\forall \theta \in \Theta$,

"$\mathbb{E}^{P^{\theta}}[Y|T] = \mathbb{E}^{\cdot}[Y|T]$". (→ una sola v.a.r.!)

In questo modo, appunto, conoscere P^{θ} su tutta $\mathcal{F}$ significa conoscere P^{θ} sulla sola $\sigma(T)$, appunto perché $\forall A \in \mathcal{F}$

$$P^{\theta}[A] = \int_{\Omega} \mathbb{E}^{\cdot}[\mathbb{1}_A | T](\omega)\, d(P^{\theta}|_{\sigma(T)})(\omega) \qquad (\forall \theta \in \Theta, \text{sempre}).$$

$\boxed{\text{Caso particolare.}}$ Scelti $m \in \mathbb{N}^*$, Θ un insieme $\neq \emptyset$, e μ^{θ} leggi di probabilità (assolutamente continue) su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ per ogni $\theta \in \Theta$, sia $(\Omega, \mathcal{F}, \{P^{\theta} | \theta \in \Theta\})$ il mod. stat. parametrico dominato (dal quale) esiste un campione $(X_1, \dots, X_m)$ di taglia m e di legge μ^{θ} $(\forall \theta \in \Theta)$. Allora, com'è chiaro, $\mathcal{F}$ è esattamente $\sigma(X_1, \dots, X_m)$, per cui tutte le statistiche sono della forma $\varphi(X_1, \dots, X_m)$ (in virtù del criterio di misurabilità di Doob), (~~φ INDIP. da $\theta \in \Theta$~~) dove naturalmente $\varphi : (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)) \to (E, \mathcal{E})$ è misurabile ed INDIP. da $\theta \in \Theta$. In questo caso, una statistica T è esaustiva se, per ciascuna di tali φ che sono $\underline{\text{im } (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ e ivi}}$ limitate, risulta $\mathbb{E}^{P^\vartheta}[\varphi(X_1, \dots, X_n) \mid T]$ indipendente da $\vartheta \in \Theta$.

(**NOTA** In questo caso, anche T sarà a sua volta funzione mis. delle X_i, naturalmente!)

ES. Fissiamo: $n \in \mathbb{N}^*_*$, $\vartheta \in \Theta \doteq (0, \infty)$, e μ^ϑ la legge di Poisson di parametro ϑ ("$\mathcal{P}(\vartheta)$", o "$\Pi(\vartheta)$"). In altri termini, μ^ϑ è quella legge di probabilità "su $\mathbb{N}$" (su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) così definita:

$\forall k = 0, 1, 2, \dots$, $\quad \mu^\vartheta(\{k\}) \doteq \dfrac{\vartheta^k}{k!} e^{-\vartheta}$, mentre

$\mu^\vartheta(\{x\}) \doteq 0 \quad \forall x \in \mathbb{R} \setminus \mathbb{N}$.

$\left(\mu(\{k\}) = \begin{cases} 1 & x \in \mathbb{N} \\ 0 & \text{altrimenti} \end{cases} \right)$

$\left[\boxed{\text{NOTA}} \quad \sum_{k=0}^{\infty} \dfrac{\vartheta^k}{k!} = 1 + \vartheta + \dfrac{\vartheta^2}{2} + \dfrac{\vartheta^3}{3!} + \dots + \dfrac{\vartheta^n}{n!} + \dots = e^\vartheta \ ! \right]$

Consideriamo, appunto, ~~un~~ il mod. stat. param. dominato $(\Omega, \mathcal{F}, \{P^\vartheta \mid \vartheta \in \Theta\})$ canonico nel quale esiste un campione $(X_1, \dots, X_n)$ di taglia n e di legge Π_ϑ. In particolare, $\forall \vartheta \in \Theta$ e $\forall i = 1, \dots, n$, è

$\mathbb{E}^{P^\vartheta}[X_i] = \mathrm{Var}^{P^\vartheta}[X_i] = \vartheta$, e ~~rispetto~~ sfruttando

▶ $X_1, \dots, X_n \sim \Pi_\vartheta$ INDIP. $\Rightarrow$ $X_1 + \dots + X_n = S_n \sim \Pi_{n\vartheta}$

(somma di v.a.r. indip. di legge Poisson è v.a.r. Poisson di parametro la somma dei parametri).

Vediamo "a mano" che la statistica $T \doteq S_n = X_1 + \dots + X_n$ è esaustiva (sfruttando pure che $T \sim \Pi_{n\vartheta}$, appunto), verificando precisamente che $\forall x_1, \dots, x_n, t \in \mathbb{R}$ (ovvero $x_i, t \in \mathbb{N}$) risulti

$$P^\vartheta[X_1 = x_1, \dots, X_n = x_n \mid T = t] \text{ sia INDIP. da } \vartheta \in \Theta.$$

Vediamolo: $P^\theta[X_1=x_1,\dots,X_n=x_n \mid T=t] = \frac{P^\theta[X_1=x_1,\dots,X_n=x_n, T=t]}{P^\theta[T=t]} =$ (10)

$$= \begin{cases} 0 & \text{se } t \neq x_1+\dots+x_n \quad (\text{perché } T=X_1+\dots+X_n) \ \underline{\to \text{OK! (indip. } \theta)} \\ \dfrac{P^\theta[X_1=x_1,\dots,X_n=x_n]}{P^\theta[T=t]} & \text{se invece } t=x_1+\dots+x_n \end{cases}$$

Non resta che calcolare, $\forall x_1,\dots,x_n \in \mathbb{N}$, $\frac{P^\theta[X_1=x_1,\dots,X_n=x_n]}{P^\theta[T=x_1+\dots+x_n]}$:

questa frazione è $\overset{(\text{INDIP.})}{=} \dfrac{\prod_{i=1}^n \left(e^{-\theta}\frac{\theta^{x_i}}{x_i!}\right)}{e^{-n\theta}\frac{(n\theta)^{x_1+\dots+x_n}}{(x_1+\dots+x_n)!}} \overset{(\text{IMM.})}{=} \dfrac{(x_1+\dots+x_n)!}{x_1!\cdots x_n!\cdot n^{x_1+\dots+x_n}}$, che

effettivamente NON dipende da θ !! ✓ Dalla def. e dall'esempio, vediamo chiaramente come possa esser complicato il riconoscere operativamente una statistica sufficiente ... Tuttavia, esiste un'elegante caratterizzazione (un teorema) che lega l'sufficienza di una certa statistica con un'opportuna fattorizzazione della verosimiglianza del modello (dominato), in realtà non troppo sorprendente: una fattorizzazione, appunto, tale per cui la verosimiglianza sia ~~[illegible]~~ dipendente da θ solo "attraverso" la statistica in questione.

Prima dell'enunciato preciso, il lemma tecnico chiave (comunque di per sé utile in generale).

$\boxed{\text{Lemma (FORMULA DI BAYES PER LA SPERANZA CONDIZIONALE)}}$ Sia $(\Omega,\mathcal{F})$ mis., e siano Q_1, Q_2 due probabilità su $\mathcal{F}$ nella relazione di ass. cont. $Q_1 \ll Q_2$. Sia quindi $Z := \frac{dQ_1}{dQ_2}$ (grazie a Radon-Nikodym), sia X una v.a.r. in $\mathcal{L}^1(Q_1)$ (e, cioè, tale che $XZ \in \mathcal{L}^1(Q_2)$ con $E^{Q_1}[X] = E^{Q_2}[XZ]$) e sia $\mathcal{B} \subseteq \mathcal{F}$ una sotto-σ-algebra. Allora, vale che:

→ (1) $\mathbb{E}^{Q_2}[Z|\mathcal{B}] \neq 0$ Q_1-q.c., e

→ (2) $$\boxed{\mathbb{E}^{Q_1}[X|\mathcal{B}] = \frac{\mathbb{E}^{Q_2}[XZ|\mathcal{B}]}{\mathbb{E}^{Q_2}[Z|\mathcal{B}]} \quad Q_1\text{-q.c.}}$$

[Dim. (EX) (1) Vediamo che $Q_1[\,\mathbb{E}^{Q_2}[Z|\mathcal{B}] = 0\,] = 0$, ossia che $Q_1[A] = 0$ posto $A := \{\underbrace{\mathbb{E}^{Q_2}[Z|\mathcal{B}]}_{(\mathcal{B}-\text{mis.})} = 0\}$; notiamo subito che $A \in \mathcal{B}$. Ma, infatti, $Q_1[A] = \mathbb{E}^{Q_1}[\mathbb{1}_A] = \mathbb{E}^{Q_2}[\mathbb{1}_A \cdot Z] =$
$= \mathbb{E}^{Q_2}\left[\mathbb{1}_A\, \mathbb{E}^{Q_2}[Z|\mathcal{B}]\right] = 0$ per def. di A.

(def. speranza cond., dato che $A \in \mathcal{B}$)

(2) Poniamo $U := \dfrac{\mathbb{E}^{Q_2}[XZ|\mathcal{B}]}{\mathbb{E}^{Q_2}[Z|\mathcal{B}]}$ Q_1-q.c., e dimostriamo che, $\forall B \in \mathcal{B}$, $\mathbb{E}^{Q_1}[\mathbb{1}_B U] = \mathbb{E}^{Q_1}[\mathbb{1}_B X]$. Calcoliamo:

$$\mathbb{E}^{Q_1}[\mathbb{1}_B U] = \mathbb{E}^{Q_2}[\mathbb{1}_B U Z] =$$

(sfruttiamo che $\mathbb{1}_B \cdot U$ è $\mathcal{B}$-misurabile)

$$= \mathbb{E}^{Q_2}\left[\,\mathbb{E}^{Q_2}[\mathbb{1}_B U Z|\mathcal{B}]\,\right] =$$

$$= \mathbb{E}^{Q_2}\left[\mathbb{1}_B U \cdot \mathbb{E}^{Q_2}[Z|\mathcal{B}]\right] = \quad \text{(def. di } U\text{!)}$$

$$= \mathbb{E}^{Q_2}[\mathbb{1}_B\, \mathbb{E}^{Q_2}[XZ|\mathcal{B}]] = \mathbb{E}^{Q_2}[\mathbb{1}_B XZ] =$$

(analogo, e rituoso)

$$= \mathbb{E}^{Q_1}[\mathbb{1}_B X] \quad \text{(sempre e rituoso)} \quad . \; \square]$$

Presentiamo finalmente il teorema fondamentale sui risarcimenti aleatori, dimostrandolo ed applicandolo.

Teorema (di fattorizzazione di NEYMANN-FISHER). Sia $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ un mod. stat. (param.) dominato da una misura σ-finita (⇔ probabilità) μ su $\mathcal{F}$, e sia $P_0 = \sum_{n=1}^{\infty} c_n P^{\theta_n}$ una probabilità dominante privilegiata ($\leftarrow c_n \geq 0$, $\sum_{n=1}^{\infty} c_n = 1$, $(\theta_n)_{n\geq 1}$ in Θ). Data una statistica $T: (\Omega, \mathcal{F}) \to (E, \mathcal{E})$, sono $\boxed{\text{equivalenti}}$ le tre seguenti condizioni:

(1) T è un riassunto esaustivo.

(2) Esistono $h(\omega): (\Omega, \mathcal{F}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ mis. e, $\forall \theta \in \Theta$, $g^\theta: (E, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ mis. tali che sussista la fattorizzazione

$$\boxed{L_\mu(\theta, \omega) := \frac{dP^\theta}{d\mu}(\omega) = g^\theta(T(\omega))\, h(\omega)} \quad \forall (\theta, \omega) \in \Theta \times \Omega .$$

(3) Esistono, $\forall \theta \in \Theta$, $g^\theta: (E, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ mis. tali che

$$\boxed{L_{P_0}(\theta, \omega) \doteq \frac{dP^\theta}{dP_0}(\omega) = g^\theta(T(\omega))} \quad \forall (\theta, \omega) \in \Theta \times \Omega .$$

[*Dim.* Notiamo subito che $\boxed{(3) \Rightarrow (2)}$ è facilmente immediata, in quanto $P_0 \ll \mu$ e allora esiste $\frac{dP_0}{d\mu}$ ed è tale che, $\forall (\theta, \omega) \in \Theta \times \Omega$,

$$\frac{dP^\theta}{d\mu}(\omega) = \frac{dP^\theta}{dP_0}(\omega) \cdot \frac{dP_0}{d\mu}(\omega)$$

("transitività di $\ll$"): di conseguenza, posto $h(\omega) := \frac{dP_0}{d\mu}(\omega)$, se $\frac{dP^\theta}{dP_0}(\omega) = g^\theta(T(\omega))$ (ipotesi) allora $\frac{dP^\theta}{d\mu}(\omega) = g^\theta(T(\omega)) \cdot h(\omega)$. ✓

Allora proviamo ad ottenere $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1)$.

$\boxed{(1) \Rightarrow (3)}$ Poniamo, $\forall \theta \in \Theta$ e $\forall \omega \in \Omega$, $L^\theta(\omega) = \frac{dP^\theta}{dP_0}(\omega)$, e quindi $L^\theta(\omega)$ è la v.a.r. in $\mathcal{L}^1(P_0)$ tale che, per ogni v.a.r. X limitata, risulti $\quad \mathbb{E}^{P^\theta}[X] = \mathbb{E}^{P_0}[X \cdot L^\theta]$ per ogni $\theta \in \Theta$. Vogliamo verificare che, se T è sufficiente e se "pongo" $\mathbb{E}^{P_0}[L^\theta | T] = g^\theta(T)$ per ogni $\theta \in \Theta$ (grazie a Doob), allora accade che, $\forall$ v.a. X limitata,

$$\boxed{\mathbb{E}^{P^\theta}[X] = \mathbb{E}^{P_0}[X\, g^\theta(T)]}.$$

(illusione solo per la fattorizzazione!)

$\rightarrow$ Dapprima, vediamo che, $\forall Y$ v.a. limitata, e $\forall \theta \in \Theta$,

$$\mathbb{E}^{\cdot}[Y | T] \underset{(T \text{ sufficiente})}{=} \mathbb{E}^{P^\theta}[Y|T] = \mathbb{E}^{P_0}[Y|T] \quad : \quad \text{ma, infatti,}$$

per ogni $B \in \sigma(T)$, $\quad \mathbb{E}^{P_0}[\mathbb{1}_B Y] \overset{(\text{def. } P_0)}{=} \sum_{n=1}^{\infty} c_n \mathbb{E}^{P^{\theta_n}}[\mathbb{1}_B Y] \underset{(B \in \sigma(T))}{=}$

$$= \sum_{n=1}^{\infty} c_n \mathbb{E}^{P^{\theta_n}}[\mathbb{1}_B \mathbb{E}^{P^{\theta_n}}[Y|T]] \underset{(\text{sufficiente})}{=} \sum_{n=1}^{\infty} c_n \mathbb{E}^{P^{\theta_n}}[\mathbb{1}_B \mathbb{E}^{\cdot}[Y|T]] =$$

$$= \mathbb{E}^{P_0}[\mathbb{1}_B \mathbb{E}^{\cdot}[Y|T]] \quad . \quad \checkmark$$

$\rightarrow$ Allora: $\quad \mathbb{E}^{P^\theta}[X] = \mathbb{E}^{P^\theta}[\mathbb{E}^{P^\theta}[X|T]] = \quad (T \text{ sufficiente})$

$$= \mathbb{E}^{P^\theta}[\mathbb{E}^{\cdot}[X|T]] \quad , \quad = \quad (P^\theta = L^\theta \cdot P_0)$$

$$= \mathbb{E}^{P_0}[L^\theta\, \mathbb{E}^{\cdot}[X|T]] \,, =$$

(possiamo fare $\mathbb{E}^{\cdot}[X|T] = \mathbb{E}^{P_0}[X|T]$, MA mancherebbe un altro condizionamento rispetto a $\sigma(T)$ per far "uscire" g^θ ...!)

$$= \mathbb{E}^{P_0}[\mathbb{E}^{P_0}[L^\theta\, \underset{(P_0!)}{\mathbb{E}^{\cdot}}[X|T] \,|\, T]] \,, = \quad \left(\begin{array}{c}\mathbb{E}^{\cdot}[X|T] = \\ = \mathbb{E}^{P_0}[X|T] \\ (T\text{-misurabile})\end{array}\right)$$

$$= \mathbb{E}^{P_0}[\mathbb{E}^{P_0}[X|T] \cdot \mathbb{E}^{P_0}[L^\theta | T]] \overset{(\text{def. } g^\theta)}{=} \mathbb{E}^{P_0}[g^\theta(T)\, \mathbb{E}^{P_0}[X|T]] =$$

$$= \mathbb{E}^{P_0}[\mathbb{E}^{P_0}[X g^\theta(T) | T]] = \mathbb{E}^{P_0}[X g^\theta(T)] \quad . \quad \checkmark$$

$\boxed{2) \Rightarrow 1)}$ Vogliamo ottenere che, $\forall \vartheta \in \Theta$, e $\forall$ v.a.r. X limitata, esiste μ-q.o. $\mathbb{E}^{P^\vartheta}[X|T]$ e che NON dipende in realtà da $\vartheta \in \Theta$. Per questo, possiamo supporre che μ sia una probabilità dominante, e sappiamo che (perché è equiv. ad una prob. dominante)

$$\frac{dP^\vartheta}{d\mu}(\omega) = g^\vartheta(T(\omega)) \cdot h(\omega) \qquad (\text{a meno di modificare } h(\omega)).$$

Ponendo $L^\vartheta(\omega) \doteq \frac{dP^\vartheta}{d\mu}(\omega)$ $\forall (\vartheta, \omega) \in \Theta \times \Omega$, la formula di Bayes per la speranza condizionale (LEMMA) garantisce che $\mathbb{E}^\mu[L^\vartheta | T] \neq 0$ (P^ϑ-q.c.) e che

$$\mathbb{E}^{P^\vartheta}[X|T] = \frac{\mathbb{E}^\mu[X L^\vartheta | T]}{\mathbb{E}^\mu[L^\vartheta | T]} \quad (P^\vartheta\text{-q.c.}) \quad (\forall \vartheta \in \Theta).$$

Ma $L^\vartheta(\cdot) = g^\vartheta(T(\cdot)) h(\cdot)$, e $g^\vartheta(T(\cdot))$ è T-misurabile, per cui avremo subito

$$\mathbb{E}^{P^\vartheta}[X|T] = \frac{\mathbb{E}^\mu[X h | T]}{\mathbb{E}^\mu[h|T]}, \text{ indip. da } \vartheta \in \Theta. \quad \square$$

? Consideriamo un mod. stat. param. $(\Omega, \mathcal{F}, \{P^\vartheta | \vartheta \in \Theta\})$ dominato da μ, e supponiamolo "riassumibile", nel senso che esiste una statistica (esaustiva, per il TFO.) $T: (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ tale che $\frac{dP^\vartheta}{d\mu}(\omega) = g^\vartheta(T(\omega)) h(\omega)$ $\forall (\vartheta, \omega) \in \Theta \times \Omega$ per opportune $g^\vartheta: (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e $h: (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabili ($\Rightarrow h$ è statistica!). La domanda è: esisterà qualche legame fra la statistica esaustiva T e certi stimatori "buoni" (di una prefissata funzione di ϑ)? Sì ...! Ma, prima, esempi.

$\boxed{\text{Oss.}}$ (Corollario.) Sia $(\Omega, \mathcal{F}, \{P^\vartheta | \vartheta \in \Theta\})$ un mod. stat. param. dominato da una misura σ-finita, o probabilità, μ su ~~[illegible]~~ $\mathcal{F}$, e sia $T: (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ una statistica esaustiva per tale modello: equivalentemente, per il teorema di fattorizzazione di Neyman–Fisher, esistono $h(\omega): (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mis. e, $\forall \vartheta \in \Theta$, $g^\vartheta: (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mis. tale che, $\forall (\vartheta, \omega) \in \Theta \times \Omega$, sia

$L_\mu(\vartheta,\omega) := \frac{dP^\vartheta}{d\mu}(\omega) = h(\omega)\, g^\vartheta(T(\omega))$. Ovviamente: funzione biiettiva di una statistica esaustiva T resta una statistica esaustiva. In altri termini, se $\varphi : (E,\mathcal{E}) \to (G,\mathcal{G})$ è mis. e biiettiva (fra spazi mis.), allora $S := \varphi(T) : (\Omega,\mathcal{A}) \xrightarrow{\varphi^{-1}} (G,\mathcal{G})$ è una statistica esaustiva per il modello. [Ciò è immediato, semplicemente perché, $\forall(\vartheta,\omega)\in\Theta\times\Omega$, è
$$L(\vartheta,\omega) := \frac{dP^\vartheta}{d\mu}(\omega) = h(\omega)\, g^\vartheta(T(\omega)) \underset{(S=\varphi(T))}{=} h(\omega)\,(g^\vartheta\circ\varphi^{-1})(S(\omega)).$$
Dunque otteniamo usando il teorema di Neyman-Fisher. ✓] ✓

(ES.) Consideriamo il mod. stat. (dunque dominato) $(\Omega,\mathcal{A},\{P^\vartheta|\vartheta\in\Theta\})$ canonico, dove $\Theta=(0,1)$, nel quale esiste un campione $(X_1,\dots,X_n)$ di taglia $n\in\mathbb{N}_*$ pre-fissata e di legge $B(1,\vartheta)$. ~~[illegible]~~ (Più esplicitamente, è $\Omega=\{0,1\}^n$, $\mathcal{A}=\mathcal{P}(\Omega)$, ($\Theta=(0,1)$,) $P^\vartheta[\{\omega\}] = P^\vartheta(\{(\underbrace{\omega_1,\dots,\omega_n}_{(\in\{0,1\})})\}) := \vartheta^{\sum_{i=1}^n \omega_i}(1-\vartheta)^{n-\sum_{i=1}^n\omega_i}$ ($\mu(\{\omega\})$, μ contatrice),

$X_i(\omega_1,\dots,\omega_n) = \omega_i \quad \forall i=1,\dots,n$.)

Allora, la verosimiglianza di questo modello (rispetto alla misura che conta i punti su $\mathcal{A}$) è evidentemente ($\forall\vartheta\in(0,1)$ e $\forall(\omega_1,\dots,\omega_n)\in\{0,1\}^n$)
$$L(\vartheta;\omega_1,\dots,\omega_n) = \vartheta^{\sum_{i=1}^n\omega_i}(1-\vartheta)^{n-\sum_{i=1}^n\omega_i}$$
: pertanto, la statistica $T := S_n = X_1+\dots+X_n$ (e, cioè, $T(\omega_1,\dots,\omega_n)=\sum_{i=1}^n\omega_i$) è esaustiva, in quanto effetto $L(\vartheta;\omega_1,\dots,\omega_n)$ è del tipo $g^\vartheta(T(\omega_1,\dots,\omega_n))$ (ed usiamo il teorema di Neyman-Fisher). ✓

(ES.) Consideriamo il mod. dominato $(\Omega,\mathcal{A},\{P^\vartheta|\vartheta\in\Theta\})$ canonico, dove $\Theta=(0,\infty)$, nel quale esiste un campione $(X_1,\dots,X_n)$ di *n*-upla n ($\in \mathbb{N}_*$ finita) e di legge Poisson Π_θ. In pratica, Ω coincide con "$\mathbb{N}^n$" (σ, ordinando, $\mathbb{R}^n$), la misura dominante μ è quella che conta i punti di $\mathbb{N}^n$, e per ogni $\theta \in \Theta$ è ($\forall k_1, \dots, k_n \in \mathbb{N}$)

$$P^\theta[\{(k_1, \dots, k_n)\}] = e^{-n\theta} \frac{\theta^{\sum_{i=1}^n k_i}}{\prod_{i=1}^n k_i!}$$

Inoltre, $\forall i = 1, \dots, n$, $X_i(k_1, \dots, k_n) = k_i$. Pertanto, usando questo e Neymann-Fischer, possiamo subito stabilire che $T := S_n = X_1 + \dots + X_n$ (ossia, $T(k_1, \dots, k_n) = \sum_{i=1}^n k_i$) è una statistica esaustiva per il modello. $\int$

ES. Consideriamo il mod. dominato $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$ canonico, ove $\Theta = (-1, \infty)$, nel quale esiste un campione $(X_1, \dots, X_n)$ di taglia $n \in \mathbb{N}_*$ (pre-fissata) e di legge μ^θ, $\theta \in \Theta$, su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ date da

$$\mu^\theta = f^\theta \cdot \underbrace{\lambda}_{\substack{\text{(mis. di Leb.}\\ \text{su } \mathcal{B}(\mathbb{R}))}}, \quad f^\theta(x) := (\theta+1) x^\theta \, 1\!\!1_{(0,1)}(x) \quad \text{per ogni } x \in \mathbb{R}.$$
($\Rightarrow \mu^\theta \ll$ Lebesgue)

$\forall \theta > -1$, e $\forall x \in \mathbb{R}$, $f^\theta(x) = \begin{cases} 0 & \text{se } x \leq 0 \text{ o } x \geq 1 \\ (\theta+1)x^\theta & \text{se } x \in (0,1) \end{cases}$, ed in particolare $f^\theta(x) \geq 0$.

Inoltre, $\int_{-\infty}^{+\infty} f^\theta(x)\,dx = \int_0^1 (\theta+1)x^\theta\,dx = \left[x^{\theta+1}\right]_{x=0}^{x=1} = 1$: in conclusione, μ^θ è ovviamente una legge di probabilità su $\mathcal{B}(\mathbb{R})$. $\int$

In pratica, $\Omega = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, μ dominante di Lebesgue (su $\mathcal{B}(\mathbb{R}^n)$),

$$\underset{(\forall x_i \in \mathbb{R})}{dP^\theta[\{(x_1, \dots, x_n)\}]} = \left[(\theta+1)^n \left(\prod_{i=1}^n x_i\right)^\theta 1\!\!1_{(0,1)^n}(x_1, \dots, x_n)\right] \cdot d\mu(\{(x_1, \dots, x_n)\}),$$

o insomma $L(\theta; x_1, \dots, x_n) = (\theta+1)^n \left(\prod_{i=1}^n x_i\right)^\theta 1\!\!1_{(0,1)^n}((x_1, \dots, x_n))$, e al solito $X_i((x_1, \dots, x_n)) = x_i$: grazie a Neymann-Fischer, segue subito che $\boxed{T = \prod_{i=1}^n X_i}$ è una statistica esaustiva per il modello. $\int$

(ES.) Consideriamo il mod. regolare canonico $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$, dove $\Theta = \mathbb{R}_m \times (0,\infty)_{\sigma^2}$, sul quale esiste un campione $(X_1, \dots, X_n)$ di taglia $n \in \mathbb{N}_*$ scelta e di legge $N(m, \sigma^2)$, con $(m, \sigma^2) \in \Theta$.

Di fatto, $\Omega = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, e $\underbrace{P^\theta}_{(P^{(m,\sigma^2)})}$ ha densità rispetto alla Lebesgue data da $(\forall x_i \in \mathbb{R})$ (e $X_i((x_1,\dots,x_n)) = x_i$)

$$L(m, \sigma^2 ; x_1, \dots, x_n) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i - m)^2}{\sigma^2}\right)\right) =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right).$$

Ma $\sum_{i=1}^{n}(x_i - m)^2 = \sum_{i=1}^{n}(x_i^2 - 2mx_i + m^2) = \sum_{i=1}^{n} x_i^2 - 2m\sum_{i=1}^{n} x_i + nm^2$,

e allora $L(m, \sigma^2 ; x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2m\sum_{i=1}^{n} x_i + nm^2\right) - n\log\sigma\right)$: segue, dunque, che la statistica

$T := \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ (a valori in $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$) è sufficiente per il modello! (O anche che $\sum_{i=1}^{n} X_i$ e $\sum_{i=1}^{n} X_i^2$ sono congiuntamente sufficienti.) ($\boxed{\text{L'IDEA}}$, in generale, è che per N parametri incogniti da "stimare"/"riassumere" occorrono N statistiche!)

(ES.) Consideriamo il mod. dominato $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$ canonico sul quale esiste un campione $(X_1, \dots, X_n)$ di legge assolutamente continua rispetto alla misura di Lebesgue (su $\mathcal{B}(\mathbb{R})$) avente densità rispettiva uniforme sull'intervallo (θ_1, θ_2), dove $\underbrace{(\theta_1, \theta_2) \in \Theta}_{\text{(come coppia)}} = \{(x,y) \in \mathbb{R}^2 \mid y > x\}$, e cioè $A^{(\vartheta_1,\vartheta_2)}(x) = \frac{1}{\vartheta_2-\vartheta_1} \mathbb{1}_{(\vartheta_1,\vartheta_2)}(x)$, $x \in \mathbb{R}$. Dunque in pratica, $\Omega = \mathbb{R}^n$, $\mathcal{Y} = \mathcal{B}(\mathbb{R}^n)$, la misura dominante μ su $\mathcal{Y}$ è quella di Lebesgue n-dim., $X_i((x_1,\dots,x_n)) = x_i$ $(\forall i=1,\dots,n$ e $\forall x_i \in \mathbb{R})$, e la verosimiglianza è

$$L(\vartheta_1,\vartheta_2\,;\,x_1,\dots,x_n) = (\vartheta_2-\vartheta_1)^{-n}\, \mathbb{1}_{(\vartheta_1,\vartheta_2)^n}((x_1,\dots,x_n)) \ .$$

($\triangledown$! Anche il supporto delle densità dipende dai parametri ... !)

Ma , chiaramente , $\mathbb{1}_{(\vartheta_1,\vartheta_2)^n}((x_1,\dots,x_n)) = \mathbb{1}_{(\vartheta_1,\vartheta_2)^2}(x_1\wedge\dots\wedge x_n,\, x_1\vee\dots\vee x_n)$, dove (come sempre) $x_1\wedge\dots\wedge x_n = \min\{x_1,\dots,x_n\}$ e $x_1\vee\dots\vee x_n = \max\{x_1,\dots,x_n\}$.

Segue pertanto che le due statistiche $\min_{i=1,\dots,n} X_i$ e $\max_{i=1,\dots,n} X_i$ risultano congiuntamente sufficienti nel riassumere il modello in questione . ✓

Approfondiremo la teoria delle statistiche esaustive su di un modello stat. dominato ~~[illegible]~~ a riguardo di una classe particolarmente speciale di mod. stat. dominati , i quali per la loro stessa definizione posseggono una statistica esaustiva (Neyman-Fisher) : i modelli stat. (dominati) "esponenziali" , cosiddetti .

Dapprima dovremo richiamare/dare per buono il concetto di trasformata di Laplace di una misura σ-finita su $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, e quindi le sue proprietà basilari più importanti .

Tuttavia , vogliamo prima studiare il rapporto tra statistiche esaustive e statistiche stimatori "buoni" , provando pure ad incrementare le qualità di una stat. esaustiva . ($\to$ STATISTICA COMPLETA)

($\boxed{\text{Promemoria}}$ (sulla speranza condizionale) Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità, e sia $\mathcal{B} \subseteq \mathcal{F}$ una sotto-σ-algebra. Dunque, evidentemente, vale $\mathcal{L}^2(\Omega, \mathcal{B}, P) < \underset{(\mathbb{R}\text{-sp. di Hilbert})}{\mathcal{L}^2(\Omega, \mathcal{F}, P)}$. Ebbene, vale che

→ il sottosp. $\mathcal{L}^2(\Omega, \mathcal{B}, P)$ è CHIUSO in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, e

→ $\forall X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, la proiezione ortogonale di X sul sottosp. chiuso $\mathcal{L}^2(\Omega, \mathcal{B}, P)$ coincide con $\mathbb{E}^P[X|\mathcal{B}]$.
(in $\mathcal{L}^2$, per speranza condizionale)

Corollario (immediato). Nella precedente situazione, vale che

$$\| \mathbb{E}^P[X | \mathcal{B}] \|_{\mathcal{L}^2(P)} \leq \| X \|_{\mathcal{L}^2(P)}, \text{ e vale } = \text{ se,}$$

e solo se, X è in realtà $\mathcal{B}$-misurabile)

$\boxed{\text{Teorema (di BLACKWELL-RAO).}}$ Sia $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ un mod. stat. dominato, e supponiamo che esista una statistica sufficiente $T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ per tale modello. Consideriamo nel problema di stima per una funzione $g(\theta) : \Theta \to D$, $D \subseteq \mathbb{R}$ aperto $\neq \emptyset$ fissato, considerando come candidati stimatori ammessi tutte le v.a.r. $\Omega \to D$ che sono in $\mathcal{L}^2(P^\theta)$ per ogni $\theta \in \Theta$. Allora, gli stimatori corretti di g in $\mathcal{L}^2(P^\theta)\ \forall \theta \in \Theta$ risultano "~~preferibili~~" quelli ottenuti condizionando rispetto a $\sigma(T)$.
($\Rightarrow$ della forma $\phi(T)$)
"Rao-Blackwell", più precisamente:

(1) se $U : \Omega \to D$ è v.a.r. in $\mathcal{L}^2(P^\theta)$, per ogni $\theta \in \Theta$, allora così è per $\boxed{V \doteq \underset{\substack{\uparrow \\ (T \text{ sufficiente!})}}{\mathbb{E}^{\bullet}}[U | T]}$ ed in più V è preferibile ad U

(nel senso che, $\forall \theta \in \Theta$, $\mathbb{E}^{P^\theta}[|V - g(\theta)|^2] \leq \mathbb{E}^{P^\theta}[|U - g(\theta)|^2]$), ed inoltre V è strettamente preferibile ad U se, e solo se, U NON è $\sigma(T)$-misurabile ;

(2) inoltre, se U è corretto (nel senso che, $\forall \vartheta \in \Theta$, $\mathbb{E}^{P^\vartheta}[U] = g(\vartheta)$), allora anche V è corretto.

Dim. (1) Fissiamo $\vartheta \in \Theta$, e consideriamo $\overbrace{U}^{U-g(\vartheta)} \in \mathcal{L}^2(\Omega, \mathcal{F}, P^\vartheta)$ ($g(\vartheta) \in \mathbb{R}$!). Abbiamo che $\mathcal{B} = \sigma(T)$ è una sotto-σ-algebra di $\mathcal{F}$, e quindi (come richiamato poco fa) si ha che $\| \mathbb{E}^{P^\vartheta}[\overbrace{U}^{U-g(\vartheta)} | T] \|_{\mathcal{L}^2(P^\vartheta)} \leq \| U - g(\vartheta) \|_{\mathcal{L}^2(P^\vartheta)}$, col $<$ se, e solo se, U non è $\sigma(T)$-misurabile. Ma, adesso,

$\| U - g(\vartheta) \|_{\mathcal{L}^2(P^\vartheta)} = \mathbb{E}^{P^\vartheta}[|U - g(\vartheta)|^2]$, ed anche

$\mathbb{E}^{P^\vartheta}[U - g(\vartheta) | T] \underset{(g(\vartheta) \text{ è costante})}{=} \mathbb{E}^{P^\vartheta}[U | T] - g(\vartheta) \overset{(\text{def. } V)}{=} V - g(\vartheta)$ che

$\| \cdot \|_{\mathcal{L}^2(P^\vartheta)} = \mathbb{E}^{P^\vartheta}[|V - g(\vartheta)|^2]$. ✓

(2) ~~Basta~~ Osserviamo infine che, $\forall \vartheta \in \Theta$, $\mathbb{E}^{P^\vartheta}[U] = \mathbb{E}^{P^\vartheta}[V]$, in quanto $V = \mathbb{E}^{P^\vartheta}[U | T]$. □

Dunque, naturalmente, se T fosse „più speciale", allora potremmo fare molto di più che confrontare $\mathbb{E}^{\cdot}[U | T]$ solo con U stesso (in $\mathcal{L}^2(P^\vartheta)$ $\forall \vartheta \in \Theta$ e corretto) - ...

Introduciamo una statistica $T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ come COMPLETA se, quale che sia Y v.a.r. in $\mathcal{L}^1(P^\vartheta)$ per ogni $\vartheta \in \Theta$ con $\mathbb{E}^{P^\vartheta}[Y] = 0$ per ogni $\vartheta \in \Theta$, se Y è T-misurabile (ossia $Y = \varphi(T)$), allora $Y = 0$ μ-q.o. (se μ è dominante).

$\boxed{\text{Corollario (Blackwell-Rao).}}$ Sia $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ un mod. stat. dominato da una misura μ su $\mathcal{F}$ (σ-finita, o probabilità), e supponiamo che tale modello ammetta una statistica esaustiva e completa $T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$. Occupiamoci del problema di stima per una funzione $g(\theta)$, cercando stimatori in $\mathcal{L}^2(P^\theta)$ per ogni $\theta \in \Theta$. Ebbene, tra tutti gli stimatori corretti di g in $\mathcal{L}^2(P^\theta)$ $\forall \theta \in \Theta$, risultano OTTIMALI quelli ottenuti condizionando rispetto a T, nel senso che: scelto uno qualsiasi $U \in \mathcal{L}^2(P^\theta)\ \forall \theta \in \Theta)$ che sia stimatore corretto di g, $V := \mathbb{E}^\cdot[U \mid T]$ $(\in \mathcal{L}^2(P^\theta)\ \forall \theta \in \Theta)$ è uno stimatore corretto di g preferibile non soltanto ad U, ma preferibile rispetto ad ogni altro stimatore corretto di g che sia in $\mathcal{L}^2(P^\theta)$ $\forall \theta \in \Theta$.

[Dim. Fissiamo una v.a. U in $\mathcal{L}^2(P^\theta)$ $\forall \theta \in \Theta$ tale che $\mathbb{E}^{P^\theta}[U] = g(\theta)$ $\forall \theta \in \Theta$. Allora, in virtù del teorema di Blackwell-Rao, possiamo affermare che anche $V := \mathbb{E}^\cdot[U \mid T]$ sta in $\mathcal{L}^2(P^\theta)$ $\forall \theta \in \Theta$ e che $\mathbb{E}^{P^\theta}[V] = \mathbb{E}^{P^\theta}[U] = g(\theta)$ $\forall \theta \in \Theta$ e che V è preferibile ad U, cioè $\underbrace{\mathbb{E}^{P^\theta}[|V - g(\theta)|^2]}_{(R_V(\theta))} \leq \underbrace{\mathbb{E}^{P^\theta}[|U - g(\theta)|^2]}_{(R_U(\theta))}$ $\forall \theta \in \Theta$.

La tesi è che T completa (oltre che esaustiva) $\Rightarrow$ $\forall \tilde{U} \in \bigcap_{\theta \in \Theta} \mathcal{L}^2(\theta)$ che abbia $\mathbb{E}^{P^\theta}[\tilde{U}] = g(\theta)$ $\forall \theta \in \Theta$, resta $R_V(\theta) \leq R_{\tilde{U}}(\theta)$ $\forall \theta \in \Theta$.

▶ Claim: $\mathbb{E}^\cdot[\tilde{U} \mid T] = \mathbb{E}^\cdot[U \mid T]$ (da cui la tesi).

Questo, ora, segue subito dalla completezza della T, perché la v.a.r. $E^{\cdot}[U|T]-E^{\cdot}[\tilde U|T] \overset{(lin.)}{=} E^{\cdot}[U-\tilde U|T]$ sta in $L^1(P^\theta)$ per ogni $\theta\in\Theta$ ed è T-misurabile ; per ciò, quindi, basta (ed occorre) dimostrare che $E^{P^\theta}[E^{\cdot}[U-\tilde U|T]]=0\ \forall\theta\in\Theta$.

Ma questo è evidente, perché $E^{P^\theta}[U]=E^{P^\theta}[\tilde U]=g(\theta)$ per ogni $\theta\in\Theta$, e allora $E^{P^\theta}[E^{P^\theta}[U-\tilde U|T]]=E^{P^\theta}[E^{P^\theta}[U|T]]-E^{P^\theta}[E^{P^\theta}[\tilde U|T]]=$
$=g(\theta)-g(\theta)=0$. □]

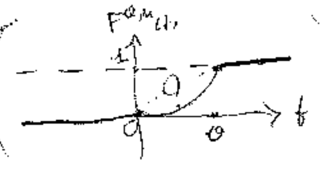
Di solito, se già non è facile trovare statistiche esaustive, e meglio ancora non è facile trovare statistiche che siano contemporaneamente sufficienti e complete. Anticipiamo tuttavia che, di solito, per i modelli stat. esponenziali sarà automatico l'avere esplicitamente una statistica esaustiva e completa.

ES. Consideriamo $\Theta=(0,\infty)$ ed il mod. stat. dominato canonico $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\})$ relativo ad un campione $(X_1,\dots,X_n)$ di legge uniforme $\mathcal{U}((0,\theta))$ su $(0,\theta)$: insomma, $\Omega=(\mathbb{R}_+)^n$, $\mathcal{A}=\mathcal{B}(\mathbb{R}_+^n)$, μ di Lebesgue (su $\mathcal{A}$), $L(\theta;x_1,\dots,x_n)=\theta^{-n}\mathbb{1}_{(0,\theta]}(x_1\vee\dots\vee x_n)$ $\forall\theta\in\Theta$ e $\forall x_1,\dots,x_n\in\mathbb{R}_+$, $X_i(x_1,\dots,x_n)=x_i$ $\forall\dots$.

Sappiamo già, e comunque è (almeno) evidente, che $T=\max_{i=1,\dots,n}X_i$ è una statistica esaustiva per il modello. Verifichiamo ora (e meno) che è pure completa : posto che $T:(\Omega,\mathcal{A})\to(\mathbb{R}_+,\mathcal{B}(\mathbb{R}_+))$, sia $h:(\mathbb{R}_+,\mathcal{B}(\mathbb{R}_+))\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mis. tale che $Y=h(T)$ sia in $\bigcap_{\theta\in\Theta\,(\theta>0)}L^1(P^\theta)$, e supponiamo precisamente che $E^{P^\theta}[Y]=0\ \forall\theta\in\Theta\ (\theta>0)$; vogliamo dedurre che $h\equiv 0$ (θ in $m(0,\theta)$ Lebesgue-q.o. $-$q.o. $\forall\theta>0$) Partiamo dunque da

$$0 = \mathbb{E}^{P^\theta}[h(T)] = \int_0^\infty h(t) f^{\theta,n}(t)\,dt \quad (\forall \theta > 0)$$, se $f^{\theta,n}(t)$ è la densità di $T(P^\theta)$ rispetto alla mis. di Lebesgue 1-dim.

$\boxed{\text{Calcoliamo } f^{\theta,n}(t)}$: indicando con $F^{\theta,n}(t)$ la funzione di ripartizione di T, otteniamo che $\underset{(\forall \theta\in\Theta, \forall n\in\mathbb{N})}{\forall t\in\mathbb{R}}$ $F^{\theta,n}(t) = P^\theta[T \le t] =$

$$= P^\theta[\max_{i=1,\dots,n} X_i \le t] \begin{cases} = 0 & \text{se } t \le 0 \\ = 1 & \text{se } t \ge \theta \end{cases}$$ (),

mentre, $\forall t \in (0,\theta)$, $F^{\theta,n}(t) \overset{(iid)}{=} P^\theta[X_1 \le t, \dots, X_n \le t] = (\text{INDIP.})$

$$= \prod_{i=1}^{n} P^\theta[X_i \le t] = \big(\underbrace{P^\theta[X_1 \le t]}_{\left(\frac{t}{\theta}\right)}\big)^n = \theta^{-n} t^n$$, e pertanto

$$f^{\theta,n}(t) = \begin{cases} n\theta^{-n} t^{n-1} & \text{per ogni } t \in (0,\theta) \\ 0 & \text{altrimenti} \end{cases}$$. ✓

Tornando quindi sopra, $0 = \mathbb{E}^{P^\theta}[h(T)] = \int_0^\infty h(t) f^{\theta,n}(t)\,dt =$

$$= \int_0^\theta n\theta^{-n} t^{n-1} h(t)\,dt = n\theta^{-n}\int_0^\theta h(t)\, t^{n-1}\,dt \quad \forall \theta > 0$$,

e cioè $\int_{\theta_1}^{\theta_2} h(t)\, t^{n-1}\,dt = 0$ per ogni $0 \le \theta_1 < \theta_2 (< \infty)$, da cui

$h(t)\, t^{n-1} \equiv 0$ q.o., e così $h(t) \equiv 0$ q.o. //

$\boxed{\text{NOTA}}$ A questo punto, avremmo uno stimatore corretto "ottimale" di una qualche $g(\theta)$, scegliendolo della forma $\Phi(T)$...

$\rightsquigarrow$ Osserviamo che, $\forall \theta \in \Theta$, $\mathbb{E}^{P^\theta}[T] =$ (VISTO!)

$$= \int_0^\vartheta t\,(n\vartheta^{-n})\,t^{n-1}\,dt = \vartheta^{-n}\frac{n}{n+1}\int_0^\vartheta (n+1)\,t^n\,dt =$$

(t^n)

$$= \vartheta^{-n}\frac{n}{n+1}\left[t^{n+1}\right]_{t=0}^{t=\vartheta} = \frac{n}{n+1}\vartheta$$, e cioè la statistica

$U := \frac{n+1}{n}\cdot T = \left(1+\frac{1}{n}\right)\max_{i=1,\dots,n} X_i$ che $\mathbb{E}^{P^\vartheta}[U] = \vartheta$ per ogni $\vartheta\in\Theta$:

U è per ciò uno stimatore corretto di ϑ (osservato anche che è a valori in $(0,\infty)$, come ϑ). Ma, adesso, T è chiaramente in $\mathcal{L}^\infty(P^\vartheta)$ per ogni $\vartheta\in\Theta$, (in quanto assume valori in $(0,\vartheta)$ P^ϑ-q.c.), e allora a maggior ragione sta in $\mathcal{L}^2(P^\vartheta)$ per ogni $\vartheta\in\Theta$, quindi anche U : possiamo concludere infine che $U = \left(\frac{n+1}{n}\right)T$ è uno stimatore corretto di ϑ che sta in $\bigcap_{\vartheta\in\Theta}\mathcal{L}^2(P^\vartheta)$ ed ottimale fra tutti gli stimatori corretti di ϑ che stanno in $\bigcap_{\vartheta\in\Theta}\mathcal{L}^2(P^\vartheta)$. √

Se di un modello stat. $(\Omega,\mathcal{F},\{P^\vartheta|\vartheta\in\Theta\})$, possibilmente dominato, esistono una moltitudine di statistiche più o meno interessanti rispetto a determinati scopi. Dal punto di vista del riassumere il modello, troviamo come "migliori" le statistiche esaustive (ancora meglio se complete). ? All'altro "estremo" cosa troviamo? Cioè: quali sono le statistiche che non danno alcuna informazione sul modello? E che legame hanno con quelle esaustive? ... Sia $S : (\Omega,\mathcal{F}) \to (E,\mathcal{E})$ una qualsiasi statistica nel nostro modello. Allora S è **LIBERA** (del modello) se le sue leggi $S(P^\vartheta)$ rispetto a P^ϑ al variare di ϑ in Θ (su $\mathcal{E}$) sono in realtà indipendenti da $\vartheta\in\Theta$, e cioè se $\{S(P^\theta)\}_{\theta\in\Theta}$ è in realtà costituito da una sola legge (su $\mathcal{E}$)

(Dunque, NON potrà essere per noi molto importante...!)

$\boxed{\text{NOTA}}$ Ad es., banalmente, una statistica costante è giustamente libera da ogni modello: infatti, se $S(\omega)\equiv x_0$ per un certo $x_0\in E$, allora, $\forall\theta\in\Theta$, $S(P^\theta)=\varepsilon_{x_0}$ (perché, $\forall A\in\mathcal{E}$, $S^{-1}(A)=\begin{cases}\Omega & \text{se } x_0\in A\\ \emptyset & \text{altrimenti}\end{cases}$). ✓ Vediamo invece come una statistica esaustiva non potrà essere anche libera, almeno in un caso particolare: prendiamo $T:(\Omega,\mathcal{A})\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ una statistica reale, e supponiamo che il nostro modello $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\})$ sia dominato; sia allora P_0 una probabilità dominante privilegiata, e sia $L^\theta(\omega)=\frac{dP^\theta}{dP_0}(\omega)$ per ogni $(\theta,\omega)\in\Theta\times\Omega$. Allora, per ogni $\theta\in\Theta$ e $A\in\mathcal{B}(\mathbb{R})$, abbiamo

$$T(P^\theta)[A] = P^\theta[T^{-1}(A)] = \int_\Omega \underbrace{1_{T^{-1}(A)}(\omega)}_{=(1_A(T(\omega)))}dP^\theta(\omega) =$$

$$= \int_{\mathbb{R}} 1_A(t)\, dT(P^\theta)(t)$$

; se T fosse esaustiva, allora per Neyman-Fischer sarebbe $L^\theta(\omega)=g^\theta(T(\omega))$ e allora

$$P^\theta[T^{-1}(A)] = \int_\Omega 1_A(T(\omega))\, g^\theta(T(\omega))\, dP_0(\omega) =$$

(teo. d'integr. rispetto ad una prob.-immagine)

(stessa T!!)

$$= \int_{\mathbb{R}} 1_A(t)\, g^\theta(t)\, dT(P_0)(t)$$

, da cui

$$T(P^\theta)[A] = \int_A g^\theta(t)\, dT(P_0)(t)$$

, chiaramente dipendente da $\theta\in\Theta$ (in generale). ✓

(ES.) Mettiamoci nel modello stat. $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\})$ regolare canonico nel quale esiste un campione $(X_1,\dots,X_n)$ di una certa taglia n ($\theta\in\Theta$)

e di legge $N(\vartheta,1)$: dunque $\Omega=\mathbb{R}^n$, $\mathcal{A}=\mathcal{B}(\mathbb{R}^n)$, $\Theta=\mathbb{R}$, mis. dominante quella di Lebesgue n-dimensionale, $X_i((x_1,\dots,x_n))=x_i$ per ogni $x_1,\dots,x_n\in\mathbb{R}$ e $i=1,\dots,n$, e, $\forall\vartheta\in\Theta$, P^ϑ ha densità rispetto a μ data da $\mathcal{L}^\vartheta(x_1,\dots,x_n)=(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^n(x_i-\vartheta)^2\right)$.

Dunque, se come sempre $\overline{X}:=\frac{S_n}{n}=\frac{1}{n}\sum_{i=1}^n X_i$, allora $\mathbb{E}^{P^\vartheta}[\overline{X}]=\vartheta$ $\forall\vartheta\in\Theta$, mentre se $S^2:=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})^2$ allora $\forall\vartheta\in\Theta$

$\boxed{\mathbb{E}^{P^\vartheta}[S^2]=1}$ (indip. da ϑ) ... Poniamo:

(stima costantemente un qualcosa indip. da ϑ!)

$$\rightarrow \boxed{Y_i := X_i-\vartheta}$$ ($\Rightarrow (Y_1,\dots,Y_n)$ è campione di legge $N(0,1)$ (indip. da ϑ))

$\Rightarrow$ (per cui) $\overline{Y}=\overline{X}-\vartheta$ (cioè $\frac{1}{n}\sum_{i=1}^n(X_i-\vartheta)=\left(\frac{1}{n}\sum_{i=1}^n X_i\right)-\vartheta$), e

allora chiaramente $\sum_{i=1}^n(\underset{(\sim N(0,1))}{Y_i}-\underset{(\sim N(0,\frac{1}{n}))}{\overline{Y}})^2=\sum_{i=1}^n(X_i-\overline{X})^2$ ha legge indipendente da $\vartheta\in\Theta$. ✓

Oppure, ad esempio, anche $\max_{i=1,\dots,n}Y_i=\max_{i=1,\dots,n}(X_i-\vartheta)$ ha legge che non dipende da $\vartheta\in\Theta$ (ϑ $\min_{i=1,\dots,n}$). ✓

$\boxed{\text{NOTA}}$ Ricordiamo che una v.a.r. ha legge $\chi^2(n)$ nel caso sia somma dei quadrati di n v.a.r. di legge $N(0,1)$ indipendenti:

$W\sim\chi^2(n) \iff \exists\, X_1,\dots,X_n\sim N(0,1)$ indipendenti tali che $W=\sum_{i=1}^n X_i^2 = \|X\|_2^2$ $\quad (X:=(X_1,\dots,X_n))$.

Dunque, la norma euclidea al quadrato di un vettore gaussiano standard ha una legge del chi-quadro.

Ora, osserviamo che $\sum_{i=1}^n(Y_i-\overline{Y})^2=\sum_{i=1}^n Y_i^2+n\overline{Y}^2-\underset{(-2n\overline{Y}^2)}{2\overline{Y}\sum_{i=1}^n Y_i}=$

$$= \left(\sum_{i=1}^{n} Y_i^2\right) - n\overline{Y}^2 \quad : \text{ sarà } \chi^2(n-1) \text{ ? } \dots$$

($Y_i \sim N(0,1) \Rightarrow \sim \chi^2(n)$) ($= (\sqrt{n}\,\overline{Y})^2$, e $\sqrt{n}\,\overline{Y} \sim N(0,1)$)

Teorema. Sul modello stat. $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ dominato da μ, consideriamo due statistiche $S, T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$. Se T è esaustiva e completa, e se S è libera, allora S e T sono indipendenti (rispetto a P^θ, per ogni $\theta \in \Theta$).

Dim. La tesi è che, per ogni $\varphi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mis. limitata, risulta $E^{P^\theta}[\varphi(S) \mid T] = E^{P^\theta}[\varphi(S)]$ P^θ-q.c. per ogni $\theta \in \Theta$ (o μ-q.o.).

Ma, infatti, per ogni $\theta \in \Theta$ fissato, da una parte $E^{P^\theta}[\varphi(S)]$ è uno scalare reale, ed è indipendente da $\theta \in \Theta$ in quanto S è libera; d'altra parte, essendo T esaustiva, la v.a.r. $E^{P^\theta}[\varphi(S) \mid T]$ è anch'essa indipendente da $\theta \in \Theta$ (è T-misurabile), per cui possiamo porre $Y := E^{\cdot}[\varphi(S) \mid T] - E^{\cdot}[\varphi(S)]$, essendo effettivamente una sola v.a.r.. Ma questa v.a.r. Y è chiaramente T-misurabile e in $\bigcap_{\theta \in \Theta} \mathcal{L}^1(P^\theta)$, con anzi $E^{P^\theta}[Y] = 0$ per ogni $\theta \in \Theta$: la completezza di T porta quindi a concludere che $Y = 0$ μ-q.o., che è la tesi. □

Concludiamo vedendo una sorta di "inversione" del precedente teorema. Prima, però, ricordiamo ~~una~~ due nozioni.

(PARENTESI...)

Sia μ una misura σ-finita su $(\Omega, \mathcal{F})$, e sia $A \in \mathcal{F}$. Allora μ è "concentrata" su A (o "portata" da A) se $\mu(A^c) = 0$, e cioè [essendo $\mu(B) = \mu(A \cap B) + \mu(A^c \cap B)$ $\forall B \in \mathcal{F}$] se $\mu(B) = \mu(A \cap B)$ per ogni $B \in \mathcal{F}$, e cioè se $\mu = 1_A \cdot \mu$. || Sia ν un'altra misura σ-finita su $(\Omega, \mathcal{F})$. Allora μ e ν sono "singolari" (o "ortogonali") se esiste un evento $A \in \mathcal{F}$ tale che μ sia concentrata su A mentre ν sia concentrata su A^c: $\mu(A^c) = \nu(A) = 0$.

una (sottolineata)

Teorema. Sul modello stat. $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ dominato da μ, siano $S, T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ due statistiche indipendenti delle quali T esaustiva. Se le P^θ sono tali che, $\forall \theta_1, \theta_2 \in \Theta$, P^{θ_1} e P^{θ_2} non sono singolari tra loro, allora S è libera.

(ricordati!)

Dim. La tesi è che, per ogni $\varphi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mis. limitata e per ogni $\theta_1, \theta_2 \in \Theta$, risulti $\mathbb{E}^{P^{\theta_1}}[\varphi(S)] = \mathbb{E}^{P^{\theta_2}}[\varphi(S)]$.

Cominciamo col usare le ipotesi: dato che T è esaustiva, esiste $A \in \mathcal{F}$ con $\mu[A] = 1$ tale che, per ogni $\omega \in A$, sia
$$\mathbb{E}^{P^{\theta_1}}[\varphi(S) | T](\omega) = \mathbb{E}^{P^{\theta_2}}[\varphi(S) | T](\omega).$$

Ma S e T sono indipendenti, per cui

$\to$ esiste $A_1 \in \mathcal{F}$ con $P^{\theta_1}[A_1] = 1$ tale che, $\forall \omega \in A_1$, è
$$\mathbb{E}^{P^{\theta_1}}[\varphi(S) | T](\omega) = \mathbb{E}^{P^{\theta_1}}[\varphi(S)] \quad \text{(costante)};$$

$\rightarrow$ esiste $A_2 \in \mathcal{Y}$ con $P^{\theta_2}[A_2]=1$ tale che, $\forall \omega \in A_2$, è

$$\mathbb{E}^{P^{\theta_2}}[\varphi(S) \mid T](\omega) = \mathbb{E}^{P^{\theta_2}}[\varphi(S)] \quad \text{(costante)}$$

Concludiamo dimostrando che $\boxed{A \cap A_1 \cap A_2 \neq \emptyset}$, usando naturalmente la non-singolarità di P^{θ_1} e P^{θ_2}.

Cominciamo osservando di poter supporre $A_1 \subseteq A$ e anche $A_2 \subseteq A$.

[Infatti, ragionando (per simmetria) solo su A_1, $A \cap A_1$ sta in $\mathcal{Y}$, è $\subseteq A$ e ha ancora $\boxed{P^{\theta_1}[A \cap A_1] = 1}$ (oltre al fatto ovviamente che $\mathbb{E}^{P^{\theta_1}}[\varphi(S) \mid T] \equiv \underset{(\text{cost})}{\mathbb{E}^{P^{\theta_1}}[\varphi(S)]}$ su $A \cap A_1$, valendo su tutto A_1, ed essendo $A \cap A_1 \subseteq A_1$), e questo perché $\mu[A^c]=0 \overset{(P^{\theta_1} \ll \mu)}{\Rightarrow} P^{\theta_1}[A^c]=0$, e così $P^{\theta_1}[A]=1$).]

$[P^{\theta_1}[(A \cap A_1)^c] = P^{\theta_1}[A^c \cup A_1^c] \leq P^{\theta_1}[A^c] + P^{\theta_1}[A_1^c] = 0+0=0]$

Ma allora la tesi è che $\boxed{A_1 \cap A_2 \neq \emptyset}$. Questo segue appunto dalla non-singolarità di P^{θ_1} e P^{θ_2}, perché $P^{\theta_1}[A_1^c]=0$ e $P^{\theta_2}[A_2^c]=0$, per cui, se fosse $A_1 \cap A_2 = \emptyset$, allora (ad es.) $A_2 \subseteq A_1^c$ o così $\boxed{A_1 \subseteq A_2^c}$ e allora $\begin{cases} P^{\theta_1}[A_1^c]=0 \\ P^{\theta_2}[A_1]=0 \end{cases}$, che è contro l'ipotesi di non-singolarità di P^{θ_1} e P^{θ_2}. $\square$

$\boxed{\text{COROLLARIO}}$. Sia $(\Omega, \mathcal{Y}, \{P^\theta \mid \theta \in \Theta\})$ un mod. stat. dominato e con le P^θ mutuamente non-singolari, che ammette una statistica T esaustiva e completa. Allora tutte e sole le statistiche S libere (del mod.) sono indipendenti da T (rispetto a P^θ, per ogni $\theta \in \Theta$)

Un'applicazione. Consideriamo il modello stat. regolare canonico $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\})$ sul quale esiste un campione $(X_1,\dots,X_n)$ di una certa taglia $n\in\mathbb{N}_*$ fissata e di legge $N(m,\sigma^2)$, con σ^2 NOTA (per esempio, $=1$): dunque, $\Omega=\mathbb{R}^n$, $\mathcal{F}=\mathcal{B}(\mathbb{R}^n)$, $\Theta=\mathbb{R}$ $(\theta=m)$, μ dominante di Lebesgue n-dim., verosimiglianza $\mathcal{L}^m(x_1,\dots,x_n)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-m)^2\right)$ per ogni $x_i\in\mathbb{R}$, $X_i((x_1,\dots,x_n))=x_i$ $\forall\dots$. Essendo appunto σ^2 un numero fissato (>0), grazie a Neyman-Fisher, è chiaro che $T:=S_n=\sum_{i=1}^n X_i$ sia stat. esaustiva, e così che $\overline{X}:=\frac{S_n}{n}=\frac{1}{n}\sum_{i=1}^n X_i$ sia stat. esaustiva. Mostriamo che è anche completa, cioè che: per ogni $h:(\mathbb{R},\mathcal{B}(\mathbb{R}))\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mis. "regolare" (...), se $\mathbb{E}^{P^m}[h(\overline{X})]=0$ per ogni $m\in\mathbb{R}$, allora $h\equiv 0$ q.o. Ma infatti, semplicemente, per indipendenza delle X_i, abbiamo che $\overline{X}\sim N(m,\frac{\sigma^2}{n})$, per cui $\forall m\in\mathbb{R}$

$$\mathbb{E}^{P^m}[h(\overline{X})]=\int_{-\infty}^{\infty}h(x)\left(2\pi\frac{\sigma^2}{n}\right)^{-\frac{1}{2}}\exp\left(-\frac{n}{2\sigma^2}(x-m)^2\right)dx\ .\ \checkmark$$

Visto poi che ovviamente le P^m sono mutuamente NON-singolari, possiamo affermare che, per ogni statistica $S:(\Omega,\mathcal{F})\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$, S è libera $\Longleftrightarrow$ S e $\overline{X}$ sono indipendenti. $\checkmark$

L' **ESERCIZIO** che proponiamo adesso sfrutta la precedente equivalenza generale per dimostrare la seguente equivalenza particolare: quella che sia $A\in\mathbb{R}^{n\times n}$ simmetrica e semidefinita positiva, dove che

$\overline{X}$ e ${}^tXAX$ sono indipendenti $\Longleftrightarrow$ $\forall j=1,\dots,n$, $\sum_{i=1}^n a_{ij}=0$

(se indichiamo $(a_{ij})_{i,j=1,\dots,n}=A$).

Mostriamo infatti che $\sum_{i=1}^{n} a_{ij}=0\ \forall j=1,\dots,n \Leftrightarrow {}^tXAX$ è libera (dal modello)! Siano per questo $Y_i := X_i - m\ \ \forall i=1,\dots,n\ (\sim N(0,\sigma^2))$, e siano $Y := {}^t(Y_1,\dots,Y_n)$ e $X := {}^t(X_1,\dots,X_n)$ (così colonne di $\mathbb{R}^n$): allora X è un vettore gaussiano di legge $N_n(m, \sigma^2 I_n)$, e così Y è un vettore gaussiano di legge $N_n(O_{\mathbb{R}^n}, \sigma^2 I_n)$, e son tali che $Y = X - m\cdot e$, se $e := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$. ✓ In particolare ${}^tYAY$ è una statistica libera! Inoltre,

$${}^tXAX = {}^t(Y+me)\cdot A \cdot (Y+me) =$$

$$= {}^tYAY + \underbrace{m\,{}^teAY + m\,{}^tYAe}_{} + m^2\,{}^teAe =$$

$$\overset{(A \text{ simm.})}{=} {}^tYAY + 2m({}^teAY) + m^2({}^teAe).$$

D'altra parte, chiaramente, $\sum_{i=1}^{n} a_{ij}=0\ \forall j=1,\dots,n \Leftrightarrow$

$$\underset{(\text{simm.})}{\Leftrightarrow} {}^teA = O_{\mathbb{R}^n}, \ \underset{(\text{trasponendo})}{\overset{(\text{simm.})}{\Leftrightarrow}} A\cdot e = O_{\mathbb{R}^n}. \ \checkmark$$

Dunque ${}^teA=0 \Rightarrow {}^tXAX = {}^tYAY$, che è libera: resta da dimostrare $\boxed{{}^tXAX \text{ libera} \Rightarrow {}^teA=0}$. Ma, infatti, se ${}^tXAX$ è libera, allora $\#\{\mathbb{E}^{P_m}[{}^tXAX] \mid m\in\mathbb{R}\}=1$; ma $\forall m\in\mathbb{R}$, $\mathbb{E}^{P_m}[{}^tXAX] = \mathbb{E}^{P_m}[{}^tYAY] + m\,{}^teA\,\mathbb{E}^{P_m}[Y] +$

$+ m\, \mathbb{E}^{P^{\mu}}[{}^tY] A b + m^2({}^t b A b) = $ ($\mathbb{E}^{P^{\mu}}[Y]=0$, e allora anche $\mathbb{E}^{P^{\mu}}[{}^tY] = {}^t\mathbb{E}^{P^{\mu}}[Y] = 0$)

$= \mathbb{E}^{P^{\mu}}[{}^tYAY] + m^2({}^t bAb)$, per cui ${}^tXAX$ libera $\Rightarrow$

(indip. da m! (${}^tYAY$ è libera!))

(A reale, simm., semidef. positiva)

$\Rightarrow \; {}^t bAb = 0$. Ma allora $0 = {}^t bAb = \langle Ab; b\rangle =$

$= \langle \sqrt{A}(\sqrt{A}b); b\rangle \overset{(\sqrt{A}\text{ simm.})}{=} \langle \sqrt{A}b; \sqrt{A}b\rangle = \|\sqrt{A}b\|_2^2$, e

dunque $\sqrt{A}\cdot b = 0_{\mathbb{R}^n}$, e dunque anche applicando $\sqrt{A}$ ritrovo lo $0_{\mathbb{R}^n}$, ossia appunto $A\cdot b = 0_{\mathbb{R}^n}$. ✓ ☺

$\boxed{\text{NOTA}}$ (algebra lineare) Ricordiamo il fatto dell'esistenza della radice quadrata di una matrice $A \in \mathbb{R}^{n^2}$ $\underbrace{\text{simmetrica}}_{\text{(autovalori reali)}}$ e semidefinita $\underbrace{\text{positiva}}_{(\Rightarrow\, \text{e} \geq 0)}$.

Grazie al classico teorema spettrale, sappiamo che esistono n numeri non negativi $\lambda_1, \dots, \lambda_n$ (≥ 0) ed n vettori unitari in $\mathbb{R}^n$ $b_1, \dots, b_n$ che sono mutuamente ortogonali ($\Rightarrow \{b_1, \dots, b_n\}$ base ortonormale di $\mathbb{R}^n$) tali che, $\forall i = 1, \dots, n$, risulti

$Ab_i = \lambda_i b_i$. Dunque, $\forall i$, b_i è un autovettore di A relativo all'autovalore λ_i. Allora la matrice $U \in \mathbb{R}^{n^2}$, $U = [b_1 | \dots | b_n]$, è ortogonale (nel senso che è invertibile con $U^{-1} = {}^tU$) ed è tale che

$$\boxed{AU = UD} \quad \text{se } D \coloneqq \operatorname{diag}[\lambda_1, \dots, \lambda_n] = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

(DIAGONALIZZAZIONE DI A)

Scritto altrimenti, essendo ${}^tU = U^{-1}$, $\boxed{A = UD\,{}^tU}$. Inoltre, convenendo di scrivere $B=\sqrt{C} \overset{def.}{\iff} B^2 = C$, risulta evidente che vale l'ipotesi $\sqrt{D} = diag[\sqrt{\lambda_1},\dots,\sqrt{\lambda_n}]$. Pertanto, esiste dunque anche $\sqrt{A}$, ed è

$$\boxed{\sqrt{A} = U\sqrt{D}\,{}^tU}$$

: infatti $(\sqrt{A})^2 = A$, perché

$$(\sqrt{A})^2 = (U\sqrt{D}\,\underbrace{{}^tU)(U}\sqrt{D}\,{}^tU) \underset{({}^tU=U^{-1})}{=} U(\sqrt{D})^2\,{}^tU =$$

$$= UD\,{}^tU = A \; . \; \checkmark$$

Passiamo finalmente ai modelli dei esponenziali, premettendo il concetto della trasformata di Laplace di una misura μ σ-finita su $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.

Scelti quindi $k\in\mathbb{N}_+$ e μ misura σ-finita su $(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$, definiamo LA TRASFORMATA DI LAPLACE DI μ come la funzione

$$\mathcal{L}_\mu(\theta) : D_\mu \subseteq \mathbb{R}^k \longrightarrow [0,\infty) \quad \text{tale che}$$

→ $D_\mu := \left\{\theta\in\mathbb{R}^k \;\middle|\; \underset{(0\le)}{\int_{\mathbb{R}^k}} \exp(\langle\theta;x\rangle_{\mathbb{R}^k}) \, d\mu(x) < +\infty\right\}$

→ $\forall \theta\in D_\mu$, $\mathcal{L}_\mu(\theta) := \int_{\mathbb{R}^k}\exp(\langle\theta;x\rangle_2)\, d\mu(x)$.

Scritto diversamente, $\forall\theta\in D_\mu$, $\mathcal{L}_\mu(\theta) = \int_{\mathbb{R}^k} e^{\sum_{i=1}^k \theta_i x_i}\, d\mu(x)$ $\left(\text{se } \begin{cases}\theta=(\theta_1,\dots,\theta_k)\\ x=(x_1,\dots,x_k)\end{cases}\right)$.

(In particolare, per $\boxed{k=1}$, la trasformata di Laplace di una misura μ σ-finita su $\mathcal{B}(\mathbb{R})$ è $\mathcal{L}_\mu(\theta): D_\mu \subseteq \mathbb{R} \to [0,\infty)$ con $D_\mu := \{\theta\in\mathbb{R} \mid \int_{-\infty}^{\infty} e^{\theta x}\, d\mu(x) < \infty\}$ e, allora, $\mathcal{L}_\mu(\theta) := \int_{-\infty}^{\infty} e^{\theta x}\, d\mu(x)$ $\forall \theta \in D_\mu$.)

Richiamiamo / diamo per buone le tre, seguenti proprietà basilari della trasformata di Laplace $\mathcal{L}_\mu(\theta): D_\mu \to [0,\infty)$ di μ :

(1) (CONVESSITÀ) D_μ è un convesso di $\mathbb{R}^k$ (eventualmente $\emptyset$), e $\mathcal{L}_\mu(\theta)$ è ivi convessa!

(2) (REGOLARITÀ) Se $D_\mu \neq \emptyset$, allora $\mathcal{L}_\mu(\theta)\big|_{\mathring{D}_\mu} : \mathring{D}_\mu \to [0,\infty)$ è di classe C^∞ su tutto l'aperto $\mathring{D}_\mu$ di $\mathbb{R}^k$, e per ogni $i=1,\dots,k$ vale l'integrazione sotto il segno di integrale, nel senso che $(\forall i)$ $(\forall\theta\in\mathring{D}_\mu)$

$$\frac{\partial \mathcal{L}_\mu(\theta)}{\partial\theta_i} = \int_{\mathbb{R}^k} x_i\, e^{\langle\theta;x\rangle_k}\, d\mu(x) .$$

(In particolare, integrali del tipo $\int_{\mathbb{R}^k} x_{1}^{\alpha_1}\cdots x_{k}^{\alpha_k} e^{\langle\theta;x\rangle}\, d\mu(x)$ sono FINITI!)

(3) ("TRASFORMATA": caratterizza le misure!) Se $D_\mu\neq\emptyset$, e se $\mathcal{L}_\nu(\theta): D_\nu\to[0,\infty)$ è la trasformata di Laplace di un'altra misura ν σ-finita su $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, se anche $D_\nu \neq\emptyset$, e se esiste inoltre un aperto $\neq\emptyset$

$A \subseteq D_\mu\cap D_\nu$ tale che, per ogni $\theta\in A$, sia $\mathcal{L}_\mu(\theta)=\mathcal{L}_\nu(\theta)$ (o, cioè, $\mathcal{L}_\mu\big|_A \equiv \mathcal{L}_\nu\big|_A$), allora

$\boxed{\mu=\nu}$ $(\Rightarrow D_\mu = D_\nu)$. ✓

→ Sia $k\in\mathbb{N}_*$, sia $(\Omega,\mathcal{F})$ ~~misurabile~~ probabilizzabile, sia $T: (\Omega,\mathcal{F})\to(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$ misurabile, sia μ una misura σ-finita su $\mathcal{F}$, e sia

(H) un sottoinsieme aperto e convesso (e possibilmente $\neq\emptyset$) di

$$D_{(T)} := \left\{ \theta \in \mathbb{R}^k \;\middle|\; \int_\Omega e^{\langle\theta;T(\omega)\rangle_2} d\mu(\omega) < \infty \right\} \quad (\subseteq \mathbb{R}^k).$$
$(0\in)$

Osserviamo subito che $D = D_{T(\mu)}$ (nelle notazioni della trasf. di Laplace di $T(\mu)$) se ~~...~~ $T(\mu)$ [forse] è ancora ~~...~~ σ-finita (su $\mathcal{B}(\mathbb{R}^k)$)), ~~...~~ perché

$$\int_\Omega \exp(\langle\theta;T(\omega)\rangle) d\mu(\omega) \overset{\text{(TEO. INTEGR. MIS. IMMAGINE)}}{\underset{(\alpha<\infty,\ \text{almeno})}{=}} \int_{\mathbb{R}^k} \exp(\langle\theta;t\rangle) dT(\mu)(t) \quad (= \mathcal{L}_{T(\mu)}(\theta)).$$

In particolare, resterebbe definita la trasformata di Laplace di $T(\mu)$ su D, e quindi su Θ (che resta convesso, ma che è pure aperto).

Abbiamo, un modello stat. $(\Omega, \mathcal{F}, \{P^\theta \mid \theta\in\Theta\})$ che sia dominato da μ, con un tale $\Theta \subseteq D$ $(\subseteq\mathbb{R}^k)$ aperto e convesso ($\subseteq D$, aperto e convesso), è un $\boxed{\text{MODELLO STAT. ESPONENZIALE}}$ se esiste una funzione $C(\theta): \Theta \to (0,\infty)$ $(C(\theta)>0)$ (↳ a meno di restringere Θ) tale che la verosimiglianza (rispetto a μ) abbia la forma

$$\boxed{\mathcal{L}^\theta(\omega) = C(\theta)\exp(\langle\theta;T(\omega)\rangle_2)} \quad \forall(\theta,\omega)\in\Theta\times\Omega.$$

Osserviamo che, in effetti, $\mathcal{L}^\theta(\cdot)$ è una v.a.r. ≥ 0 con $\int_\Omega \mathcal{L}^\theta(\omega) d\mu(\omega) < \infty$ per definizione stessa di D, e quindi:

$\longrightarrow$ $\mathcal{L}^\theta(\omega) > 0$, per cui il modello è REGOLARE;

$\longrightarrow$ deve essere $\int_\Omega \mathcal{L}^\theta(\omega) d\mu(\omega) = 1$ $\forall\theta\in\Theta$, per cui necessariamente $C(\theta) = \dfrac{1}{\int_\Omega e^{\langle\theta;T(\omega)\rangle_2}\, d\mu(\omega)}$ per ogni $\theta\in\Theta$ (ottenuto imponendo che, $\forall\theta\in\Theta$, $\mathcal{L}_{T(\mu)}(\theta)$ sia >0 (non $=0$))

Inoltre, questa statistica $T:(\Omega,\mathcal{A})\to(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$ che in un certo senso "genera" il mod. stat. esponenziale in questione, risulta automaticamente esaustiva (grazie al teorema di Neyman-Fisher per un mod. stat. dominato).

Per cui, infatti, un mod. stat. esponenziale è sicuramente regolare e con automaticamente una statistica esaustiva. Scrivendo in forma "più propriamente esponenziale" (nel senso di scrivere $\mathcal{L}^\theta(\omega) = \exp(\ldots)$) : ponendo

$$\boxed{\forall\theta\in\Theta,\quad \psi(\theta) := \log\left(\int_\Omega e^{\langle\theta;T(\omega)\rangle} d\mu(\omega)\right)},$$

abbiamo chiaramente che $C(\theta) = \exp(-\psi(\theta))$ $\forall\theta\in\Theta$, da cui la forma canonica della verosimiglianza

$$\boxed{\mathcal{L}^\theta(\omega) = \exp\left(\langle\theta;T(\omega)\rangle_2 - \psi(\theta)\right) \quad \forall(\theta,\omega)\in\Theta\times\Omega.}$$

(**NOTA** Dunque, in un mod. esponenziale ci sono "facili [illegible] queste statistiche esaustive"!)

NOTA Per avere $\mathcal{L}^\theta(\omega)=\exp(\langle\theta;T(\omega)\rangle)C(\theta)$ (con $C(\theta)$ costante necessaria, e $\theta\in\Theta$ fissato, per avere che $\mathcal{L}^\theta(\cdot)$ sia una densità di probabilità), occorre almeno che $C(\theta)\geq 0$ $\forall\theta\in\Theta$. Allora la proprietà $C(\theta)>0$ $\forall\theta\in\Theta$ risulta automatica, perché, se esistesse $\theta_0\in\Theta$ tale che $C(\theta_0)=0$, allora $\mathcal{L}^{\theta_0}(\cdot)\equiv 0$ su Ω, che è assurdo.

Il primo risultato fondamentale è che, se $T(\mu)$ fosse σ-finita (su $\mathcal{B}(\mathbb{R}^k)$), allora T sarebbe pure completa per il mod. esponenziale, ed è per dimostrare tale affermazione che cominciamo col usare le proprietà basilari enunciate prima sulle trasformate di Laplace di una misura σ-finita su $\mathcal{B}(\mathbb{R}^k)$ ($T(\mu)$, in questo caso). Vediamole.

Teorema Sia $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$ un modello stat. esponenziale con verosimiglianza $\boxed{L^\theta(\omega) = C(\theta)\exp(\langle\theta;T(\omega)\rangle)}$ $\forall \theta,\omega \in \Theta\times\Omega$, nel senso che: sullo spazio mis. $(\Omega,\mathcal{A})$ esistono una misura σ-finita μ (dominante) ed una "statistica" $T : (\Omega,\mathcal{A}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ tale che, posto $D := \{\theta\in\mathbb{R}^k \mid \int_\Omega e^{\langle\theta;T(\omega)\rangle}d\mu(\omega) < \infty\}$, risulti $\Theta \subseteq D$ $(\subseteq \mathbb{R}^k)$ e Θ aperto e convesso ($\neq\emptyset$) tale che, $\forall\theta\in\Theta$, $C(\theta) > 0$ (θ, cioè, $\neq 0$), per cui $C(\theta) = \left(\int_\Omega e^{\langle\theta;T(\omega)\rangle}d\mu(\omega)\right)^{-1}$ (reciproco).
In particolare, il modello è regolare e la statistica T è esaustiva.
Allora, se $T(\mu)$ è σ-finita su $(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$, per cui
$$\int_\Omega e^{\langle\theta;T(\omega)\rangle}d\mu(\omega) = \int_{\mathbb{R}^k} e^{\langle\theta;t\rangle}dT(\mu)(t) \overset{?}{=} \mathcal{L}_{T(\mu)}(\theta) \quad \forall\theta\in D = D_{T(\mu)},$$
allora T è anche COMPLETA.

[Dim. La tesi è che, per ogni $h : (\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k)) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ mis. tale che $E^{P^\theta}[h(T)] = 0$ (sapere finito ed è) per ogni $\theta\in\Theta$, risulti $h\equiv 0$ $\mathcal{L}$-q.o.
Ora, infatti, $\forall\theta\in\Theta$, $E^{P^\theta}[h(T)] = \int_\Omega h(T(\omega))\underbrace{L^\theta(\omega)}d\mu(\omega) =$
$$= C(\theta)\int_\Omega h(T(\omega))e^{\langle\theta;T(\omega)\rangle}d\mu(\omega) = \underbrace{C(\theta)}_{(>0!)}\int_{\mathbb{R}^k} h(t)e^{\langle\theta;t\rangle}dT(\mu)(t),$$ e

per ciò, $\forall \theta \in \Theta$, $E^{P^\theta}[h(T)] = 0$ equivale ad avere che, $\forall \theta \in \Theta$,

$$\int_{\mathbb{R}^k} e^{\langle\theta;t\rangle} h^+(t)\, dT(\mu)(t) = \int_{\mathbb{R}^k} e^{\langle\theta;t\rangle} h^-(t)\, dT(\mu)(t).$$ Allora,

se effettivo $T(\mu)$ è σ-finita su $B(\mathbb{R}^k)$, allora anche le due misure su $B(\mathbb{R}^k)$ $h^+ . T(\mu)$, $h^- . T(\mu)$ sono σ-finite, e effettivamente tale che $\mathcal{L}_{h^+ . T(\mu)}(\theta) = \mathcal{L}_{h^- . T(\mu)}(\theta)$ per ogni $\theta \in \Theta$ (cioè ossia che $\Theta \subseteq D_{h^+ . T(\mu)} \cap D_{h^- . T(\mu)}$). Ma Θ è un aperto $\neq \phi$, per ciò deduciamo (per (3)) che $h^+ . (T(\mu)) = h^- . (T(\mu))$, e cioè che $h^+ = h^-$ $\mathcal{L}$. q.o. $\square$]

$\boxed{\text{OSS.}}$ (Diamo un'idea del perché in diversi ESEMPI visti (di modello stat.), scegli e campioni $(X_1, \dots, X_m)$ di una certa legge (dominata), avevamo che $T = X_1 + \dots + X_n$ fosse stat. esaustiva.)

Consideriamo $k \in \mathbb{N}_*$, $\Theta \subseteq \mathbb{R}^k$ un insieme aperto e convesso $\neq \phi$, e per ogni $\theta \in \Theta$ una legge di probabilità μ^θ su $(\mathbb{R}, B(\mathbb{R}))$, dominate da ν, e consideriamo il modello stat. $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$ dominato canonico sul quale esiste un campione $(X_1, \dots, X_m)$ di taglia $m \in \mathbb{N}_*$ fissata e di legge μ^θ (per ogni $\theta \in \Theta$ fissato): dunque $\Omega = \mathbb{R}^m$, $\mathcal{A} = B(\mathbb{R}^m)$, $X_i((x_1, \dots, x_m)) = x_i$ ($\forall i = 1, \dots, m$ e $\forall x_1, \dots, x_m \in \mathbb{R}$), $P^\theta = (\mu^\theta)^{\otimes m}$, e la misura dominante è $\mu = (\nu)^{\otimes m}$ (quella che domina qui $\forall \mu^\theta$) (una). In particolare, la verosimiglianza del modello (rispetto a μ) sarebbe

$$\mathcal{L}^\theta(x_1, \dots, x_m) = \prod_{i=1}^{m} \frac{d\mu^\theta}{d\nu}(x_i) \text{ “}=\text{” } \prod_{i=1}^{m} \mathcal{L}^\theta(x_i) \text{”}.$$

Ebbene, supponiamo che i singoli fattori del modello siano modelli esponenziali, nel senso: ~~[illegible]~~ $\forall i=1,\dots,M$ esiste

$T_i(x) : (\mathbb{R}, \mathcal{B}(\mathbb{R}))_{\mu^\alpha} \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ (indip. da $\vartheta \in \Theta$) tale

che, $\forall \vartheta \in \Theta$ e $\forall x \in \mathbb{R}$, risulti $\frac{d\mu^\vartheta}{d\nu}(x) = C_i(\vartheta) e^{\langle \vartheta; T_i(x) \rangle}$ per

un'opportuna $C_i(\vartheta) : \Theta \to (0,\infty)$ (nell'ipotesi aggiuntiva che $\Theta \subseteq$

$\subseteq \bigcap_{i=1}^{M} D_i$, $D_i \doteq \{\vartheta \in \mathbb{R}^k \mid \int_{\mathbb{R}} e^{\langle \vartheta; T_i(x) \rangle} d\nu(x) < \infty\}$). Allora,

$\forall x = (x_1, \dots, x_k) \in \mathbb{R}^M = \Omega$, $T(x) \doteq \sum_{i=1}^{M} T_i(x_i)$, o

cioè $\boxed{T \doteq \sum_{i=1}^{M} T_i(X_i)}$, è una statistica (per il modello (prodotto))

tale che, $\forall \vartheta \in \mathbb{R}^k$, $\int_{\mathbb{R}^M (\Omega)} \underbrace{e^{\langle \vartheta; T(x) \rangle}}_{\prod_{i=1}^{M} e^{\langle \vartheta; T_i(x_i) \rangle}} \underbrace{d\mu(x)}_{d(\nu)^{\otimes M}(x) \; (d\nu(x_1) \dots d\nu(x_M))}$ è $< \infty$ se e solo

se $\vartheta \in \bigcap_{i=1}^{M} D_i$, ossia $D = \bigcap_{i=1}^{M} D_i$, e quindi allora

$\Theta \subseteq D$, e risulta ($\forall \vartheta \in \Theta$ e $\forall x_1, \dots, x_M \in \mathbb{R}$)

$$L^\vartheta(x_1, \dots, x_M) = \prod_{i=1}^{M} \frac{d\mu^\vartheta}{d\nu}(x_i) = \prod_{i=1}^{M} C_i(\vartheta) e^{\langle \vartheta; T_i(x_i) \rangle} =$$

$$= \underbrace{\left(\prod_{i=1}^{M} C_i(\vartheta) \right)}_{(=: C(\vartheta) > 0!)} \cdot \exp(\langle \vartheta; T(x_1, \dots, x_M) \rangle)$$ ed anche il

($\boxed{\text{OSS.}}$ $\int_{\mathbb{R}^M} e^{\langle \vartheta; \sum_{i=1}^{M} T_i(x_i) \rangle} d\mu(x_1, \dots, x_M) = \prod_{i=1}^{M} \int_{-\infty}^{+\infty} e^{\langle \vartheta; T_i(x) \rangle} d\nu(x)$!)

mod. prodotto è esponenziale. / In particolare, se i "modelli fattori" sono esponenziali di statistica generatrice T", allora il "modello prodotto" è esponenziale di statistica generatrice $T(x_1) + \dots + T(x_M)$, $x_i \in \mathbb{R}$, la quale in particolare è <u>invariante</u>.

Prima degli esempi/esercizi, qualche calcolo differenziale.

Proposizione Sia $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ un modello stat. esponenziale di verosimiglianza $L^\theta(\omega) = C(\theta)\exp(\langle \theta ; T(\omega)\rangle)$ $\forall (\theta,\omega) \in \Theta \times \Omega$.
Con questo, intendiamo: sullo spazio mis. $(\Omega, \mathcal{F})$ esistono una misura μ σ-finita (dominante) ed una statistica $T : (\Omega,\mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ tale che, posto $D := \{\theta \in \mathbb{R}^k \mid \int_\Omega e^{\langle \theta ; T(\omega)\rangle} d\mu(\omega) < \infty\}$ $(\subseteq \mathbb{R}^k)$, risulti $\Theta \subseteq D$ ed inoltre che Θ sia un aperto, convesso e $\neq \emptyset$ tale che $C(\theta) > 0$ per ogni $\theta \in \Theta$. In particolare, il modello è regolare e T è esaustiva. Inoltre, in realtà,

$$C(\theta) = \left(\int_\Omega e^{\langle \theta ; T(\omega)\rangle} d\mu(\omega) \right)^{-1} = \exp(-f(\theta)) \quad \forall \theta \in \Theta$$

se $f(\theta) := \log\left(\int_\Omega e^{\langle \theta ; T(\omega)\rangle} d\mu(\omega)\right)$ $(\forall \theta)$, e possiamo scrivere

$$L^\theta(\omega) = \exp(\langle \theta ; T(\omega)\rangle - f(\theta)) \quad \forall (\theta,\omega) \in \Theta \times \Omega .$$

Sappiamo, sempre se $T(\mu)$ è σ-finita su $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, per cui [$D = D_{T(\mu)}$ ed infatti] $\int_\Omega e^{\langle \theta ; T(\omega)\rangle} d\mu(\omega) = \int_{\mathbb{R}^k} e^{\langle \theta ; t\rangle} dT(\mu)(t) = \mathcal{L}_{T(\mu)}(\theta)$ $\forall \theta \in \Theta$, allora T è completa (OK) e f è C^∞ nell'aperto Θ. Inoltre valgono le due seguenti formule:
se $\theta := (\theta_1, \dots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$ e se $T = (T_1, \dots, T_k) \in (\mathbb{R}^k)^\Omega$ (cioè $T_i : (\Omega,\mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, \dots, k$, sono le componenti di T), allora, $\forall\, i, j = 1, \dots, k$, vale

$$\boxed{\boxed{\frac{\partial \psi}{\partial \vartheta_i}(\vartheta) = \mathbb{E}^{P^\vartheta}[T_i]}\ , e\ \boxed{\frac{\partial \psi}{\partial \vartheta_i \partial \vartheta_j}(\vartheta) = \mathrm{Cov}^{P^\vartheta}[T_i, T_j]}\ .}$$

$\boxed{\text{NOTA}}$ Date due v.a.r. $X, Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la Covarianza tra X e Y rispetto ad una probabilità P su $\mathcal{F}$ resta definita se X e Y stanno in $L^2(P)$, ed in tal caso vale

$$\mathrm{Cov}^P[X, Y] := \mathbb{E}^P[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}^P[XY] - \mathbb{E}^P[X]\mathbb{E}^P[Y].$$

[Dim. Vediamo anzitutto che ψ è C^∞ su tutto Θ, e che $T_{(i)}$ ammette tutti i momenti finiti (e che, anzi, queste due condizioni si equivalgono), nell'ipotesi che $T(\mu)$ sia σ-finita. Infatti, in tal caso, $\int_\Omega e^{\langle \vartheta; T(\omega)\rangle} d\mu(\omega) = \mathcal{L}_{T(\mu)}(\vartheta)$ per ogni $\vartheta \in \Theta$, per cui

$\psi(\vartheta) = \log(\mathcal{L}_{T(\mu)}(\vartheta))$ per ogni $\vartheta \in \Theta$. Ora, sappiamo che (per ②) $\mathcal{L}_{T(\mu)}$ è C^∞ su tutto $\mathring{D}_{T(\mu)} = \mathring{D}$ $(\supseteq \Theta)$, per cui ψ è C^∞ su tutto $\mathring{D}$ e, a maggior ragione, su tutto Θ.

Non solo: sappiamo di poter derivare sotto il segno d'integrale! Dunque, $\forall j = 1, \dots, k$, $\frac{\partial}{\partial \vartheta_j} \int_{\mathbb{R}^k} e^{\langle \vartheta; t\rangle} dT(\mu)(t) = \int_{\mathbb{R}^k} t_j e^{\langle \vartheta; t\rangle} dT(\mu)(t)$,

e poi allora, $\forall i = 1, \dots, k$, $\frac{\partial^2}{\partial \vartheta_j \partial \vartheta_i} \int_{\mathbb{R}^k} e^{\langle \vartheta; t\rangle} dT(\mu)(t) =$

$$= \frac{\partial}{\partial \vartheta_i} \int_{\mathbb{R}^k} t_j e^{\langle \vartheta; t\rangle} dT(\mu)(t) = \int_{\mathbb{R}^k} t_i t_j e^{\langle \vartheta; t\rangle} dT(\mu)(t).$$

Scritto altrimenti, chiaramente, $\forall i, j = 1, \dots, k$, vale ($\forall \vartheta \in \Theta$)

$$\frac{\partial}{\partial \vartheta_j} \mathcal{L}_{T(\mu)}(\vartheta) = \int_{\mathbb{R}^k} t_j e^{\langle \vartheta; t\rangle} dT(\mu)(t) \quad e \quad \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \mathcal{L}_{T(\mu)}(\vartheta) = \int_{\mathbb{R}^k} t_i t_j e^{\langle \vartheta; t\rangle} dT(\mu)(t).$$

D'altra parte, allora, $\forall j=1,\dots,k$ ($e\ \forall \theta\in\Theta$), $\frac{\partial\psi}{\partial\theta_j}(\theta) = \frac{\partial}{\partial\theta_j}\left(\log L_{T(\mu)}(\theta)\right) =$

$$= \underbrace{\frac{1}{L_{T(\mu)}(\theta)}}_{(C(\theta)!)}\ \underbrace{\frac{\partial}{\partial\theta_j} L_{T(\mu)}(\theta)}_{\left(\int_{\mathbb{R}^k} t_j e^{\langle\theta;t\rangle} dT(\mu)(t)\right)} = \int_{\mathbb{R}^k} t_j \left(C(\theta) e^{\langle\theta;t\rangle}\right) dT(\mu)(t) =$$

$$= \int_\Omega T_j(\omega) \underbrace{L^\theta(\omega)\, d\mu(\omega)}_{(dP^\theta(\omega))} = \mathbb{E}^{P^\theta}[T_j]$$, effettivamente. ✓

Calcoliamo l'ulteriore derivata, derivando direttamente sotto il segno d'integrale: visto cioè che $\frac{\partial\psi}{\partial\theta_j}(\theta) = \int_{\mathbb{R}^k} t_j\, C(\theta) e^{\langle\theta;t\rangle} dT(\mu)(t) =$

$= \int_{\mathbb{R}^k} t_j \exp(\langle\theta;t\rangle - \psi(\theta))\, dT(\mu)(t)$, viene che $\frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}(\theta) =$

$$= \int_{\mathbb{R}^k} \frac{\partial}{\partial\theta_i}\left(t_j \exp(\langle\theta;t\rangle - \psi(\theta))\right) dT(\mu)(t) =$$

$$= \int_{\mathbb{R}^k} t_j \exp(\langle\theta;t\rangle - \psi(\theta)) \Big(t_i - \underbrace{\frac{\partial\psi}{\partial\theta_i}(\theta)}_{\substack{\text{(visto ora)} \\ (\mathbb{E}^{P^\theta}[T_i]! \ (\in\mathbb{R}))}}\Big)\, dT(\mu)(t) =$$

$$= \underbrace{\int_{\mathbb{R}^k} t_i t_j \exp(\langle\theta;t\rangle - \psi(\theta))\, dT(\mu)(t)} - \mathbb{E}^{P^\theta}[T_i] \underbrace{\int_{\mathbb{R}^k} t_j \exp(\cdots)\, dT(\mu)(t)} =$$

$$= \mathbb{E}^{P^\theta}[T_i T_j] - \mathbb{E}^{P^\theta}[T_i]\,\mathbb{E}^{P^\theta}[T_j] = Cov^\theta[T_i, T_j]. \quad \square$$

(Come su!)

(ES) Consideriamo il modello stat. "prodotto" canonico $(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in\Theta\})$ regolare nel quale esiste un campione $(X_1,\dots,X_n)$ di taglia $n\in\mathbb{N}_*$ formato e di legge $N(m,\sigma^2)$, $(m,\sigma^2)\in\Theta = \mathbb{R}\times(0,\infty)$ (aperto e convesso di $\mathbb{R}^2$): cioè $\Omega = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, m…

dominante μ di Lebesgue n-dim., $X_i((x_1,\dots,x_n))=x_i$, e verosimiglianza

$$\mathcal{L}^{(m,\sigma^2)}(x_1,\dots,x_n)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\Big(\frac{-1}{2\sigma^2}\underbrace{\sum_{i=1}^n(x_i-m)^2}_{\left(\sum_{i=1}^n x_i^2-2m\sum_{i=1}^n x_i+nm^2\right)}\Big)=$$

$$=(2\pi)^{-\frac{n}{2}}\exp\Big(-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^n x_i^2-2m\sum_{i=1}^n x_i+nm^2\Big)-n\log\sigma\Big)=$$

$$=\exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2+\frac{m}{\sigma^2}\sum_{i=1}^n x_i-\Big(\frac{nm^2}{2\sigma^2}+n\log\sigma+\frac{n}{2}\log(2\pi)\Big)\Big).$$

Intendendo quindi $\Theta=\mathbb{R}\times(0,\infty)=\left\{\left(\frac{m}{\sigma^2},\frac{1}{2\sigma^2}\right)\,\middle|\,m\in\mathbb{R},\sigma^2>0\right\}$, e posto $T:=\left(\sum_{i=1}^n X_i,-\sum_{i=1}^n X_i^2\right)$ (statistica $(\Omega,\mathcal{B})\to(\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2))$) e

$\psi\left(\frac{m}{\sigma^2},\frac{1}{2\sigma^2}\right):=\frac{n}{2}\frac{m^2}{\sigma^2}+n\log(\sigma)+\frac{n}{2}\log(2\pi)$, allora abbiamo effettivamente trovato che, $\forall\vartheta\in\Theta$ e $\forall x_i\in\mathbb{R}$,

$$\mathcal{L}^\vartheta(x_1,\dots,x_n)=\exp\left(\langle\vartheta;T(x_1,\dots,x_n)\rangle_{\mathbb{R}^2,2}-\psi(\vartheta)\right),$$ e per ciò

il modello stat. in questione è esponenziale ! (Il che è più del dire che T è esaustiva (e questo l'avevamo anche già verificato)).

$\boxed{\text{NOTA}}$ Questo NON è affatto sorprendente se ripensiamo al mod. "fattore" (cioè $n=1$) ! In effetti, come capito, per un modello campionario relativo ad un campione possiamo sempre supporre $\boxed{n=1}$!

Restringiamoci pertanto al caso $n=1$: $(\Omega,\mathcal{B})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$, μ di Lebesgue su $\mathcal{B}(\mathbb{R})$, X l'identità su $\mathbb{R}$, e densità di prob. rispetto a μ data da $\mathcal{L}^\vartheta(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$ $\forall x\in\mathbb{R}$ e $\forall\vartheta\in\Theta$ ($\Theta\subseteq\mathbb{R}^2$ lo stesso), per cui ovviamente $T=(X,-X^2)$ lo "genera" in "modo esponenziale", essendo

$\mathcal{L}^\vartheta(x) = \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{m}{\sigma^2}x - \left(\frac{m^2}{2\sigma^2} + \log(\sigma) + \frac{1}{2}\log(2\pi)\right)\right)$. Allora, (21)

osserviamo che nel caso o $m \in \mathbb{R}$ noto oppure $\sigma^2 \in (0,\infty)$ noto si hanno delle semplificazioni: se $\boxed{m \in \mathbb{R} \text{ è noto}}$, allora semplicemente torniamo indietro a $\mathcal{L}^\vartheta(x) = (2\pi\sigma^2)^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$ col intendere $\Theta := (0,\infty) = \left\{\frac{1}{2\sigma^2} \mid \sigma^2 > 0\right\}$, per cui $T := -(X-m)^2$ è tale che $\mathcal{L}^\vartheta(x) = C(\vartheta)\exp(\vartheta \cdot T(x))$. ✓ Se invece $\boxed{\sigma^2 > 0 \text{ è noto}}$ (è fissato),

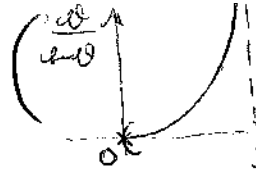
allora osserviamo che $\mathcal{L}^\vartheta(x) = \exp\left(\frac{m}{\sigma^2}x - \left(\frac{m^2}{2\sigma^2} + \log(\sigma) + \frac{1}{2}\log(2\pi)\right)\right) \cdot$

$\cdot \exp\left(-\frac{1}{2\sigma^2}x^2\right)$, per cui passiamo alla nuova dominante (equivalente) di densità $\exp\left(-\frac{x^2}{2\sigma^2}\right)$ rispetto alla μ di Lebesgue, e allora intendiamo $\Theta := \mathbb{R} = \left\{\frac{m}{\sigma^2} \mid m \in \mathbb{R}\right\}$ e $T := X$ per cui la verosimiglianza diviene $\exp(\vartheta \cdot T(x) - f(\vartheta))$. //
(rispetto quindi alla nuova mis. dominante)

ES. Consideriamo il mod. stat. $(\Omega, \mathcal{F}, \{P^\vartheta \mid \vartheta \in \Theta\})$ canonicamente associato ad un campione $(X_1, \dots, X_n)$ di taglia $n \in \mathbb{N}_*$ (scelta) e di legge $B(1,\vartheta)$, $\vartheta \in \Theta = (0,1)$ (aperto, convesso di $\mathbb{R}$) : $\Omega = \{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$, $X_i((x_1, \dots, x_n)) = x_i$, mis. dominante μ che conta i punti, verosimiglianza relativa $\mathcal{L}^\vartheta(x_1, \dots, x_n) = \prod_{i=1}^n \vartheta^{x_i}(1-\vartheta)^{1-x_i} =$

$= \vartheta^{\sum_{i=1}^n x_i}(1-\vartheta)^{1-\sum_{i=1}^n x_i}$. Scritto altrimenti, se $x = (x_1, \dots, x_n)$ e se $\boxed{T = \sum_{i=1}^n X_i}$, allora $\mathcal{L}^\vartheta(x) = \vartheta^{T(x)}(1-\vartheta)^{1-T(x)}$. Ora

$\vartheta^{T(x)}(1-\vartheta)^{1-T(x)} \overset{(ovvio)}{=} \vartheta^{T(x)}\frac{1}{(1-\vartheta)^{T(x)-1}} = \left(\frac{\vartheta}{1-\vartheta}\right)^{T(x)}(1-\vartheta)$,

e allora $\mathcal{L}^\vartheta(x) = \exp\left(T(x) \cdot \log\left(\frac{\vartheta}{1-\vartheta}\right) + \log(1-\vartheta)\right)$: pertanto, ponendo e $\Theta = \mathbb{R} = \{\log(\frac{\vartheta}{1-\vartheta}) \mid \vartheta \in (0,1)\}$  ,

che comunque resta in biiezione con $(0,1)$, possiamo scrivere

$L^\vartheta(x) = \exp(\vartheta \cdot T(x) - f(\vartheta))$ ed il modello è esponenziale. ✓

(opportuna ...)

$\boxed{\text{NOTA}}$ Di meno, era equivalente fare lo stesso nel caso $n=1$, ossia per le densità $\vartheta^x (1-\vartheta)^{1-x} = \exp(x \cdot \log(\frac{\vartheta}{1-\vartheta}) + \log(1-\vartheta))$ ✓

$\boxed{\text{IDEA}}$ Per arrivare ad un mod. esponenziale, può essere necessario cambiare Θ, μ dominante, oppure entrambi!

ES. Partiamo direttamente dalle leggi Π_ϑ di Poisson di parametro $\vartheta \in (0,\infty)$ su $\mathbb{N}$, o meglio su $\mathbb{R}$ ma in relazione di equivalenza con la misura μ su $\mathcal{B}(\mathbb{R})$ concentrata su $\mathbb{N}$ che conta i punti di $\mathbb{N}$: dunque le densità ~~di~~ $L^\vartheta(x) = e^{-\vartheta} \frac{\vartheta^x}{x!}$ $\forall \vartheta > 0$ e $\forall x \in \mathbb{N}$. Quindi, cambiando, passando da μ a $\frac{1}{x!} \mu$ $(\sim \mu)$, la densità diviene $L^\vartheta(x) = e^{-\vartheta} \vartheta^x$. A questo punto, $\vartheta^x = \exp(x \log(\vartheta))$ implica $L^\vartheta(x) = \exp(x \cdot \log(\vartheta) - \vartheta)$, che è tutto esponenziale se passiamo ai parametri $\Theta = \mathbb{R} = \{\log(\vartheta) \mid \vartheta > 0\}$ (in biiezione con $(0,\infty)$, comunque).

Da quanto fin ora scoperto, deduciamo che il modello stat. prodotto canonico $(\Omega, \mathcal{A}, \{P^\vartheta \mid \vartheta \in \Theta\})$ relativo associato ad un campione ~~di legge~~ $(X_1, \dots, X_n)$ di taglia $n \in \mathbb{N}^*$ estratto e di legge Π_λ di Poisson di parametro $\lambda > 0$, risulta esponenziale se consideriamo come misura dominante "$\frac{1}{x_1! \dots x_n!} \cdot \mu$" con μ su $\mathcal{B}(\mathbb{R}^n)$ concentrata su $\mathbb{N}^n$ che ivi conta i punti $(x_i \in \mathbb{N})$, se $\Theta = \mathbb{R} = \{\log(\lambda) \mid \lambda > 0\}$, e se $T = \sum_{i=1}^n X_i$. ✓

ES. Questo esempio si basa su v.a. reali di ~~legge~~ densità Gamma di parametri $r, \lambda \in (0,\infty)$, la quale pertanto ricordiamo preliminarmente.

→ La densità Gamma $\Gamma(r,\lambda)$ di parametri $r,\lambda > 0$ (reali) che definizione che si basa sulla funzione Gamma $\Gamma(r)$,

$\Gamma(r) : (0,+\infty) \to (0,+\infty)$, data da

$$\boxed{\Gamma(r) := \int_0^{+\infty} x^{r-1} e^{-x}\,dx \quad \forall r>0}$$

(integrale potenzialmente improprio, che però è effettivamente reale ($e>0$) in quanto x^{r-1} è integrabile in $q_{\text{ro}}(0,\varepsilon)$ per quanto $r-1>-1$, e affatto $r>0$; inoltre, a $+\infty$, domina il e^{-x}!) ; questo integrale NON è calcolabile „esplicitamente" (o, cioè, non può che esser lasciato in forma integrale), ma qualcosa si può dire: (1) $\Gamma(1) = 1$.

$\left[\Gamma(1) = \int_0^{+\infty} e^{-x}\,dx = \left[e^{-x}\right]_{x=+\infty}^{x=0} = 1 - 0 = 1 . \checkmark\right]$

(2) $\forall r>1$, $\Gamma(r) = (r-1)\Gamma(r-1)$. $(\Rightarrow \forall r>0, \Gamma(r+1) = r\Gamma(r))$

$(r-1>0)$

$\left[\Gamma(r) = \int_0^{+\infty} \underbrace{x^{r-1}}_{\text{(deriva)}}\, \underbrace{e^{-x}}_{\text{(integro)}}\,dx \overset{\text{(INT. PARTI)}}{=} \underbrace{\left[x^{r-1}e^{-x}\right]_{x=+\infty}^{x=0}}_{\overset{\parallel}{0}\ (r-1>0)!} + (r-1)\int_0^{+\infty} x^{(r-1)-1} e^{-x}\,dx =\right.$ $(r-1>0!)$

$\left. = (r-1)\Gamma(r-1) . \checkmark\right]$

(3) $\Rightarrow \forall n \in \mathbb{N}_*$, $\boxed{\Gamma(n) = (n-1)!}$.

$\left[\forall n \geq 1\right.$, se $n=1$ allora, in effetti, $\Gamma(1) \overset{(1)}{=} 1 = 0!$; se, invece, $n>1$, allora $\Gamma(n) \overset{(2)}{=} (n-1)\Gamma(n-1)$, e „così via" (induzione). $\left.\checkmark\right]$

→ Finalmente, la densità Gamma di parametri r e $\lambda > 0$ (fissati) (su $\mathbb{R}$, rispetto alla mis. di Lebesgue) è data da, $\forall x \in \mathbb{R}$,

$$\boxed{\Gamma(r,\lambda)(x) := \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x)}$$

(ovvero, vale $\frac{1}{\Gamma(r)}\lambda^r x^{r-1} e^{-\lambda x}$ per ogni $x>0$, e vale 0 altrimenti).
Osserviamo che, in effetti, $\int_0^{+\infty} \Gamma(r,\lambda)(x)\,dx = 1$.

$$\left[\int_0^{+\infty} \underbrace{\lambda^r x^{r-1}}_{(=(\lambda x)^{r-1}\cdot\lambda)} e^{-\lambda x}\,dx \underset{\substack{\lambda x =: y \\ dx = \frac{1}{\lambda}dy \\ (0,+\infty)\mapsto(0,\infty)}}{=} \int_0^{+\infty} y^{r-1} e^{-y}\,dy = \Gamma(r). \checkmark \right]$$

In particolare, $\forall n \in \mathbb{N}_*$ e $\forall \lambda > 0$, e $\forall x \in \mathbb{R}$,

$$\Gamma(n,\lambda)(x) = \frac{1}{(n-1)!}\lambda^n x^{n-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x)$$

, ed ancor più in particolare, $\forall \lambda>0$ e $\forall x \in \mathbb{R}$, $(n=1)$

$$\boxed{\Gamma(1,\lambda)(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x)}$$

, che è la densità esponenziale "$\mathcal{E}(\lambda)$" di parametro $\lambda > 0$ (su $\mathbb{R}$).

$\boxed{\text{OSS}}$ Una v.a.r. $X \sim \Gamma(r,\lambda)$. Lebesgue (su $(\Omega,\mathcal{A},P)$, discreta) ammette momenti di ogni ordine, e più precisamente vale che, $\forall \beta > 0$,

$$\mathbb{E}^P[X^\beta] = \frac{\Gamma(r+\beta)}{\Gamma(r)\cdot\lambda^\beta}$$

, ed in particolare X che speranza $(\beta=1)$ $\boxed{\mathbb{E}^P[X] = \frac{r}{\lambda}}$. (Ritroviamo, in particolare,

che una v.a. di densità $\mathcal{E}(\lambda)$ ha speranza $\frac{1}{\lambda}$.) Inoltre, se $X \sim \Gamma(r,\lambda)$ e $Y \sim \Gamma(s,\lambda)$ $(r,s,\lambda>0)$ sono indipendenti, allora $X+Y \sim \Gamma(r+s,\lambda)$.
(→ di varianza $1/\lambda^2$, perché $\mathbb{E}[X^2]=\frac{\Gamma(3)}{\lambda^2}=2/\lambda^2$)

[Per esercizio, verifichiamo solo che $\mathbb{E}[X^\beta]=\frac{\Gamma(r+\beta)}{\Gamma(r)\cdot\lambda^\beta}$: ma, infatti,

$$\mathbb{E}[X^\beta]=\int_0^{+\infty} x^\beta\cdot\frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x}dx,$$ per cui la tesi è

$$\int_0^{+\infty} x^\beta\lambda^r x^{r-1}e^{-\lambda x}dx=\frac{\Gamma(r+\beta)}{\lambda^\beta},$$ e cioè che

$$\int_0^{+\infty}\lambda^{(r+\beta)}x^{(r+\beta)-1}e^{-\lambda x}dx=\Gamma(r+\beta),$$ che è evidente. $\square$]

$$\left(=(\lambda x)^{(r+\beta)-1}\cdot e^{-\lambda x}\cdot d(\lambda x)\,!\right)$$

ES. Consideriamo il modello statistico regolare $(\Omega,\mathcal{A},\{P_\vartheta\,|\,\vartheta\in\Theta\})$ canonicamente associato ad un campione $(X_1,\dots,X_n)$ di taglia $n\in\mathbb{N}_*$ fissato e di legge $\Gamma(r,\lambda)$, $r,\lambda>0$ „dipendenti" da $\vartheta\in\Theta$: dunque $\Omega=(\mathbb{R}_+)^n$, $\mathcal{A}=\mathcal{B}((\mathbb{R}_+)^n)$, $\Theta=(0,\infty)^2$ (formalmente da cambiare...), $X_i((x_1,\dots,x_n))=x_i$, mis. dominante μ di Lebesgue, verosimiglianza $\mathcal{L}^\vartheta(x_1,\dots,x_n)=\prod_{i=1}^n\mathcal{L}^\vartheta(x_i)$ dove $\forall x>0$

$$\mathcal{L}^\vartheta(x)=\frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x}.$$ Osserviamo che

$$x^{r-1}=\exp((r-1)\log(x)),$$ per cui

$$\mathcal{L}^\vartheta(x)=\underbrace{\frac{\lambda^r}{\Gamma(r)}}_{(\text{dip. da }\Theta)}\cdot\exp\big(\lambda\cdot(-x)+(r-1)\log(x)\big),$$ che è

di tipo esponenziale se intendiamo $\Theta=(0,\infty)\times(-1,\infty)=$ (aperto, convesso di $\mathbb{R}^2$)
$=\{\vartheta=(\lambda,r-1)\,|\,\lambda>0 \text{ e } r>0\}$ e se consideriamo

$T := (-X, \log(X))$ (la quale, pertanto, risulta sufficiente per il modello con $n=1$). Equivalentemente, il modello prodotto sempre con $\Theta = (0,\infty) \times (-1,\infty)$ è di tipo esponenziale con

$$T = \left(-\sum_{i=1}^{n} X_i \,, \sum_{i=1}^{n} \log(X_i)\right).$$

(ES.) Consideriamo "il" mod. stat. $(\Omega, \mathcal{F}, [P^\theta | \theta \in \Theta])$ regolare canonicamente associato ad un campione ~~di taglia~~ $(X_1, \dots, X_n)$ di taglia $n \in \mathbb{N}_*$ (fissata) e di legge avente densità rispetto alla Lebesgue (1-dim.) data da $f^\theta(x) := (\theta+1) x^\theta \mathbb{1}_{(0,1)}(x)$ $\forall \theta \in \underset{(-1,\infty)}{\Theta}$ e $\forall x \in \mathbb{R}$. Dunque $(\Omega, \mathcal{F}) = ((0,1)^n, \mathcal{B}((0,1)^n))$, dominante di Lebesgue (n-dim.), $X_i((x_1, \dots, x_n)) = x_i$, verosimiglianza ($\forall \theta > -1$ e $\forall x_i \in (0,1)$)

$$L^\theta(x) = L^\theta((x_1, \dots, x_n)) = (\theta+1)^n \left(\prod_{i=1}^{n} x_i\right)^\theta =$$

$$= (\theta+1)^n \exp\left(\theta \cdot \sum_{i=1}^{n} \log(x_i)\right)$$: per ciò, il modello è esponenziale se $\boxed{T := \sum_{i=1}^{n} \log(X_i)}$. (Sapevamo più che era statistica esaustiva!)

EX. / OSS. ~~RIASSUNTINO~~ / RIASSUNTINO. Consideriamo "il" modello statistico $(\Omega, \mathcal{F}, [P^\theta | \theta \in \Theta])$ "prodotto" canonicamente associato ad un campione $(X_1, \dots, X_n)$ di una certa taglia $n \in \mathbb{N}_*$ e di legge μ^θ dominata da una misura σ-finita ν su $\mathcal{B}(\mathbb{R})$ in modo che, per ogni $\theta \in \Theta$, le μ^θ siano mutuamente equivalenti (e ν). In pratica, stiamo parlando

(30)

$(\Omega, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $P^\theta = (\mu^\theta)^{\otimes n}$ (su $\mathcal{F}$) , $\mu = \nu^{\otimes n}$ (su $\mathcal{F}$) chiaramente, e $X_i((x_1, \dots, x_n)) = x_i$. In particolare, anche $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ è regolare, e che verosimiglianza

$L^\theta((x_1, \dots, x_n)) = \prod_{i=1}^{n} \frac{d\mu^\theta}{d\nu}(x_i)$. Supponiamo adesso che, in realtà, l'insieme dei parametri Θ $(\neq \emptyset)$ sia $\subseteq \mathbb{R}^k$ per un certo $k \in \mathbb{N}_+$ (che fissiamo), e che anzi sia aperto e convesso in $\mathbb{R}^k$. Supponiamo pure qualcosa di più: che esiste una funzione $S : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ misurabile tale che, per ogni $\theta \in \Theta$, risulti

$$0 \underset{\substack{\uparrow \\ \text{(invece che} \\ \text{"}0 \leq\text{")}}}{<} \int_{-\infty}^{+\infty} \exp\left(\langle \theta ; S(x) \rangle_{\mathbb{R}^k, 2}\right) d\nu(x) < +\infty \quad , \text{ e}$$

che allora risulti anche $\boxed{\frac{d\mu^\theta}{d\nu}(x) = B(\theta) \exp(\langle \theta ; S(x) \rangle)}$ per un'applicazione $B(\theta) : \Theta \to (0, \infty)$, la quale fissi data come $B(\theta) = \left(\int_{-\infty}^{\infty} \exp(\langle \theta ; S(x) \rangle) d\nu(x) \right)^{-1}$ ($\leftarrow$ reciproco), essendo μ^θ una legge di probabilità $\forall \theta \in \Theta$.

Dunque, il modello "fattore" è di tipo esponenziale con statistica esaustiva S. Allora (come anche già visto) il modello "prodotto" $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$ è anch'esso di tipo esponenziale con statistica esaustiva $\boxed{T := \sum_{i=1}^{n} S(X_i)}$ (precisamente $\underset{((\Omega, \mathcal{F}))}{(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))} \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$) : infatti,

$$\forall \theta \in \Theta, \quad \int_{\mathbb{R}^n} \exp\big(\langle \theta ; \underbrace{T(x_1,\dots,x_n)}_{\left(=\sum_{i=1}^n S(x_i)\right)} \rangle\big)\, \underbrace{d\mu(x_1,\dots,x_n)}_{\left(= d\nu(x_1)\cdots d\nu(x_n)\right)} =$$

$$= \left(\int_{-\infty}^{+\infty} \exp\big(\langle \theta ; S(x) \rangle\big)\, d\nu(x) \right)^n \in (0,+\infty) \quad , \text{ ed inoltre}$$

$$\mathcal{L}^\theta(x_1,\dots,x_n) = \prod_{i=1}^n \frac{d\mu^\theta}{d\nu}(x_i) = \underbrace{(B(\theta))^n}_{(=: C(\theta))} \prod_{i=1}^n \exp\big(\langle \theta ; S(x_i) \rangle\big) =$$

$$= C(\theta) \exp\big(\langle \theta ; T(x_1,\dots,x_n) \rangle\big) \quad . \quad \checkmark\checkmark$$

$\boxed{\text{Adesso}}$ la questione è la seguente: "quando" possiamo essere certi che la misura immagine $T(\mu)$ (su $\mathcal{B}(\mathbb{R}^k)$) sia σ-finita? ($\Leftarrow$ completa)

Per esempio, dato che la mis. di Lebesgue k-dim. è ovviamente σ-finita, se ~~...~~ T avesse "densità" rispetto alla Lebesgue k-dim. ("partendo da μ"), e cioè se esistesse una funzione misurabile

$f : \mathbb{R}^k \to [0,\infty)$ $\left(\text{con } \int_{\mathbb{R}^k} f(x)\,dx = \mu(\mathbb{R}^n)\right)$ tale che,

$$\forall A \in \mathcal{B}(\mathbb{R}^k), \quad \int_A f(x)\,dx = T(\mu)[A] = \mu[T^{-1}(A)] \quad ,$$

o più brevemente $dT(\mu) = f \cdot dx$, allora $T(\mu)$ sarebbe (banalmente) σ-finita e meglio ancora.

Ma $T = \sum_{i=1}^n S(X_i)$, e poiché $S(X_1),\dots,S(X_n)$ sono indipendenti: grazie ad una formula di CONVOLUZIONE, basterà dunque che ognuna fra $S(X_1),\dots,S(X_n)$ ammetta (UNA SOLA, equiv.) densità rispetto alla Leb. k-dim. ("partendo da μ"). $\checkmark\checkmark$

Torniamo all'ES. precedente, dove in particolare $T=\sum_{i=1}^{n}\log(X_i)$ e X_i ha legge μ^ϑ di densità rispetto alla Leb. 1-dim. data da "$f^\vartheta(x) = (\vartheta+1)x^\vartheta$, su $(0,1)$". Allora, per quanto appreso or ora, T è **Completa** se $\log(X_i)$ ammette densità rispetto alla Leb 1-dim. (prodotto della Leb n-dim. su $\mathcal{B}((0,1)^n)$).

Ma questo, in effetti, è vero, in quanto $X_i : \Omega \to \mathbb{R}$ con densità concentrata su un aperto di $\mathbb{R}_+$ ($(0,1)$), e

$\log : (0,1) \subseteq (0,+\infty) \to \mathbb{R}$ (anzi nell'aperto $\log((0,1))$) è un diffeomorfismo sull'immagine !! (Ci riferiamo col un teorema fondamentale sulla densità di una trasformazione diffeomorfa di una v.a. (anche multidimensionale) con densità.)

In definitiva, ogni $\log(X_i)$ ammette veramente densità, e quindi anche T, la quale per ciò risulta **Completa**.

($\boxed{\text{NOTA}}$ Quanto stiamo dicendo vale anche per questi tutti gli ES. precedenti a questo!)

($\boxed{\text{EX.}}$ Dimostrare che lo "stimatore" $U := -\frac{T}{n}$ (statistica), il quale sta in $\bigcap_{\vartheta>-1} \mathcal{L}^2(P^\vartheta)$ semplicemente perché qui X_i è limitata $(0<X_i\leq 1)$ ($\forall\vartheta\in\Theta$, $\forall M>0$, $P^\vartheta[|X_i|\leq M] =$
$= P^\vartheta[X_i \in [-M,M]] = \int_{-M}^{M} (\vartheta+1)x^\vartheta 1_{(0,1)}(x)\,dx = (\vartheta+1)\int_0^{M\wedge 1} x^\vartheta dx$, $= 1$ per $M \uparrow \infty$), stima correttamente $\frac{1}{\vartheta+1}$, e perciò è ottimale fra tutti gli stimatori (corretti) di $\frac{1}{\vartheta+1}$ in $\bigcap_{\vartheta>-1}\mathcal{L}^2(P^\vartheta)$.

($\boxed{\text{NOTA}}$ $\{\frac{1}{\vartheta+1} \mid \vartheta>-1\} = (0,\infty)$, $= U(\Omega)$ $(=\{-\log(x) \mid x\in(0,1)\})$.)

[Basta ottenere che, $\forall\vartheta\in\Theta$, $E^{P^\vartheta}[T] = -\frac{n}{\vartheta+1}$, ossia la

Correttezza, in quanto poi l'ottimalità è grazie alla (esercizio e)

Completezza della U (della T) e, quindi, ad un corollario del teorema di Blackwell–Rao. (È un calcolo :

$\forall \vartheta \in \Theta$ $(\vartheta > -1)$, $\mathbb{E}^{P^\vartheta}[T] = \sum_{i=1}^{n} \mathbb{E}^{P^\vartheta}[\log(X_i)] = n\, \mathbb{E}^{P^\vartheta}[\log(X_1)] =$

$= n(\vartheta+1)\int_0^1 \log(x)\, x^\vartheta\, dx$, ed ora (int. parti)

$$\int_0^1 \log(x)\, x^\vartheta\, dx = \underbrace{\left[\log(x)\frac{x^{\vartheta+1}}{\vartheta+1}\right]_{x=0}^{x=1}}_{=0} - \int_0^1 \underbrace{\frac{1}{x}\frac{x^{\vartheta+1}}{\vartheta+1}}_{\frac{1}{\vartheta+1}\cdot x^\vartheta}\, dx =$$

$$= -\frac{1}{\vartheta+1}\int_0^1 x^\vartheta\, dx = -\frac{1}{\vartheta+1}\left[\frac{x^{\vartheta+1}}{\vartheta+1}\right]_{x=0}^{x=1} = -\frac{1}{(\vartheta+1)^2},$$

da cui effetto $\mathbb{E}^{P^\vartheta}[T] = -\dfrac{n}{\vartheta+1}$. ✓]

⚠ Per quanto sia OVVIO, NON è assolutamente vero che, se $(\forall \vartheta \in \Theta)$ $\mathbb{E}^{P^\vartheta}[U] = \frac{1}{\vartheta+1}$, allora $\mathbb{E}^{P^\vartheta}[\frac{1}{U}] = \vartheta+1$ (o, cioè, $\mathbb{E}^{P^\vartheta}[\frac{1-U}{U}] = \vartheta$) !! Detto altrimenti, è falso che, presa $Y : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ v.a.r. integrabile $\neq$ ovunque nulla, valutata (NON costante) $\frac{1}{Y}$ integrabile con $\mathbb{E}^P[\frac{1}{Y}] = \frac{1}{\mathbb{E}^P[Y]}$!! (1.000.000 di controesempi)

$\boxed{\text{EX.}}$ A riguardo del modello stat. prodotto $(\Omega, \mathcal{A}, (P^\vartheta)_{\vartheta \in \Theta})$ canonicamente associato ad un campione $(X_1, \dots, X_n)$ $(n \in \mathbb{N}_*)$ di legge Π_ϑ, $\vartheta > 0$, dimostrare che $\overline{X} = \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ è uno stimatore corretto di ϑ stesso, in $L^2(P^\vartheta)$ $\forall \vartheta \in \Theta$ $(\vartheta>0)$, e che è ottimale a tutti gli stimatori corretti di ϑ in $\bigcap_{\substack{\vartheta\in\Theta\\(\vartheta>0)}} L^2(P^\vartheta)$. ✓

[Consideriamo $\Omega = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, $\Theta = (0,+\infty)$, μ dominante concentrata su $\mathbb{N}^n$ che conta ivi i punti, $X_i((x_1,\dots,x_n)) = x_i$, e ovviamente

$$L^\vartheta(x_1,\dots,x_n) = \prod_{i=1}^{n}\left(e^{-\vartheta}\frac{\vartheta^{x_i}}{x_i!}\right) = e^{-n\vartheta}\frac{\vartheta^{\sum_{i=1}^n x_i}}{x_1!\cdots x_n!} \quad (\forall x_i \in \mathbb{N}).$$

Visto che $\vartheta^{\sum_{i=1}^n x_i} = \exp\left(\log(\vartheta)\cdot\sum_{i=1}^n x_i\right)$, risulta evidente che, cambiando $\Theta = (0,\infty)_\vartheta$ in $\Theta = \mathbb{R}_{\log(\vartheta)}$ e μ dominante in $\text{“}\frac{1}{x_1!\cdots x_n!}\mu((x_1,\dots,x_n))\text{”}$, il modello è esponenziale con statistica $\boxed{T = \sum_{i=1}^n X_i}$, la quale per ciò risulta esaustiva. Inoltre, avendo le X_i indipendenti e con densità, anche T ammette densità e per ciò $T(\mu)$ risulta σ-finita, e quindi (TEO.) T è anche completa. Equivalentemente, $\overline{X} = \frac{T}{n}$ è stat. esaustiva e completa. Inoltre $T \in \bigcap_{\substack{\log \vartheta\\ \vartheta>0}} L^2(P^\vartheta)$, perché le Poisson ammettono sempre momento secondo $(=\vartheta^2+\vartheta)$, per cui grazie a Blackwell-Rao basta trovare che, $\forall \vartheta \in \Theta$, $\mathbb{E}^{P^\vartheta}[T] = n\vartheta$ per concludere. Ma questo è ovvio, perché $T = X_1 + \dots + X_n$ e $X_i \sim \Pi_\vartheta$, per cui che speranza propria $=\vartheta$. □]

→ Passiamo finalmente al concetto di <u>informazione</u> "locale" ("puntuale") di un modello statistico dominato, con l'idea principale di avere un altro strumento per "riassumere" il modello che non sia una statistica esaustiva "da trovare": bensì un numero non-negativo, o più in generale una matrice quadrata simmetrica e semi-definita positiva, calcolabile facilmente per mezzo della sola verosimiglianza del modello. (Naturalmente, non per tutti i mod. dominati potremo fare i ragionamenti che faremo...)

ESEMPIO MOTIVANTE Sia $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$ un modello stat. esponenziale di verosimiglianza $L^\theta(\omega) = C(\theta)\exp(\langle \theta; T(\omega)\rangle_{\mathbb{R}^k}) \quad \forall (\theta,\omega) \in \Theta \times \Omega$.
Con questo intendiamo che sullo spazio mis. $(\Omega, \mathcal{A})$ esiste una misura σ-finita μ dominante le P^θ (e rispetto alla quale calcoliamo la verosimiglianza), e una statistica $T : (\Omega,\mathcal{A}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, tale che Θ sia un aperto convesso $\neq \emptyset$ di $\mathbb{R}^k$ tale che, $\forall \theta \in \Theta$, risulti

$$0 < \int_\Omega \exp\langle \theta; T(\omega)\rangle \, d\mu(\omega) = \int_{\mathbb{R}^k} \exp\langle \theta; t\rangle \, dT(\mu)(t) < +\infty$$

ed anche $C(\theta) > 0$, se effettivamente la verosimiglianza è della forma

$L^\theta(\omega) = C(\theta)\exp(\langle \theta; T(\omega)\rangle) \quad \forall (\theta,\omega) \in \Theta \times \Omega$. Supponiamo in più che la misura immagine $T(\mu)$ "resti" σ-finita su $\mathcal{B}(\mathbb{R}^k)$ (ad esempio, che $T(\mu)$ sia $\ll$ rispetto alla Lebesgue k-dim.). Allora sappiamo che il modello è regolare e che la statistica T è esaustiva, completa e che ammette momenti di ogni ordine. In effetti, posto $f(\theta) := \log\left(\int_\Omega \exp\langle \theta; T(\omega)\rangle \, d\mu(\omega)\right) \quad \forall \theta \in \Theta$, ove $C(\theta) = \left(\int_\Omega \exp(\langle\theta; T(\omega)\rangle)\, d\mu(\omega)\right)^{-1} = \exp(-\psi(\theta))$ $\forall \theta\in\Theta$,

ossia che $\boxed{L^\theta(\omega) = \exp\left(\langle\theta;T(\omega)\rangle - \psi(\theta)\right)}$ $\forall (\theta,\omega)\in\Theta\times\Omega$, e

se effetto $T(\mu)$ è σ-finita su $\mathcal{B}(\mathbb{R}^k)$ allora ψ è C^∞ su tutto Θ, con in particolare $\frac{\partial\psi}{\partial\theta_i}(\theta) = E^{P^\theta}[T_i]$ e

$$\frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}(\theta) = \mathrm{Cov}^{P^\theta}[T_i\,;T_j] \quad \forall i,j=1,\dots,k \text{ e } \forall\theta\in\Theta$$

(se $\theta \doteq (\theta_1,\dots,\theta_k)$ e se $T \doteq (T_1,\dots,T_k)$). Ecco che viene la

$\boxed{\text{IDEA}}$: usare la matrice $k\times k$, simmetrica e semi-def. positiva delle covarianze fra le componenti di T, per ogni $\theta\in\Theta$, come effettiva l'informazione relativa al modello nel "punto" θ, per ogni $\theta\in\Theta$: cioè considerare come informazione in θ la matrice $(I(\theta))_{i,j=1,\dots,k}$,

$$I(\theta)_{i,j} := \mathrm{Cov}^{P^\theta}[T_i\,;T_j] \quad \forall i,j=1,\dots,k,\ \theta\in\Theta.$$

(Infatti, le T_i riassumono completamente il modello !!)

? Come arrivarci dalla verosimiglianza $L^\theta(\omega)$?

Osserviamo che $\log L^\theta(\omega) = \langle\theta;T(\omega)\rangle - \psi(\theta)$ resta C^∞ su Θ ($\forall\omega\in\Omega$),

con $\boxed{\frac{\partial}{\partial\theta_i}\log L^\theta(\omega) = T_i(\omega) - \frac{\partial\psi}{\partial\theta_i}(\theta)}$ $\forall i=1,\dots,k$, $\forall(\theta,\omega)\in\Theta\times\Omega$,

e dunque con $-\frac{\partial^2\log L^\theta(\omega)}{\partial\theta_i\partial\theta_j} = \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}(\theta) = \mathrm{Cov}^{P^\theta}[T_i\,,T_j]$

$\forall i,j=1,\dots,k$, $\forall\theta\in\Theta$ e $\forall\omega\in\Omega$, col che è costante in $\omega\in\Omega$,

per cui avendo $E^{P^\theta}\left[-\frac{\partial^2\log L^\theta(\cdot)}{\partial\theta_i\partial\theta_j}\right] = \mathrm{Cov}^{P^\theta}[T_i\,;T_j]$ $\forall i,j=1,\dots,k$,

$\forall\theta\in\Theta$. √ In effetti, allo stesso risultato arriviamo subito calcolando

$E^{P^\theta}\left[\frac{\partial}{\partial\theta_i}\log L^\theta(\cdot)\,\frac{\partial}{\partial\theta_j}\log L^\theta(\cdot)\right]$, perché $\frac{\partial\psi}{\partial\theta_h}(\theta) = E^{P^\theta}[T_h]$.

→ In generale, c'interesserà poter calcolare il logaritmo (naturale) della verosimiglianza, quando il suo predicato e la sua natura lo permettano! Definito così il concetto d'informazione "locale" (in θ) legata al modello, vedremo (prossimamente) estendendo il concetto di informazione legata ad una statistica sul modello stesso (guardando l'"immagine" del modello mediante tale statistica): sempre "inferiori" a quella del modello, e massima per una statistica esaustiva, mentre nulla per una statistica libera! ✓

Procediamo!

$\boxed{\text{CONTESTO (STANDARD)}}$: sia $(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in\Theta\})$ un modello stat. regolare, dove $\Theta\subseteq\mathbb{R}^k$ è aperto $\neq\emptyset$, con misura σ-finita / probabilità dominante (equivalente) μ, e probabilità dominante privilegiata P_0. Dunque $L^\theta(\omega) = \frac{dP^\theta}{d\mu}(\omega)$ è >0 su tutto Ω per ogni $\theta\in\Theta$, e quindi possiamo calcolare $\log L^\theta(\cdot)$ per ogni $\theta\in\Theta$. ($\rightarrow$ Per ogni $\theta\in\Theta$, l'evento $A_\theta := \{\omega\in\Omega \mid L^\theta(\omega)>0\}$ è indipendente da θ perché, in queste ipotesi, $A_\theta=\Omega$ $\forall\theta\in\Theta$ ...!)$^{(*)}$ Supponiamo inoltre che, $\forall\omega\in\Omega$, $\theta\mapsto L^\theta(\omega)$ sia di classe C^2 su tutto Θ con "tutte le derivate che restano ecc.", cioè: con $L^\theta(\omega)$ e $\frac{\partial L^\theta(\omega)}{\partial\theta_i}$ in $L^2(P^\theta)$ e $\frac{\partial^2 L^\theta(\omega)}{\partial\theta_i\partial\theta_j}$ in $L^1(P^\theta)$ per ogni $\theta\in\Theta$ e $i,j=1,\dots,k$.

Allora, naturalmente, $\forall\omega\in\Omega$, anche $\theta\mapsto\log(L^\theta(\omega))$ è C^2 su Θ: supponiamo che pure $\log L^\theta(\omega)$, $\frac{\partial\log L^\theta(\omega)}{\partial\theta_i}$ stiano in $L^2(P^\theta)$, e $\frac{\partial^2\log L^\theta(\omega)}{\partial\theta_i\partial\theta_j}$ stiano in $L^1(P^\theta)$ per ogni $\theta\in\Theta$ e $i,j=1,\dots,k$.

$\boxed{\text{NOTA}}$ Vale $\frac{\partial}{\partial\theta_i}\log L^\theta(\omega) = \frac{1}{L^\theta(\omega)}\frac{\partial L^\theta(\omega)}{\partial\theta_i}$, e che

$$\frac{\partial^2 \log L^\theta(\omega)}{\partial\theta_i \partial\theta_j} = \frac{\partial}{\partial\theta_j}\left[\frac{1}{L^\theta(\omega)} \frac{\partial L^\theta(\omega)}{\partial\theta_i}\right] = \frac{1}{L^\theta(\omega)} \frac{\partial^2 L^\theta(\omega)}{\partial\theta_i \partial\theta_j} -$$

$$- \frac{1}{(L^\theta(\omega))^2} \frac{\partial L^\theta(\omega)}{\partial\theta_j} \frac{\partial L^\theta(\omega)}{\partial\theta_i}$$, per ogni $\theta\in\Theta$, $\omega\in\Omega$ e $i,j=1,\dots,k$.

Per ciò, se $\frac{\partial L^\theta(.)}{\partial\theta_i}$ e $\frac{\partial^2 L^\theta(.)}{\partial\theta_i\partial\theta_j}$ stanno ~~tutte~~ in $L^2(p^\theta)$ (e $L^1(p^\theta)$ risp.), $\theta\in\Theta$, e se ad esempio $\frac{1}{L^\theta(\omega)}$ è una v.a. in L^∞ $\forall\theta\in\Theta$ ($\partial L^\theta(\omega)$ limitata) ($\partial L^2(p^\theta)$), allora automaticamente $\frac{\partial \log L^\theta(.)}{\partial\theta_i}$ e $\frac{\partial^2 \log L^\theta(.)}{\partial\theta_i\partial\theta_j}$ stanno tutte in $L^2(p^\theta)$ e $L^1(p^\theta)$ rispettivamente.

OSS. Sempre nel supporre $L^\theta(.)$ e $\nabla_\theta L^\theta(.)^t$ in $L^2(p^\theta)$, $\forall\theta\in\Theta$, siamo sicuri della seguente formula d'integrazione sotto il segno d'integrale: per ogni $\theta\in\Theta$ ed $i=1,\dots,k$, e quale che sia $X(\omega)\in L^2(\mu)$ (statistica), vale

$$\frac{\partial}{\partial\theta_i}\int_\Omega L^\theta(\omega)X(\omega)\,d\mu(\omega) = \int_\Omega \frac{\partial L^\theta(\omega)}{\partial\theta_i}X(\omega)\,d\mu(\omega)$$, cioè

(usando $dp^\theta(\omega) = L^\theta(\omega)d\mu(\omega)$, o anche $d\mu(\omega) = \frac{1}{L^\theta(\omega)}dp^\theta(\omega)$)

$$\boxed{\frac{\partial}{\partial\theta_i}E^{p^\theta}[X] = E^{p^\theta}\Big[\underbrace{\frac{\partial \log L^\theta}{\partial\theta_i}}_{\left(\frac{1}{L^\theta}\frac{\partial L^\theta}{\partial\theta_i}\right)}X\Big]}$$.

In particolare, chiaramente, $E^\mu\left[\frac{\partial L^\theta}{\partial\theta_i}\right] = E^{p^\theta}\left[\frac{\partial\log L^\theta}{\partial\theta_i}\right] = 0$ per ogni $\theta\in\Theta$ ed $i=1,\dots,k$. (Detto altrimenti, le v.a. $\frac{\partial\log L^\theta}{\partial\theta_i}(.)$ sono tutte p^θ-centrate.)

▷ Sotto tutte queste ipotesi sul modello, definiamo "LA MATRICE DI INFORMAZIONE SECONDO FISHER DI TALE MODELLO NEL PUNTO $\theta\in\Theta$" come la matrice $(I(\theta)_{i,j})_{i,j=1,\dots,k}$, $\theta\in\Theta$, delle covarianze fra le v.a. in $L^2(p^\theta)$ $\frac{\partial\log L^\theta(.)}{\partial\theta_h}$, calcolate rispetto a p^θ: cioè,

$\forall \theta \in \Theta$ e $\forall i,j = 1,\dots,k$, $I(\theta)_{i,j} \doteq Cov^{P^\theta}\left[\frac{\partial \log L^\theta}{\partial \theta_i}, \frac{\partial \log L^\theta}{\partial \theta_j}\right] =$

$$= E^{P^\theta}\left[\frac{\partial \log L^\theta}{\partial \theta_i} \cdot \frac{\partial \log L^\theta}{\partial \theta_j}\right].$$

($\boxed{\text{NOTA}}$ $I(\theta)_{i,j} = E^{P^\theta}\left[\frac{1}{(L^\theta)^2} \frac{\partial L^\theta}{\partial \theta_i} \frac{\partial L^\theta}{\partial \theta_j}\right] = E^\mu\left[\frac{1}{L^\theta} \frac{\partial L^\theta}{\partial \theta_i} \frac{\partial L^\theta}{\partial \theta_j}\right]$.
($\Rightarrow$ è continua in θ, dunque è informazione LOCALE più che soltanto puntuale!))

Dunque, $(I(\theta)_{i,j})_{i,j=1,\dots,k}$ è la matrice delle covarianze del vettore aleatorio (in $\mathbb{R}^k$) $\nabla_\theta \log L^\theta(\cdot)$, e come tale è necessariamente simmetrica e semi-definita positiva.

NOTAZIONI / OSSERVAZIONI — Dato un vettore aleatorio in $\mathbb{R}^k$ $X \doteq (X_1,\dots,X_k)^t$ su $(\Omega,\mathcal{F},P)$ (sp. di probabilità), dove X_h sta in $L^2(P)$ per ogni $h=1,\dots,k$, indichiamo con

$$\text{“}Cov^P[X]\text{”} := \left(Cov^P[X_i, X_j]\right)_{i,j=1,\dots,k} =$$

$$= \left(E^P[X_i X_j] - E^P[X_i]E^P[X_j]\right)_{i,j=1,\dots,k}$$

la matrice $k \times k$ delle P-covarianze tra le X_h.

Ricordiamo, in particolare, che vale la formula

$$\forall x = (x_1,\dots,x_k)^t \in \mathbb{R}^k, \quad x^t Cov^P[X] x = Var^P\Big[\underbrace{\sum_{i=1}^k x_i X_i}_{(\langle X; x\rangle_{\mathbb{R}^k})}\Big] \overset{(imm.)}{=}$$

$$\overset{(imm.)}{=} E^P\left[\langle X - E^P[X]; x\rangle_{\mathbb{R}^k}^2\right].$$

($\Rightarrow$ Se X è centrato, allora $x^t Cov^P[X] x = E^P[\langle X; x\rangle^2]$ $\forall x \in \mathbb{R}^k$.)

Inoltre, se $Y := (Y_1,\dots,Y_k)^t$ è un altro vettore aleatorio su $\mathbb{R}^k$ su $(\Omega,\mathcal{F},P)$, in $L^2(P)$, e se X e Y sono vettori indipendenti, allora vale

$$\boxed{Cov^P[X+Y] = Cov^P[X] + Cov^P[Y]}.$$

[La tesi è che, $\forall i,j=1,\dots,k$, risulta $\mathbb{E}^P[(X_i+Y_i)(X_j+Y_j)] - \mathbb{E}^P[X_i+Y_i]\,\mathbb{E}^P[X_j+Y_j] = \mathbb{E}^P[X_iX_j] - \mathbb{E}^P[X_i]\,\mathbb{E}^P[X_j] + \mathbb{E}^P[Y_iY_j] - \mathbb{E}^P[Y_i]\,\mathbb{E}^P[Y_j]$, e ciò è possibile evidente ricordando che, per indipendenza, $\mathbb{E}^P[X_{h_1}Y_{h_2}] = \mathbb{E}^P[X_{h_1}]\,\mathbb{E}^P[X_{h_2}]$ $\forall h_1,h_2=1,\dots,k$.]

Dunque, da $\dfrac{\partial^2 \log L^\theta}{\partial\theta_i\partial\theta_j} = -\dfrac{1}{(L^\theta)^2}\dfrac{\partial L^\theta}{\partial\theta_i}\dfrac{\partial L^\theta}{\partial\theta_j} + \dfrac{1}{L^\theta}\dfrac{\partial^2 L^\theta}{\partial\theta_i\partial\theta_j}$ e da

$(I(\theta))_{i,j} = \mathbb{E}^{P^\theta}\left[\dfrac{1}{(L^\theta)^2}\dfrac{\partial L^\theta}{\partial\theta_i}\dfrac{\partial L^\theta}{\partial\theta_j}\right]$, deduciamo subito che

$$I(\theta)_{i,j} = \mathbb{E}^{P^\theta}\left[-\frac{\partial^2 \log L^\theta}{\partial\theta_i\partial\theta_j}\right] + \mathbb{E}^{P^\theta}\left[\frac{1}{L^\theta}\frac{\partial^2 L^\theta}{\partial\theta_i\partial\theta_j}\right].$$

Proposizione. All'interno del contesto statistico dove è possibile definire l'informazione secondo Fischer, e nelle notazioni precedenti, vale che

$$\boxed{I(\theta)_{i,j} = -\mathbb{E}^{P^\theta}\left[\frac{\partial^2 \log L^\theta}{\partial\theta_i\partial\theta_j}\right]} \quad \forall\theta\in\Theta,\ \forall i,j=1,\dots,k.$$

[Dim. Dimostriamo infatti che $\mathbb{E}^{P^\theta}\left[\dfrac{1}{L^\theta}\dfrac{\partial^2 L^\theta}{\partial\theta_i\partial\theta_j}\right] = 0$ ($\forall\theta\in\Theta$ e $\forall i,j=1,\dots,k$): semplicemente, $\mathbb{E}^{P^\theta}\left[\dfrac{1}{L^\theta}\dfrac{\partial^2 L^\theta}{\partial\theta_i\partial\theta_j}\right] = \mathbb{E}^{\mu}\Big[\underbrace{\dfrac{\partial^2 L^\theta}{\partial\theta_i\partial\theta_j}}_{(\in L^2(\mu))}\Big] =$

$$= \frac{\partial}{\partial\theta_j}\underbrace{\mathbb{E}^{\mu}\left[\frac{\partial L^\theta}{\partial\theta_i}\right]}_{(\equiv 0)} = 0 \quad . \qquad \square\,]$$

OSS. Consideriamo un modello stat. $(\Omega,\mathcal{F},\{P^\theta\mid\theta\in\Theta\})$ canonicamente associato ad un campione $(X_1,\dots,X_n)$ di taglia n (pre-scelta) e di legge μ^θ su $\mathcal{B}(\mathbb{R})$, $\theta\in\Theta$: dunque $(\Omega,\mathcal{F}) = (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$, $P^\theta = (\mu^\theta)^{\otimes n}$, e $X_i(x_1,\dots,x_n) = x_i$. Supponiamo che le μ^θ siano fra loro equivalenti, e sia ν una probabilità che le domina: allora il modello è regolare, e $\mu \doteq \nu^{\otimes n}$ è una probabilità dominante. Rispetto a questa, la verosimiglianza è $L^\theta(x_1,\dots,x_n) = \prod_{i=1}^{n} \frac{d\mu^\theta}{d\nu}(x_i)$. Ebbene, vogliamo ora convincerci del fatto che se i "modelli fattori" sono in "contesto standard" per poter definire il concetto di informazione secondo Fisher (in ogni $\theta\in\Theta$), allora così è per il modello prodotto ed inoltre le informazioni si sommano. Precisiamo: mettiamoci sul "modello fattore" $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\mu^\theta|\theta\in\Theta\}, \nu)$", supponiamo che Θ sia un aperto connesso $\neq\emptyset$ di $\mathbb{R}^k$, $k\in\mathbb{N}_*$ opportuno, mentre abbiamo già supposto il modello regolare con ν dominante, per cui $\frac{d\mu^\theta}{d\nu}(x)>0$ per ogni $\theta,x\in\Theta\times\mathbb{R}$; supponiamo quindi che $\theta\mapsto\frac{d\mu^\theta}{d\nu}(x)$ sia in $C^2(\Theta)$ per ogni $x\in\mathbb{R}$, e supponiamo che, $\forall\theta\in\Theta$ e $\forall i,j=1,\dots,k$, siano $\frac{d\mu^\theta}{d\nu}(\cdot)$ limitata, $\frac{\partial}{\partial\theta_i}\left(\frac{d\mu^\theta}{d\nu}(\cdot)\right)$ in $L^2(\mu^\theta)$, $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(\frac{d\mu^\theta}{d\nu}(\cdot)\right)$ in $L^1(\mu^\theta)$. In particolare, resta ben definito il $\log\left(\frac{d\mu^\theta}{d\nu}(x)\right)$ per ogni $(\theta,x)\in\Theta\times\mathbb{R}$, e per ogni $\theta\in\Theta$ $\log\left(\frac{d\mu^\theta}{d\nu}(\cdot)\right)$ è limitato, mentre $\forall\theta\in\Theta$ e $\forall i,j=1,\dots,k$, $\frac{\partial}{\partial\theta_i}\log\left(\frac{d\mu^\theta}{d\nu}(\cdot)\right)$ sta in $L^2(\mu^\theta)$ e $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\left(\frac{d\mu^\theta}{d\nu}(\cdot)\right)$ sta in $L^1(\mu^\theta)$.

Allora resta definita la matrice d'informazione secondo Fisher di $(\mathbb{R},\mathcal{B}(\mathbb{R}),\{\mu^\theta|\theta\in\Theta\},\nu)$ in ogni $\theta\in\Theta$, sia $(I_0(\theta)_{i,j})_{i,j=1,\dots,k}$, ed è data da

$$I_0(\theta)_{i,j} \doteq Cov^{\mu^\theta}\left[\frac{\partial}{\partial\theta_i}\log\frac{d\mu^\theta}{d\nu}, \frac{\partial}{\partial\theta_j}\log\frac{d\mu^\theta}{d\nu}\right] \overset{\text{(visto)}}{=}$$

$$= \int_{\mathbb{R}} \frac{\partial}{\partial\theta_i}\log\frac{d\mu^\theta}{d\nu}(x_1)\frac{\partial}{\partial\theta_j}\log\frac{d\mu^\theta}{d\nu}(x_1)\,d\mu^\theta(x_1) \overset{\text{(INTEGR. RISP. IMMAGINE)}}{=} \int_{\mathbb{R}^n}\frac{\partial}{\partial\theta_i}\log\frac{d\mu^\theta}{d\nu}(X_h(\omega))\frac{\partial}{\partial\theta_j}\log\frac{d\mu^\theta}{d\nu}(X_h(\omega))\,dP^\theta(\omega) =$$

$$= Cov^{P^\theta}\left[\frac{\partial}{\partial\theta_i}\left(\log\frac{d\mu^\theta}{d\nu}(X_h)\right), \frac{\partial}{\partial\theta_j}\left(\log\frac{d\mu^\theta}{d\nu}(X_h)\right)\right]$$ per ogni $h=1,\dots,n$ (è

veramente una convenzione, risulta quella $\int_{\mathbb{R}^n} \frac{\partial}{\partial \theta_i}\left(\log \frac{d\mu^\theta}{d\nu}(X_h(\omega))\right) dP^\theta(\omega) =$

$= \int_{\mathbb{R}} \frac{\partial}{\partial \theta_i}\left(\log \frac{d\mu^\theta}{d\nu}(x)\right) d\mu^\theta(x) \overset{(\text{visto})}{=} 0 \quad \forall \theta \in \Theta, \forall i=1,\dots,k, \forall h=1,\dots,m)$.

Scritto in modo conciso, pertanto, è $\boxed{I_0(\theta) = \mathrm{Cov}^{P^\theta}\left[\nabla_\theta \log \frac{d\mu^\theta}{d\nu}(X_h)\right]}$

per ogni $\theta \in \Theta$ e $h=1,\dots,m$. Tornando quindi al nostro modello "prodotto" $(\Omega, \mathcal{A}, \{P^\theta | \theta \in \Theta\}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mu^{\theta \otimes n} | \theta \in \Theta\}, \nu^{\otimes n})$

di verosimiglianza $\mathcal{L}^\theta(x_1,\dots,x_n) = \prod_{h=1}^{m} \frac{d\mu^\theta}{d\nu}(X_h(x_1,\dots,x_n)) \quad \left(= \prod_{h=1}^{m} \frac{d\mu^\theta}{d\nu}(x_h)\right)$,

è chiaro che: $\mathcal{L}^\theta(\omega) > 0 \quad \forall (\theta,\omega) \in \Theta \times \Omega$, che $\theta \mapsto \mathcal{L}^\theta(\omega)$ sia $C^2(\Theta)$

$\forall \omega \in \Omega$ e limitata per ogni $\theta \in \Theta$, e che, $\forall \theta \in \Theta$ e $\forall i,j=1,\dots,k$,

$\frac{\partial \mathcal{L}^\theta(\cdot)}{\partial \theta_i}$ e $\frac{\partial^2 \mathcal{L}^\theta(\cdot)}{\partial \theta_i \partial \theta_j}$ stanno in $L^2(P^\theta)$ e $L^1(P^\theta)$ rispettivamente.

Allora resta definita $I(\theta) = \mathrm{Cov}^{P^\theta}[\nabla_\theta \log \mathcal{L}^\theta] \quad \forall \theta \in \Theta$, come ormai ben sappiamo. Ma, adesso, osserviamo che, $\forall (\theta,\omega) \in \Theta \times \Omega$,

$$\log \mathcal{L}^\theta(\omega) = \sum_{h=1}^{m} \log \frac{d\mu^\theta}{d\nu}(X_h(\omega))$$, da cui, $\forall i=1,\dots,k$,

$$\frac{\partial}{\partial \theta_i} \log \mathcal{L}^\theta(\omega) = \sum_{h=1}^{m} \frac{\partial}{\partial \theta_i} \log \frac{d\mu^\theta}{d\nu}(X_h(\omega))$$, cioè

$$\boxed{\nabla_\theta \log \mathcal{L}^\theta(\omega) = \sum_{h=1}^{m} \nabla_\theta \log \frac{d\mu^\theta}{d\nu}(X_h(\omega))}$$, da cui (per

indipendenza delle X_h) $\underset{(\text{visto})}{}$ $\mathrm{Cov}^{P^\theta}[\nabla_\theta \log \mathcal{L}^\theta] = \sum_{h=1}^{m} \mathrm{Cov}^{P^\theta}\left[\nabla_\theta \log \frac{d\mu^\theta}{d\nu}(X_h)\right]$

per ogni $\theta \in \Theta$, e cioè, $\boxed{\forall \theta \in \Theta, I(\theta) = m I_0(\theta)}$. //

($\boxed{\text{NOTA}}$ "$\log(\prod) = \sum(\log)$" e "$\nabla(\sum) = \sum(\nabla)$" sono a favore della bontà della matrice di Fisher come matrice di informazione!) (risultato fondamentale)

(ES.) Mettiamoci nel modello statistico canonico per un campione $(X_1,\dots,X_n)$ di legge μ^θ esponenziale $\mathcal{E}(\theta)$, $\theta > 0$: dunque,

$(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in\Theta\}, \mu) = \left((0,+\infty)^n, \mathcal{B}((0,+\infty)^n), \{\mu^{\theta\,\otimes n} \mid \theta \in (0,+\infty)\}, \nu^{\otimes n}\right)$,

dove ν è di Lebesgue su $\mathcal{B}((0,+\infty))$ e $\frac{d\mu^\theta}{d\nu}(x) = \theta e^{-\theta x}$ $\forall \theta>0$ e $x>0$, effettivamente

e $X_i(x_1,\dots,x_n)=x_i$. Allora il modello fattore è esponenziale con statistica canonica $-X$, X l'identità su $(0,+\infty)$, anche completa perché avente densità (SOLITO TEOREMA),

e dunque, nelle notazioni precedenti, $I_0(\theta) = \mathrm{Var}^{\mu^\theta}[-X] = \frac{1}{\theta^2}$ $\forall\theta>0$, da cui $I(\theta) = \frac{n}{\theta^2}$ $\forall \theta>0$. ✓

(ES.) Consideriamo invece $\left((0,1)^n, \mathcal{B}((0,1)^n), \{\mu^{\theta\,\otimes n} \mid \theta\in(-1,+\infty)\}, \nu^{\otimes n}\right)$ con ν di Lebesgue su $\mathcal{B}((0,1))$ e $\frac{d\mu^\theta}{d\nu}(x) = (\theta+1)x^\theta$ $\forall\theta>-1$ e $\forall x\in(0,1)$, e $X_i(x_1,\dots,x_n)=x_i$. Allora il modello fattore è esponenziale con statistica $\log(X)$, X l'identità su $(0,1)$, che ha densità e dunque $I_0(\theta) = \mathrm{Var}^{\mu^\theta}[\log(X)]$, $\theta>-1$, è calcolabile. ✓

(ES.) Mettiamoci nel modello stat. canonico per un campione $(X_1,\dots,X_n)$ di legge μ^θ $N(m,\sigma^2)$ su $\mathbb{R}$, $\theta=(m,\sigma^2)\in\mathbb{R}\times(0,+\infty)$: cioè, $(\Omega,\mathcal{A},\{P^\theta\mid\theta\in\Theta\},\mu) = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mu^{\theta\,\otimes n} \mid \theta=(m,\sigma^2)\in\mathbb{R}\times(0,+\infty)\}, \nu^{\otimes n}\right)$ dove ν è di Lebesgue 1-dim. e $\frac{d\mu^\theta}{d\nu}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$ per ogni $\theta=(m,\sigma^2)\in\mathbb{R}\times(0,+\infty)$ e $x\in\mathbb{R}$, e poi $X_i(x_1,\dots,x_n)=x_i$. Col mantenere $\Theta=\mathbb{R}\times(0,+\infty)$ (comunque aperto $(\neq\emptyset)$ in $\mathbb{R}^2$), il modello

NON è esponenziale e allora verifichiamo "a mano" che l'informazione secondo Fisher è ben definita, e calcoliamola. Ebbene, $\frac{d\mu^\theta}{d\nu}(x)$ è >0 "sempre", è limitata su $\mathbb{R}$ per ogni $\theta\in\Theta$, ~~e~~ e

$\frac{d\mu^\theta}{d\nu}(x) = \exp\left(-\frac{1}{2\sigma^2}(x-m)^2 - \log\sqrt{\sigma^2} - \log\sqrt{2\pi}\right)$ e quindi che

$(m,\sigma^2)\mapsto\frac{d\mu^{(m,\sigma^2)}}{d\nu}(x)$ in $C^2(\Theta)$ per ogni $x\in\mathbb{R}$, con

$\frac{\partial}{\partial m}\left(\frac{d\mu^\theta}{d\nu}(x)\right) = \frac{d\mu^\theta}{d\nu}(x) \cdot \frac{1}{\sigma^2}(x-m)$ e

$\frac{\partial}{\partial \sigma^2}\left(\frac{d\mu^\theta}{d\nu}(x)\right) = \frac{d\mu^\theta}{d\nu}(x) \cdot \left(\frac{1}{2\sigma^4}(x-m)^2 - \frac{1}{2\sigma^2}\right)$ chiaramente in $L^2(\nu)$,

così come analogamente per $\frac{\partial^2}{\partial m^2}\frac{d\mu^\theta}{d\nu}(\cdot)$, $\frac{\partial^2}{\partial m \partial \sigma^2}\frac{d\mu^\theta}{d\nu}(\cdot)$ e $\frac{\partial^2}{\partial (\sigma^2)^2}\frac{d\mu^\theta}{d\nu}(\cdot)$,

pertanto in effetti restano ben definite le matrici 2×2 d'informazione secondo Fisher in ogni $\theta \in \Theta$ sia del modello fattore sia del modello prodotto, ed indicandole con $I_0(\theta)$ e $I(\theta)$ rispettivamente, vale $I(\theta) = n I_0(\theta)$ e

$$-I_0(\theta) \overset{(\text{TEO.})}{=} \begin{bmatrix} \mathbb{E}^{\mu^\theta}\left[\frac{\partial^2}{\partial m^2}\log\frac{d\mu^\theta}{d\nu}\right] & \mathbb{E}^{\mu^\theta}\left[\frac{\partial^2}{\partial m \partial \sigma^2}\log\frac{d\mu^\theta}{d\nu}\right] \\ * & \mathbb{E}^{\mu^\theta}\left[\frac{\partial^2}{\partial (\sigma^2)^2}\log\frac{d\mu^\theta}{d\nu}\right] \end{bmatrix} \text{ per ogni } \theta \in \Theta .$$

Si tratta dunque di calcolare un po' di derivate ... Ora $\log\frac{d\mu^\theta}{d\nu}(x) =$

$= -\frac{1}{2\sigma^2}(x-m)^2 - \log\sqrt{\sigma^2} - \log\sqrt{2\pi}$, per cui $\frac{\partial}{\partial m}\log\frac{d\mu^\theta}{d\nu}(x) =$

$= \frac{1}{\sigma^2}(x-m)$ e $\frac{\partial}{\partial \sigma^2}\log\frac{d\mu^\theta}{d\nu}(x) = \frac{1}{2\sigma^4}(x-m)^2 - \frac{1}{2\sigma^2}$ già calcolate

(in effetti, $\frac{\partial}{\partial \theta_i}\log\frac{d\mu^\theta}{d\nu}(x) = \frac{1}{d\mu^\theta/d\nu(x)}\frac{\partial}{\partial \theta_i}\frac{d\mu^\theta}{d\nu}(x)$), ed intanto

vediamo subito che $\frac{\partial^2}{\partial m^2}\log\frac{d\mu^\theta}{d\nu}(x) \equiv -\frac{1}{\sigma^2}$, che ha $-\mathbb{E}^{\mu^\theta}[\ldots] =$

$= 1/\sigma^2$! ✓ Poi, $\frac{\partial^2}{\partial m \partial \sigma^2}\log\frac{d\mu^\theta}{d\nu}(x) = -\frac{1}{\sigma^4}(x-m)$ e allora che

$-\mathbb{E}^{\mu^\theta}[\ldots] = \frac{1}{\sigma^4}\mathbb{E}^{\mu^\theta}[x-m] = 0$ (perché $\mathbb{E}^{\mu^\theta}[x] = m$) ! ✓

Dunque, $\frac{\partial^2}{\partial (\sigma^2)^2}\log\frac{d\mu^\theta}{d\nu}(x) \overset{(\text{imm.})}{=} \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-m)^2$ e quindi che

$-\mathbb{E}^{\mu^\theta}[\ldots] = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}\underbrace{\mathbb{E}^{\mu^\theta}[(x-m)^2]}_{(=\sigma^2!)} = \frac{1}{2\sigma^4}$! ✓ In conclusione,

$$I(m,\sigma^2) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix} \quad \forall n \in \mathbb{N}_+ , \cancel{\forall m \in \mathbb{R}} \text{ e } \forall \sigma^2 \in (0,+\infty) . \quad //$$

EX. Verificare che $I_0(\sigma) = \frac{2}{\sigma^2}$, da cui $I(\sigma) = \frac{2n}{\sigma^2}$.

[Anzitutto, osserviamo che $\frac{\partial}{\partial\sigma}\frac{d\mu^\vartheta}{d\lambda}(x) = \frac{d\mu^\vartheta}{d\lambda}(x)\cdot\left(\frac{1}{\sigma^3}(x-m)^2 - \frac{1}{\sigma}\right)$, da cui di nuovo $I_0(\sigma)$ e $I(\sigma)$ hanno perfettamente senso. Inoltre,

$I_0(\sigma) = \mathbb{E}^{\mu^\vartheta}\left[-\frac{\partial^2}{\partial\sigma^2}\log\frac{d\mu^\vartheta}{d\lambda}(\cdot)\right]$, ed essendo $\frac{\partial}{\partial\sigma}\log\frac{d\mu^\vartheta}{d\lambda}(x) =$

$= \frac{1}{\sigma^3}(x-m)^2 - \frac{1}{\sigma}$ si ha $\frac{\partial^2}{\partial\sigma^2}\log\frac{d\mu^\vartheta}{d\lambda}(x) = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x-m)^2$,

che conseguentemente dà $-\mathbb{E}^{\mu^\vartheta}[\dots] = -\frac{1}{\sigma^2} + \frac{3}{\sigma^2} = \frac{2}{\sigma^2}$. ✓]

(?) A questo punto, come ottenere informazioni sul modello a partire da una statistica su di esso? Ovvero, come estrarre informazione da una statistica? Risposta: a partire dal relativo "modello immagine"! Vediamo.

Sia $(\Omega, \mathcal{A}, \{P^\vartheta | \vartheta\in\Theta\}, \mu)$ un modello stat. dominato, sia $(E, \mathcal{E})$ uno spazio misurabile, e sia $T : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ una statistica. Allora "IL MODELLO STAT. IMMAGINE DI $(\Omega, \mathcal{A}, \{P^\vartheta | \vartheta\in\Theta\})$ MEDIANTE T" è

$(E, \mathcal{E}, \{T(P^\vartheta) | \vartheta\in\Theta\})$, dove come sempre $T(P^\vartheta)[A] =$

$= P^\vartheta[T^{-1}(A)]$ $\forall \vartheta\in\Theta$ e $\forall A\in\mathcal{E}$ (e sono tutte probabilità).

→ Supponiamo che $T(\mu)$ sia σ-finita su $\mathcal{E}$, per cui $T(\mu)$ domina le $T(P^\vartheta)$, $\vartheta\in\Theta$, ed è una probabilità se μ lo è.

[Dominaz.: $\forall A\in\mathcal{E}$, se $T(\mu)[A] = 0$, allora $T(P^\vartheta)[A] = 0$ perché $0 = T(\mu)[A] = \mu(T^{-1}(A))$
$\overset{\mu \text{ domina}}{\Rightarrow} P^\vartheta(T^{-1}(A)) = 0$, cioè appunto $T(P^\vartheta)[A] = 0$. ✓]

Introduciamo dunque $\left(L^\vartheta(\omega) = \frac{dP^\vartheta}{d\mu}(\omega) \text{ e}\right)$ $\boxed{L^\vartheta_T(x) \doteq \frac{dT(P^\vartheta)}{dT(\mu)}(x)}$ per ogni $\vartheta\in\Theta$ e $x\in E$, e mettiamoci nelle condizioni "standard" per questi due modelli dominati affinché ammettano le rispettive matrici di informazione secondo Fischer in ogni $\vartheta \in \Theta$, siano $I(\vartheta) = (I(\vartheta)_{i,j})_{i,j}$ e $I_T(\vartheta) = (I_T(\vartheta)_{i,j})_{i,j}$: dunque, supponiamo che

- esiste $k \in \mathbb{N}_+$ tale che Θ sia un aperto $(\neq \emptyset)$ di $\mathbb{R}^k$ connesso;
- $L^\vartheta(\omega) > 0$, $\vartheta \mapsto L^\vartheta(\omega)$ in $C^2(\Theta)$ $\forall \omega \in \Omega$, poi $L^\vartheta(\cdot)$ e $\frac{\partial}{\partial \vartheta_i} L^\vartheta(\cdot)$ in $L^2(p^\vartheta)$ e $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} L^\vartheta(\cdot)$ in $L^1(p^\vartheta)$ $\forall \vartheta \in \Theta$, $i,j = 1,\dots,k$, e anche $\log L^\vartheta(\cdot)$ e $\frac{\partial}{\partial \vartheta_i} \log L^\vartheta(\cdot)$ in $L^2(p^\vartheta)$ mentre $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log L^\vartheta(\cdot)$ in $L^1(p^\vartheta)$ $\forall \vartheta \in \Theta$ e $\forall i,j = 1,\dots,k$;
- $L^\vartheta_T(x) > 0$, $\vartheta \mapsto L^\vartheta_T(x)$ in $C^2(\Theta)$ $\forall x \in E$, poi $L^\vartheta_T(\cdot)$ e $\log L^\vartheta_T(\cdot)$ e $\frac{\partial}{\partial \vartheta_i} L^\vartheta_T(\cdot)$ e $\frac{\partial}{\partial \vartheta_i} \log L^\vartheta_T(\cdot)$ in $L^2(T(p^\vartheta))$ mentre $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} L^\vartheta_T(\cdot)$ e $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log L^\vartheta_T(\cdot)$ in $L^1(T(p^\vartheta))$, per ogni $\vartheta \in \Theta$ e $i,j = 1,\dots,k$.

Per ciò, le v.a. $\frac{\partial}{\partial \vartheta_h} \log L^\vartheta(\cdot)$ e $\frac{\partial}{\partial \vartheta_h} \log L^\vartheta_T(\cdot)$ di $L^2(p^\vartheta)$ e $L^2(T(p^\vartheta))$ risp. sono p^ϑ-centrate e $T(p^\vartheta)$-centrate risp., per ogni $\vartheta \in \Theta$ e $h = 1,\dots,k$, e restano ben definite le matrici di informazione secondo Fischer in $\vartheta \in \Theta$ dei due modelli, rispettivamente

$$\begin{cases} I(\vartheta) = Cov^{p^\vartheta}[\nabla_\vartheta \log L^\vartheta(\cdot)], \text{ e} \\ I_T(\vartheta) = Cov^{T(p^\vartheta)}[\nabla_\vartheta \log L^\vartheta_T(\cdot)] \ (\forall \vartheta \in \Theta), \end{cases}$$ ossia

$$\begin{cases} I(\vartheta)_{i,j} = E^{p^\vartheta}\left[\frac{\partial}{\partial \vartheta_i} \log L^\vartheta \frac{\partial}{\partial \vartheta_j} \log L^\vartheta\right] = -E^{p^\vartheta}\left[\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log L^\vartheta\right], \text{ e} \\ I_T(\vartheta)_{i,j} = E^{T(p^\vartheta)}\left[\frac{\partial}{\partial \vartheta_i} \log L^\vartheta_T \frac{\partial}{\partial \vartheta_j} \log L^\vartheta_T\right] \overset{(\text{Prop.})}{=} -E^{T(p^\vartheta)}\left[\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log L^\vartheta_T\right] \ (\forall \vartheta \in \Theta, \forall i,j = 1,\dots,k) \end{cases}$$

In particolare, anche $I_T(\vartheta)$ è semi-definita positiva $\forall \vartheta \in \Theta$, o così in simbolo "$I_T(\vartheta) \geq 0$" $\forall \vartheta \in \Theta$. Ebbene, vogliamo verificare che $\boxed{I_T(\vartheta) \leq I(\vartheta)}$ $\forall \vartheta \in \Theta$, nel senso che $I(\vartheta) - I_T(\vartheta)$ sia semi-def. positiva $\forall \vartheta \in \Theta$, e che $I_T(\vartheta) \equiv I(\vartheta)$ se e solo se T è crescente per il modello (dai fattori), mentre ovviamente $I_T(\vartheta)\equiv 0$ se e solo se T è libera dal modello (dai fattori).

NOTA Per quanto sopra visto, rimarchiamo solo che $\mathcal{L}_T^\vartheta(T(\omega)) > 0$ "sempre", $\vartheta \mapsto \mathcal{L}_T^\vartheta(T(\omega))$ in $C^2(\Theta)$ $\forall \omega\in\Omega$, così $\mathcal{L}_T^\vartheta(T(\cdot))$ e $\log\mathcal{L}_T^\vartheta(T(\cdot))$ e $\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}_T^\vartheta(T(\cdot))$ e $\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}_T^\vartheta(T(\cdot))$ sono in $L^2(P^\vartheta)$, mentre $\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\mathcal{L}_T^\vartheta(T(\cdot))$ e $\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\log\mathcal{L}_T^\vartheta(T(\cdot))$ sono in $L^1(P^\vartheta)$, per ogni $\vartheta\in\Theta$ e $i,j=1,\dots,k$.
Inoltre, in effetti, $\underset{(\mu\text{-q.o.})}{I_T(\vartheta)} = \mathrm{Cov}^{P^\vartheta}[\nabla_\vartheta \log\mathcal{L}_T^\vartheta(T)]$!

OSS. (Generalizzazione) $\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}_T^\vartheta(T(\cdot)) = E^{P^\vartheta}\left[\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}^\vartheta \,\middle|\, T\right](\cdot)$ per ogni $\vartheta\in\Theta$ e $i=1,\dots,k$, o brevemente $\nabla_\vartheta\log\mathcal{L}_T^\vartheta(T) = E^{P^\vartheta}[\nabla_\vartheta\log\mathcal{L}^\vartheta \mid T]$.

Dim. Chiaramente, la tesi è che $E^{P^\vartheta}\left[\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}^\vartheta\, 1_{T^{-1}(A)}\right] =$

$= E^{P^\vartheta}\left[\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}_T^\vartheta(T)\, 1_{T^{-1}(A)}\right]$ per ogni $\vartheta\in\Theta$, $i=1,\dots,k$ e $A\in\mathcal{E}$. Ma, infatti, $E^{P^\vartheta}\Big[\underbrace{\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}^\vartheta}_{\left(\frac{\partial}{\partial\vartheta_i}\mathcal{L}^\vartheta\cdot\frac{1}{\mathcal{L}^\vartheta}\right)} 1_{T^{-1}(A)}\Big] = \text{"}E^\mu\text{"}\left[\frac{\partial}{\partial\vartheta_i}\mathcal{L}^\vartheta 1_{T^{-1}(A)}\right] \overset{(\text{visto})}{=}$ (direttamente!)

$= \frac{\partial}{\partial\vartheta_i}E^\mu[\mathcal{L}^\vartheta \underbrace{1_{T^{-1}(A)}}_{(1_A(T))}] = \frac{\partial}{\partial\vartheta_i}E^{P^\vartheta}[1_A(T)] = \frac{\partial}{\partial\vartheta_i}E^{T(P^\vartheta)}[1_A] \overset{(\text{procedendo a ritroso})}{=}$

$= E^{T(P^\vartheta)}\left[\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}_T^\vartheta\cdot 1_A\right] = E^{P^\vartheta}\Big[\frac{\partial}{\partial\vartheta_i}\log\mathcal{L}_T^\vartheta(T)\underbrace{1_A(T)}_{(1_{T^{-1}(A)}(\cdot))}\Big]$. □

Proposizione. Per ogni $\vartheta\in\Theta$, vale $\boxed{I_T(\vartheta)\le I(\vartheta)}$.

Dim. La tesi è che $I(\vartheta)-I_T(\vartheta)$ sia semi-def positiva per ogni $\vartheta\in\Theta$, cioè che $\forall\vartheta\in\Theta$ e $\forall x\in\mathbb{R}^k$, $x^t I_T(\vartheta)x \le x^T I(\vartheta)x$. Ma, $\forall\vartheta\in\Theta$, $I_T(\vartheta) = \mathrm{Cov}^{P^\vartheta}[\nabla_\vartheta\log\mathcal{L}_T^\vartheta(T)]$ e $I(\vartheta) = \mathrm{Cov}^{P^\vartheta}[\nabla_\vartheta\log\mathcal{L}^\vartheta]$, ed entrambi i vettori aleatori $\nabla_\vartheta\log\mathcal{L}_T^\vartheta(T)$ e $\nabla_\vartheta\log\mathcal{L}^\vartheta$ sono P^ϑ-centrati, $\vartheta\in\Theta$, per cui (come sappiamo) la tesi è che, $\forall\vartheta\in\Theta$ e $\forall x\in\mathbb{R}^k$, risulti

$$\mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta_T(T); x\rangle^2_{\mathbb{R}^k}] \leq \mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta; x\rangle^2_{\mathbb{R}^k}] .$$

Questa è una immediata conseguenza del fatto che $\nabla_\theta \log L^\theta_T(T) = \mathbb{E}^{P^\theta}[\nabla_\theta \log L^\theta \mid T]$: infatti, $\langle \nabla_\theta \log L^\theta_T(T); x\rangle \overset{\text{(+ linearità)}}{=} \mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta; x\rangle \mid T]$, da cui $\langle \nabla_\theta \log L^\theta_T(T); x\rangle^2 = (\mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta; x\rangle \mid T])^2 \overset{\text{(JENSEN condiz.)}}{\leq} \mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta; x\rangle^2 \mid T]$ e quindi segue la tesi passando a $\mathbb{E}^{P^\theta}[\dots]$. $\square$]

REMARK : $\langle \nabla_\theta \log L^\theta_T(T); x\rangle = \mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta; x\rangle \mid T]$ $\forall \theta \in \Theta, \forall x \in \mathbb{R}^k$.

Occupiamoci prima del caso T libera, poi del caso T esaustiva.

$\boxed{\text{NOTA}}$ Dire che T è libera dal modello, ossia che $T(P^\theta)$ è una volta costante in $\theta \in \Theta$ (è una sola legge di probabilità), equivale chiaramente ad avere che $L^\theta_T = \frac{dT(P^\theta)}{dT(\mu)}$ è costante in $\theta \in \Theta$ (è una sola densità), anche perché $T(P^\theta)[A] = \int_A L^\theta_T(x)\, dT(\mu)(x)$ $\forall A \in \mathcal{E}$. ✓

$\boxed{\text{PROP.}}$ $I_T(\theta) \equiv 0_{\mathbb{R}^{k^2}} \iff T$ è libera dal modello .

[Dim. $\boxed{\Leftarrow}$ Se T è libera, allora ovviamente $\nabla_\theta \log L^\theta_T \equiv 0$, da cui $I_T = 0$.

$\boxed{\Rightarrow}$ Se $Cov^{T(P^\theta)}[\nabla_\theta \log L^\theta_T] = 0$ $\forall \theta \in \Theta$, allora (ricordando pure che il vettore aleatorio $\nabla_\theta \log L^\theta_T$ è $T(P^\theta)$ - centrato) è $\overset{(L^\theta_T)^{-1}\nabla_\theta L^\theta_T =}{}\nabla_\theta \log L^\theta_T = \mathbb{E}^{T(P^\theta)}[\nabla_\theta \log L^\theta_T] = 0$ $T(P^\theta)$ - q.c , $\forall \theta \in \Theta$. Esplicitamente,

$\longrightarrow$ $\forall \theta \in \Theta$, $\exists N_\theta \in \mathcal{E}$ con $T(P^\theta)[N_\theta] = 0$ tale che, $\forall x \in E \setminus N_\theta$, sia $\nabla_\theta L^\theta_T(x) = 0$, cioè (per equivalenza $T(P^\theta) \sim T(\mu)$)

$\longmapsto$ $\forall \theta \in \Theta$, $\exists N_\theta \in \mathcal{E}$ con $T(\mu)[N_\theta] = 0$ tale che $\nabla_\theta L^\theta_T(x) = 0$ $\forall x \in E \setminus N_\theta$.

Alla, $\forall x \in E$, $\theta \mapsto \nabla_\theta \mathcal{L}^\theta_T(x)$ è (abbondantemente) continua (è $C^1(\Theta)$) ...... L'idea è considerare un sottoinsieme $D \subsetneq \Theta$ che sia numerabile e denso in Θ, sia $D \equiv \{\theta_n \mid n \in \mathbb{N}_+\}$, cosicché

$N := \bigcup_{n=1}^{\infty} N_{\theta_n}$ sta in $\mathcal{E}$ e resta $T(\mu)[N] = 0$: per ciò, abbiamo che esiste $N \in \mathcal{E}$ con $T(\mu)[N] = 0$ tale che, $\forall \theta \in D$, sia $\nabla_\theta \mathcal{L}^\theta_T(x) = 0$ per ogni $x \in E \setminus N$, e per densità di D (e regolarità di $\theta \mapsto \mathcal{L}^\theta_T(x)$) deduciamo subito che $\nabla_\theta \mathcal{L}^\theta_T(\cdot) = 0$ $T(\mu)$-q.o. per ogni $\theta \in \Theta$, cioè che $\mathcal{L}^\theta_T(\cdot) \equiv$ costante $T(\mu)$-q.o. per ogni $\theta \in \Theta$. $\square$

PROP. $I_T \equiv I \iff T$ è esaustiva per il modello.

Dim. **OSS.** Per ogni $\theta \in \Theta$ e $A \in \mathcal{E}$, vale $\int_\Omega 1\!\!1_A(T(\omega)) \mathcal{L}^\theta(\omega) \, d\mu(\omega) =$

$= \int_E 1\!\!1_A(x) \mathcal{L}^\theta_T(x) \, dT(\mu)(x)$ (: infatti, semplicemente, $T(P^\theta)[A] =$

$= P^\theta[T^{-1}(A)]$ e $T(P^\theta)[A] = \int_E 1\!\!1_A(x) \mathcal{L}^\theta_T(x) \, dT(\mu)(x)$ mentre, effetto,

$P^\theta[T^{-1}(A)] = \int_\Omega \underbrace{1\!\!1_{T^{-1}(A)}(\omega)}_{(1\!\!1_A(T(\omega)))} \mathcal{L}^\theta(\omega) \, d\mu(\omega)$.) //

$\boxed{\Leftarrow}$ Se T è esaustiva, dato che il modello è dominato (ed anzi è regolare), allora per Neyman-Fischer esistono ~~che~~ funzioni (misurabili) $h(\omega) : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e, $\forall \theta \in \Theta$, $g^\theta(x) : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, tali che risulti $\mathcal{L}^\theta(\omega) \overset{(\mu\text{-q.o.})}{=} g^\theta(T(\omega)) \cdot h(\omega)$ per ogni $(\theta, \omega) \in \Theta \times \Omega$. Più precisamente, se P_0 è una probabilità dominante privilegiata per il modello, allora in realtà $h(\omega) \overset{(\mu\text{-q.o.})}{=} \frac{dP_0}{d\mu}(\omega)$ ed è > 0 per regolarità. Per ciò, segue che, $\forall \theta \in \Theta$ e $\forall A \in \mathcal{E}$,

$$\int_E 1\!\!1_A(x) \mathcal{L}^\theta_T(x) \, dT(\mu)(x) \overset{(\text{OSS.})}{=} \int_\Omega 1\!\!1_A(T(\omega)) \, g^\theta(T(\omega)) \, dP_0(\omega) \overset{(\text{INTEGR. IMM.})}{=}$$

$=\int_E \mathbb{1}_A(x)\, g^\theta(x)\, d\,T(P_0)(x)$), e dunque (ricordando anche $T(P_0) \sim T(\mu)$)

$L^\theta_T(x) = g^\theta(x) \cdot \frac{d\,T(P_0)}{d\,T(\mu)}(x)$ (per $T(\mu)$-q.o. $x \in E$) : riassumendo,

$L^\theta(\omega) \overset{(\mu\text{-q.o.})}{=} g^\theta(T(\omega)) \cdot \frac{dP_0}{d\mu}(\omega) \Rightarrow L^\theta_T(x) \overset{(T(\mu)\text{-q.o.})}{=} g^\theta(x) \cdot \frac{d\,T(P_0)}{d\,T(\mu)}(x)$. ✓

Ma, adesso, $\frac{dP_0}{d\mu}(\omega)$ e $\frac{d\,T(P_0)}{d\,T(\mu)}(\omega)$ sono indipendenti da $\theta \in \Theta$, e il log di un prodotto è la somma dei logaritmi, per cui segue subito che

$\nabla_\theta \log L^\theta(\cdot) = \nabla_\theta \log g^\theta(T(\cdot))$ e $\nabla_\theta \log L^\theta_T(\cdot) = \nabla_\theta \log g^\theta(\cdot)$,

ed ora è evidente che $\mathrm{Cov}^{T(P^\theta)}[\nabla_\theta \log g^\theta(\cdot)] = \mathrm{Cov}^{P^\theta}[\nabla_\theta \log g^\theta(T(\cdot))]$,

cioè che $I_T(\theta) = I(\theta)$, per ogni $\theta \in \Theta$. ✓

$\boxed{\Rightarrow}$ Ragionando esattamente come nella dimostrazione del fatto che $I_T \leq I$, quando in particolare abbiamo ottenuto che, $\forall \theta \in \Theta$ e $\forall x \in \mathbb{R}^k$,

$E^{P^\theta}[\langle \nabla_\theta \log L^\theta ; x\rangle^2 \mid T] - \underset{(= E^{P^\theta}[\langle \nabla_\theta \log L^\theta; x\rangle \mid T]^2)}{\langle \nabla_\theta \log L^\theta_T(T) ; x\rangle^2} \geq 0$, vediamo subito che, se questo $I_T = I$, allora tale v.a.r. ≥ 0 ha $E^{P^\theta}[\ldots] = 0$, per cui necessariamente è $\equiv 0$ e dunque (sempre per Jensen cond.) $\langle \nabla_\theta \log L^\theta(\cdot) ; x\rangle$ è in realtà $\sigma(T)$-misurabile. ✓ (Allora, $\forall \theta \in \Theta$ e $\forall x \in \mathbb{R}^k$, $\langle \nabla_\theta \log L^\theta_T(T) ; x\rangle \overset{(\text{"REMARK"})}{=} \langle \nabla_\theta \log L^\theta ; x\rangle$, e tutto questo in senso P^θ-q.c. Dunque, più esplicitamente,

$\longmapsto$ $\forall \theta \in \Theta$ e $\forall x \in \mathbb{R}^k$, esiste $N_{\theta,x} \in \mathcal{A}$ con $\underset{(\mu \sim P^\theta)}{\mu(N_{\theta,x}) = 0}$ tale che, $\forall \omega \in \Omega \setminus N_{\theta,x}$, risulti $\langle \nabla_\theta \log L^\theta_T(T(\omega)) ; x\rangle = \langle \nabla_\theta \log L^\theta(\omega) ; x\rangle$. ✓

Ma, di nuovo per regolarità su Θ delle derivate delle verosimiglianze, se $D := \{\theta_m \mid m \in \mathbb{N}_1\}$ è $\subseteq \Theta$ numerabile e ivi denso, e se $\forall x \in \mathbb{R}^k$ poniamo

$N_x \doteq \bigcup_{m=1}^\infty N_{\theta_m,x}$, allora $N_x \in \mathcal{A}$ e $\mu[N_x] = 0$ ed è chiaro che

$\langle \nabla_\theta \log L^\theta_T(T(\omega)) ; x\rangle = \langle \nabla_\theta \log L^\theta(\omega) ; x\rangle$ $\forall x \in \mathbb{R}^k$, $\forall \omega \in \Omega \setminus N_x$ e ⫽

$\forall \vartheta \in \Theta$ (valendo $\forall \vartheta \in D$), e così abbiamo ottenuto che

$\Rightarrow \forall x \in \mathbb{R}^k$, $\exists N_x \in \mathcal{F}$ con $\mu[N_x]=0$ tale che, $\forall \omega \in \Omega \setminus N_x$ e $\forall \vartheta \in \Theta$, sia $\langle \nabla_\vartheta \log L_T^\vartheta(T(\omega))\,;\, x\rangle = \langle \nabla_\vartheta \log L^\vartheta(\omega)\,;\, x\rangle$.

Adesso è immediato dedurne che $\exists N \in \mathcal{F}$ con $\mu[N]=0$ tale che, $\forall \omega \in \Omega \setminus N$ e $\forall \vartheta \in \Theta$, risulti $\nabla_\vartheta \log L_T^\vartheta(T(\omega)) = \nabla_\vartheta \log L^\vartheta(\omega)$: fissata infatti una qualsiasi base $x_1, \dots, x_k$ di $\mathbb{R}^k$, basta prendere $N := \bigcup_{i=1}^{k} N_{x_i}$ ($\in \mathcal{F}$, e con $\mu[N]=0$) per avere che, $\forall \omega \in \Omega \setminus N$ e $\forall \vartheta \in \Theta$, risulta $\langle \nabla_\vartheta \log L_T^\vartheta(T(\omega))\,;\, x_h\rangle = \langle \nabla_\vartheta \log L^\vartheta(\omega)\,;\, x_h\rangle$ per ogni $h = 1, \dots, k$, e così $\nabla_\vartheta \log L_T^\vartheta(T(\omega)) = \nabla_\vartheta \log L^\vartheta(\omega)$ in quanto appunto $x_1, \dots, x_k$ è base di $\mathbb{R}^k$.

Pertanto, $\forall \omega \notin N$ e $\forall \vartheta \in \Theta$, $\log L^\vartheta(\omega) = \log L_T^\vartheta(T(\omega)) + f(\omega)$ con $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mis. ed indipendente da Θ (infatti $f^\vartheta(\omega) \underset{(\mu\text{-q.o.})}{\equiv} \log L^\vartheta(\omega) - \log L_T^\vartheta(T(\omega))$ che $\nabla_\vartheta f^{\vartheta}(\omega) = 0$ (e Θ è convesso!)): allora

$$L^\vartheta(\omega) \overset{\mu\text{-q.o.}}{=} L_T^\vartheta(T(\omega)) \cdot e^{f(\omega)}$$ per ogni $\vartheta \in \Theta$, e dunque T è esaustiva sempre per Neyman–Fischer. $\square$

$\boxed{\text{OSS.}}$ Sia sempre $(\Omega, \mathcal{F}, \{P^\vartheta | \vartheta \in \Theta\}, \mu)$ un modello stat. dominato, anzi regolare, sul quale sia possibile definire la matrice d'informazione secondo Fischer in ogni $\vartheta \in \Theta$: dunque $L^\vartheta(\omega) = \frac{dP^\vartheta}{d\mu}(\omega) > 0$ "sempre", $\vartheta \mapsto L^\vartheta(\omega) \in C^2(\Theta)$ $\forall \omega \in \Omega$, $L^\vartheta(\cdot), \log L^\vartheta(\cdot), \frac{\partial}{\partial \vartheta_i} L^\vartheta(\cdot), \frac{\partial}{\partial \vartheta_i} \log L^\vartheta(\cdot)$ in $L^2(P^\vartheta)$ $\forall \vartheta \in \Theta$ e $\forall i = 1, \dots, k$, e $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} L^\vartheta(\cdot)$, $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log L^\vartheta(\cdot)$ in $L^1(P^\vartheta)$ $\forall \vartheta \in \Theta$ e $\forall i, j = 1, \dots, k$, con sottinteso ormai che $\Theta \subseteq \mathbb{R}^k$ sia aperto convesso $\neq \emptyset$. Sia ora $T : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ una statistica con $T(\mu)$ σ-finita su $\mathcal{E}$ che sia equivalente ad un'altra misura σ-finita ν su $\mathcal{E}$ in modo che $\frac{d\nu}{dT(\mu)}$ o $\frac{dT(\mu)}{d\nu}$ siano limitate su E, e consideriamo il modello immagine mediante T $(E,\mathcal{E},\{T(P^\theta)\mid\theta\in\Theta\})$ e quindi, $\forall(\theta,x)\in\Theta\times E$,

$$L_T^\theta(x) = \frac{dT(P^\theta)}{dT(\mu)}(x) = \frac{dT(P^\theta)}{d\nu}(x)\cdot\frac{d\nu}{dT(\mu)}(x)$$

: allora le "condizioni standard per Fischer" su $L_T^\theta(x)$ equivalgono a quelle per $\frac{dT(P^\theta)}{d\nu}(x)$, e, se sussistono, allora $I_T(\theta) := \mathrm{Cov}^{T(P^\theta)}[\nabla_\theta \log L_T^\theta(\cdot)] = \mathrm{Cov}^{T(P^\theta)}[\nabla_\theta\log\frac{dT(P^\theta)}{d\nu}(\cdot)]$ per ogni $\theta\in\Theta$. (Detto altrimenti, in pratica, si può supporre direttamente che sia $T(\mu)=\nu$.) [Dim. Infatti $L_T^\theta(x)>0 \Leftrightarrow \frac{dT(P^\theta)}{d\nu}(x)>0$, e $\frac{d\nu}{dT(\mu)}(\cdot)$ NON dipende da $\theta\in\Theta$, per cui $\theta\mapsto L_T^\theta(x)$ è $C^2(\Theta)$ $\forall x\in E \Leftrightarrow \forall x\in E$, $\theta\mapsto\frac{dT(P^\theta)}{d\nu}(x)$ è $C^2(\Theta)$, ed in tale situazione la limitatezza di $\frac{d\nu}{dT(\mu)}(x)$ (ed indip. da Θ) garantisce che $L_T^\theta(\cdot), \frac{\partial}{\partial\theta_i}L_T^\theta(\cdot)\in L^2(P^\theta) \Leftrightarrow \frac{dT(P^\theta)}{d\nu}(\cdot), \frac{\partial}{\partial\theta_i}\frac{dT(P^\theta)}{d\nu}(\cdot)\in L^2(P^\theta)$ ($\theta\in\Theta$, $i=1,\dots,n$), e anche $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_T^\theta(\cdot)\in L^1(P^\theta)\Leftrightarrow \frac{\partial^2}{\partial\theta_i\partial\theta_j}\frac{dT(P^\theta)}{d\nu}(\cdot)\in L^1(P^\theta)$ ($\theta\in\Theta$, $i,j=1,\dots,n$). Inoltre, ciò è analogo passando ai $\log L_T^\theta(x)$ e $\log\frac{dT(P^\theta)}{d\nu}(\cdot)$, che infine hanno anche esattamente lo stesso $\nabla_\theta(\cdot)$. □]

ES. Mettiamoci nel modello stat. per un campione $(X_1,\dots,X_n)$ di legge μ^θ $N(\theta,1)$ su $\mathbb{R}$, $\theta\in\mathbb{R}$: $(\Omega,\mathcal{A},\{P^\theta\mid\theta\in\Theta\},\mu)=(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\{(\mu^\theta)^{\otimes n}\mid\theta\in\mathbb{R}\},\nu^{\otimes n})$ con ν di Lebesgue 1-dim., $\frac{d\mu^\theta}{d\nu}(x)=\exp\left(-\frac{1}{2}(x-\theta)^2-\log\sqrt{2\pi}\right)$ $\forall\theta\in\mathbb{R}, \forall x\in\mathbb{R}$, e $X_i(x_1,\dots,x_n)=x_i$. Allora, come già visto (o comunque evidente), è ben definita l'informazione secondo Fischer, ed è $I(\theta)\equiv n$ ($\forall\theta\in\mathbb{R}$). D'altra parte, grazie a Neyman-Fischer, è evidente pure che la statistica reale $T:=\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ sia esaustiva per il modello: troviamo il relativo modello immagine e verifichiamo che $I_T\equiv I$ ($\equiv n$). Ma, $\forall\theta\in\mathbb{R}$, l'indipendenza delle X_i dà subito che la legge di $T(P^\theta)$ è una $N(\theta,\frac{1}{n})$, equivalente a ν, per cui il modello immagine è $(\mathbb{R},\mathcal{B}(\mathbb{R}),\{N(\theta,\frac{1}{n})\mid\theta\in\Theta\})$ e possiamo supporre per i calcoli che $T(\mu)=\nu$. Ma allora è ben definita l'informazione secondo Fischer in modo del tutto analogo a prima, e viene $I_T(\theta) \equiv 1/(1/n) = n$. ✓

(ES.) Consideriamo invece il modello stat. per un campione $(X_1, \dots, X_n)$ (n∈ℕ, n≥2) di legge μ^θ $N(m, \sigma^2)$ con $\theta = \sigma^2 \in (0, +\infty)$, mentre $m \in \mathbb{R}$ è nota: dunque $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mu^{\theta \otimes n} | \theta = \sigma^2 > 0\}, \nu^{\otimes n})$ con ν di Lebesgue 1-dim., "$\mu^\theta = N(m, \theta)$", e $X_i(x_1, \dots, x_n) = x_i$. Allora il modello ammette informazione di Fischer $I(\sigma^2) = \frac{n}{2\sigma^4}$, $\forall \sigma^2 > 0$.

Inoltre, per tale modello risulta esaustiva la statistica $-\sum_{i=1}^n X_i^2$ congiuntamente e $\sum_{i=1}^n X_i$. Dunque, considerata $T := S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, statistica ≥ 0, T "difficilmente" sarà esaustiva (per σ^2) ...

Fatto: $\frac{n-1}{\sigma^2} T = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2$ ha legge $\chi^2(n-1) \doteq \Gamma(\frac{n-1}{2}, \frac{1}{2})$,
(che diamo per OK)
ossia ha densità $f(x) = 1_{(0,+\infty)}(x) \frac{1}{\Gamma(\frac{n-1}{2})} (\frac{1}{2})^{\frac{n-1}{2}} \underbrace{x^{\frac{n-1}{2} - 1}}_{(x^{\frac{n-3}{2}})} e^{-\frac{1}{2}x}$, $x \in \mathbb{R}$.

$(\Rightarrow)$ $\frac{n-1}{\sigma^2} T$ ha speranza $\frac{\frac{n-1}{2}}{\frac{1}{2}} = n-1$ (cioè $E^{\sigma^2}[\frac{n-1}{\sigma^2} T] = n-1$)

Dunque $\frac{n-1}{\sigma^2} T$ è libera dal modello, e allora T non sarà (neanche) libera dal modello ... Dimostriamo che T è né libera né esaustiva verificando che per il relativo modello immagine è ben definita l'informazione di Fischer $I_T(\sigma^2)$ $\forall \sigma^2 > 0$, e che $0 < I_T(\sigma^2) < I(\sigma^2) = \frac{n}{2\sigma^4}$ $\forall \sigma^2 > 0$. ✓ Cominciamo osservando che

T ammette densità $f_T^{\sigma^2}(t) = 1_{(0,+\infty)}(t) \frac{n-1}{\sigma^2} f(\frac{n-1}{\sigma^2} t) =$
$= 1_{(0,+\infty)}(t) \frac{1}{\Gamma(\frac{n-1}{2}) 2^{\frac{n-1}{2}}} \left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} t^{\frac{n-3}{2}} e^{-\frac{(n-1)t}{2\sigma^2}}$, $t \in \mathbb{R}$ (per ogni $\sigma^2 > 0$).

[Infatti, $\forall t \in \mathbb{R}$, $\forall \sigma^2 > 0$, $P^{\sigma^2}[T \leq t] = P^{\sigma^2}[\underbrace{\frac{n-1}{\sigma^2} T}_{(\sim \Gamma(\frac{n-1}{2}, \frac{1}{2}))} \leq \frac{n-1}{\sigma^2} t] =$

$= \int_{-\infty}^{\frac{n-1}{\sigma^2}t} f(x)\,dx \underset{\left(x = \frac{n-1}{\sigma^2}y\right)}{=} \int_{-\infty}^{t} \frac{n-1}{\sigma^2} f\left(\frac{n-1}{\sigma^2}y\right)dy$ . ...] Dunque il

modello immagine $\left((0,+\infty), \mathcal{B}((0,+\infty)), \{f_T^{\sigma^2}\cdot d\nu \mid \sigma^2>0\}, \nu\right)$ ammette informazione secondo Fisher e possiamo supporre $T(\mu) = \nu$: dunque, $\forall \sigma^2>0$,
$\left(\Rightarrow f_T^{\sigma^2} = \frac{dT(P^{\sigma^2})}{dT(\mu)} = f_T^{\sigma^2}\right)$

$I_T(\sigma^2) = -E^{T(P^{\sigma^2})}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log f_T^{\sigma^2}\right]$. Calcoliamo: $\forall\sigma^2>0, \forall t>0$,

$\log f_T^{\sigma^2}(t) = \frac{n-1}{2}\log\frac{n-1}{\sigma^2} - \frac{(n-1)t}{2\sigma^2}$ + «costante in σ^2», da cui

$\frac{\partial}{\partial\sigma^2}\log f_T^{\sigma^2}(t) = \frac{n-1}{2}\frac{\sigma^2}{n-1}\left(-\frac{n-1}{\sigma^4}\right) + \frac{(n-1)t}{2\sigma^4} = -\frac{n-1}{2\sigma^2}+\frac{(n-1)t}{2\sigma^4}$, e

allora $\frac{\partial^2}{\partial(\sigma^2)^2}\log f_T^{\sigma^2}(t) = \frac{n-1}{2\sigma^4} - \frac{(n-1)t}{\sigma^6}$, e per questo

$I_T(\sigma^2) = -\frac{n-1}{2\sigma^4} + E^{T(P^{\sigma^2})}\left[\frac{(n-1)t}{\sigma^6}\right] = \underbrace{E^{T(P^{\sigma^2})}\left[\frac{n-1}{\sigma^2}t\right]}_{=(n-1)}\frac{1}{\sigma^4} - \frac{n-1}{2\sigma^4} =$

$= \frac{n-1}{2\sigma^4}$, effettivamente >0 e $< \frac{n}{2\sigma^4}$. ✓

Pertanto, desiderando un modo opportuno per "riassumere" un complicato modello statistico dominato, che possibilmente generalizzasse il ~~...~~ cercare una statistica esaustiva (al meglio completa) per esso, abbiamo definito e studiato il concetto di informazione secondo Fisher e possiamo volentieri sorvolarci a riguardo. Tuttavia, riassunto il modello, dobbiamo provare a fare un ritorno alla teoria delle stime, parlando particolareggiando ulteriormente il modello stesso: in effetti, se la matrice d'informazione secondo Fisher è invertibile (in ogni $\theta\in\Theta$) e così definita positiva, allora è possibile ottenere una significativa "stima" inferiore per il rischio quadratico di un dato stimatore di una data funzione di $\theta\in\Theta$! Ebbene, vedremo così e fare qualche osservazione a riguardo, ma solo dopo qualche elementare ragionamento di algebra lineare (in $\mathbb{R}^k$).

Sia $k \in \mathbb{N}_*$ e sia $A \in \mathbb{R}^{k^2}$ una matrice reale $k\times k$ simmetrica e semi-definita positiva, per cui $A^t = A$ e $\langle Ax ; x\rangle \geq 0$ per ogni $x \in \mathbb{R}^k$. Allora, grazie al classico teorema spettrale „in $\mathbb{R}^{k^2}$", esiste una base ortonormale di $\mathbb{R}^k$ costituita da autovettori di A aventi autovalori non-negativi: esistono cioè $\overset{k \text{ scalari non-negativi}}{\lambda_1, \ldots, \lambda_k} \geq 0$ e $\overset{k \text{ vettori ortonormali}}{b_1, \ldots, b_k} \in \mathbb{R}^k$ tali che $Ab_j = \lambda_j b_j$ per ogni $j = 1, \ldots, k$. Per ciò, considerate le due matrici (di $\mathbb{R}^{k^2}$) $U := [b_1 | \ldots | b_k]$ (mettendo più b_j „in colonna") e $D := \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{bmatrix} = \text{diag}[\lambda_1, \ldots, \lambda_k]$, vale che U è ortogonale (cioè invertibile con $U^{-1} = U^t$) e D è diagonale, e $\boxed{AU = UD}$, cioè $\boxed{A = UDU^t}$ o anche $\boxed{U^t A U = D}$ („diagonalizzazione di A").

Come conseguenza, l'esistenza di una „radice quadrata" $B \in \mathbb{R}^{k^2}$ di A, ossia di una matrice $B \in \mathbb{R}^{k^2}$ tale che $B^2 = B \cdot B = A$. Infatti, posta $\boxed{\sqrt{D} := \text{diag}[\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k}]} \in \mathbb{R}^{k^2}$ (tale che $(\sqrt{D})^2 = D$), $B := U\sqrt{D}U^t$ è chiaramente tale che $B^2 = A$: la indichiamo con $\boxed{B \doteq \text{„}\sqrt{A}\text{"} = U\sqrt{D}U^t}$. Dunque, per l'ovvietà, questa $\sqrt{A}$ resta anche simmetrica e semi-definita positiva: infatti, $\forall x \in \mathbb{R}^k$, $\langle \sqrt{A}x ; x\rangle \overset{(\text{def.})}{=} \langle (U\sqrt{D}U^t)x ; x\rangle \overset{(\text{ovvio})}{=} \langle \sqrt{D}(U^t x) ; (U^t x)\rangle \geq 0$ (perché $\sqrt{D}$ è semi-definita positiva). Ma in effetti, essendo $\sqrt{A}U = U\sqrt{D}$, è evidente che la base $\overset{(\text{ortonormale})}{b_1, \ldots, b_k}$ sia di autovettori anche per $\sqrt{A}$, ma relativi agli autovalori $\sqrt{\lambda_1},\dots,\sqrt{\lambda_n}$ di $\sqrt{A}$ rispettivamente, ed essendo una base così (fatto sia per A si è effettuato per $\sqrt{A}$) abbiamo automaticamente la semi-definizione positiva dal fatto che $\lambda_1,\dots,\lambda_n$ sono ≥ 0 (e quindi anche $\sqrt{\lambda_1},\dots,\sqrt{\lambda_n}$): infatti, per ogni $\alpha_1,\dots,\alpha_n \in \mathbb{R}$, $\langle \underbrace{A \sum_{j=1}^{n} \alpha_j b_j}_{\sum_{j=1}^{n} \alpha_j \lambda_j b_j} ; \sum_{j=1}^{n} \alpha_j b_j \rangle \overset{\text{(autovettori + ortonormalità)}}{=} \sum_{i=1}^{n} \left((\alpha_i)^2 \lambda_i \right) \geq 0$,

e anche $\langle \sqrt{A} \sum_{j=1}^{n} \alpha_j b_j ; \sum_{j=1}^{n} \alpha_j b_j \rangle = \sum_{j=1}^{n} \left((\alpha_j)^2 \sqrt{\lambda_j} \right) \geq 0$. Pertanto, è chiaro che una matrice $n \times n$ A simmetrica e semi-definita positiva è invertibile se, e solo se, ha tutti autovalori > 0 (cioè D è invertibile, ed in effetti $\det A = \det D$, ed in tal caso $A^{-1} = U D^{-1} U^t$), e cioè se e solo se è definita positiva. In particolare, A è invertibile (definita positiva) se e solo se $\sqrt{A}$ è invertibile (definita positiva), ed in tal caso $\boxed{(\sqrt{A})^{-1} = \sqrt{A^{-1}}}$.

[Anzitutto, $A^{-1} = U D^{-1} U^t$ è simmetrica e definita positiva, ed ammette perciò $\sqrt{A^{-1}} = U \sqrt{D^{-1}}\, U^t$, mentre $(\sqrt{A})^{-1} = U (\sqrt{D})^{-1} U^t$, per cui la tesi è $(\sqrt{D})^{-1} = \sqrt{D^{-1}}$, cioè che $\operatorname{diag}[\sqrt{\lambda_1},\dots,\sqrt{\lambda_n}]^{-1} = \operatorname{diag}[\frac{1}{\sqrt{\lambda_1}},\dots,\frac{1}{\sqrt{\lambda_n}}]$: questo è evidente. ✓ (Comunque è evidente che $((\sqrt{A})^{-1})^2 = A^{-1}$, perché $(\sqrt{A})^{-1} = U (\sqrt{D})^{-1} U^t$ che quadrato porge $U D^{-1} U^t = A^{-1}$ e basta sempre di osservare che $((\sqrt{D})^{-1})^2 = D^{-1}$). ✓]

(Ma in questa situazione_)

NOTA Fissato su $\mathbb{R}^n$ l'usuale prodotto scalare euclideo $\langle \cdot ; \cdot \rangle$, il quale induce l'usuale norma euclidea $\|\cdot\|_2$, e la quale induce a sua volta la norma matriciale "$\|\cdot\|_2$", osserviamo che $\|U\|_2 = \|U^t\|_2 = 1$ (essendo $U^t U = U U^t = I_n$) e quindi che $\|A\|_2 = \|D\|_2$ (essendo

$A = UDU^t$ ma anche $D = U^t A U$, e $\|\cdot\|_2$ è sub-moltiplicativa), così come (ad esempio) $\|A^{-1}\|_2 = \|D^{-1}\|_2$, e $\|\sqrt{A}\|_2 = \|\sqrt{D}\|_2$. Allora $\|D^{-1}\|_2 = \|D\|_2^{-1}$, mentre $\|\sqrt{D}\|_2 = \sqrt{\|D\|_2}$, da cui $\boxed{\|\sqrt{A}\|_2 = \sqrt{\|A\|_2}}$. ✓

UNA CONSEGUENZA: scelti $k \in \mathbb{N}^*$, $A \in \mathbb{R}^{k^2}$ simmetrica e ~~...~~ definita positiva e $a \in \mathbb{R}^k$, la funzione $\mathbb{R}^k \setminus \{0\} \to [0, +\infty)$,

$x \longmapsto \dfrac{\langle x; a \rangle^2}{\langle Ax; x \rangle}$ per ogni $x \in \mathbb{R}^k \setminus \{0\}$, ammette valor massimo assoluto dato da

$$\boxed{\max_{x \in \mathbb{R}^k \setminus \{0\}} \frac{\langle x; a \rangle^2}{\langle Ax; x \rangle} = \langle A^{-1} a; a \rangle}$$

(e raggiunto, per esempio, in $\bar{x} \neq 0$ qualsiasi se $a = 0$; mentre in $\bar{x} := A^{-1} a$ per ogni $a \neq 0$).

[Dim. Il fatto è che $\langle Ax; x \rangle = \|\sqrt{A} x\|_2^2$ e $\langle A^{-1} a; a \rangle = \|\sqrt{A^{-1}} a\|_2^2 =$ $= \|(\sqrt{A})^{-1} a\|_2^2$, da cui $|\langle x; a \rangle| = |\langle (\sqrt{A})^{-1} \sqrt{A} x; a \rangle| \overset{(\text{simm.})}{=}$ $= |\langle \sqrt{A} x; (\sqrt{A})^{-1} a \rangle| \overset{(\text{C.-S.})}{\leq} \|\sqrt{A} x\|_2 \|(\sqrt{A})^{-1} a\|_2$ e quindi, $\forall x \in \mathbb{R}^k \setminus \{0\}$, $\dfrac{\langle x; a \rangle^2}{\|\sqrt{A} x\|_2^2} \leq \|(\sqrt{A})^{-1} a\|_2^2$, da cui la tesi. □]

NOTA Nella situazione del precedente risultato, vale che, $\forall x \in \mathbb{R}^k \setminus \{0\}$,

$$\langle A^{-1} a; a \rangle \geq \frac{\langle x; a \rangle^2}{\|\sqrt{A} \cdot x\|^2} \overset{(\text{SUB-MOLT.})}{\geq} \frac{\langle x; a \rangle^2}{\|\sqrt{A}\|^2 \|x\|^2} \overset{(\text{VISTO})}{=} \frac{\langle x; a \rangle^2}{\|x\|^2} \cdot \frac{1}{\|A\|}. \checkmark$$

Torniamo finalmente ad un modello stat. per il quale risulti ben definibile la matrice d'informazione secondo Fisher: sia per questo $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\}, \mu)$ un modello stat. regolare con $\Theta \subseteq \mathbb{R}^k$ un aperto convesso $\neq \emptyset$ e con $L^\theta(\omega) := \frac{dP^\theta}{d\mu}(\omega)$ (>0 "sempre") tale che $\theta \mapsto L^\theta(\omega)$ sia in $C^2(\Theta)$ $\forall \omega \in \Omega$, $L^\theta(\cdot), \frac{\partial}{\partial \theta_i} L^\theta(\cdot), \log L^\theta(\cdot), \frac{\partial}{\partial \theta_i} \log L^\theta(\cdot)$ in $L^2(P^\theta)$ $\forall \theta \in \Theta$ e $\forall i = 1,\dots,k$, $\frac{\partial^2}{\partial \theta_i \partial \theta_j} L^\theta(\cdot)$ e $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L^\theta(\cdot)$ in $L^2(P^\theta)$ $\forall \theta \in \Theta$ e $\forall i,j = 1,\dots,k$.

Allora, quale che sia X in $L^2(P^\theta)$ $\forall \theta \in \Theta$, e $\forall \theta \in \Theta$ e $\forall i = 1,\dots,k$, vale

$$\frac{\partial}{\partial \theta_i} \int_\Omega L^\theta(\omega) X(\omega)\, d\mu(\omega) = \int_\Omega \frac{\partial}{\partial \theta_i} L^\theta(\omega) X(\omega)\, d\mu(\omega)$$

, e cioè (osservato che $dP^\theta(\omega) = L^\theta(\omega)\, d\mu(\omega)$ e $L^\theta(\omega) \frac{\partial}{\partial \theta_i} \log L^\theta(\omega) = \frac{\partial}{\partial \theta_i} L^\theta(\omega)$) vale

$$\frac{\partial}{\partial \theta_i} \mathbb{E}^{P^\theta}[X] = \mathbb{E}^{P^\theta}\left[X \frac{\partial}{\partial \theta_i} \log L^\theta\right]$$

, o più compattamente

$$\boxed{\nabla_\theta \mathbb{E}^{P^\theta}[X] = \mathbb{E}^{P^\theta}[X \nabla_\theta \log L^\theta]}$$

$\forall \theta \in \Theta$. Dunque, inoltre, il vettore aleatorio $\nabla_\theta \log L^\theta$ è P^θ-centrato $\forall \theta \in \Theta$, e resta ben definita la matrice d'informazione secondo Fisher $I(\theta) = (I(\theta)_{i,j})_{i,j=1,\dots,k}$ del modello in ogni $\theta \in \Theta$, data da $I(\theta)_{i,j} := \mathrm{Cov}^{P^\theta}[\nabla_\theta \log L^\theta]_{i,j}$ $\Big($ $= \mathbb{E}^{P^\theta}\left[\frac{\partial}{\partial \theta_i} \log L^\theta \frac{\partial}{\partial \theta_j} \log L^\theta\right] = -\mathbb{E}^{P^\theta}\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L^\theta\right]$ $\Big)$, e cioè

$$\boxed{I(\theta) = \mathrm{Cov}^{P^\theta}[\nabla_\theta \log L^\theta]}$$

, $\forall \theta \in \Theta$ e $\forall i,j = 1,\dots,k$. In particolare, per ogni $x \in \mathbb{R}^k$, risulta $\langle I(\theta) x ; x \rangle = \mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta ; x \rangle^2]$.

PROP. All'interno di questo contesto statistico, e nelle stesse notazioni, se $I(\theta)$ è <u>invertibile</u>, o cioè definita positiva, per ogni $\theta \in \Theta$, allora quale che sia $X \in \bigcap_{\theta \in \Theta} L^2(P^\theta)$ vale la disuguaglianza

$$\mathrm{Var}^{P^\theta}[X] \geq \langle I(\theta)^{-1} \cdot \nabla_\theta \mathbb{E}^{P^\theta}[X] ; \nabla_\theta \mathbb{E}^{P^\theta}[X] \rangle \quad \forall \theta \in \Theta.$$

[Dim. Per questo lemma/richiamato, è del tutto evidente che, $\forall \theta \in \Theta$ risulti,

$\nabla_\theta \mathbb{E}^{P^\theta}[X] = \mathbb{E}^{P^\theta}[(X - \mathbb{E}^{P^\theta}[X]) \nabla_\theta \log L^\theta]$, da cui $\forall x \in \mathbb{R}^k$ abbiamo

$\langle \nabla_\theta \mathbb{E}^{P^\theta}[X] ; x \rangle \overset{(\text{lin.})}{=} \mathbb{E}^{P^\theta}[(X - \mathbb{E}^{P^\theta}[X]) \langle \nabla_\theta \log L^\theta ; x \rangle]$, da cui

(per Hölder / Cauchy - Schwarz) $\langle \nabla_\theta \mathbb{E}^{P^\theta}[X] ; x \rangle^2 = \mathbb{E}^{P^\theta}[\dots]^2 \leq$

$\leq \mathrm{Var}^{P^\theta}[X] \, \mathbb{E}^{P^\theta}[\langle \nabla_\theta \log L^\theta ; x \rangle^2] \overset{(\text{visto})}{=} \mathrm{Var}^{P^\theta}[X] \langle I(\theta) x ; x \rangle$:

Pertanto, $\forall \theta \in \Theta$ e $\forall x \in \mathbb{R}^k \setminus \{0\}$, $\mathrm{Var}^{P^\theta}[X] \geq \dfrac{\langle x ; \nabla \mathbb{E}^{P^\theta}[X] \rangle}{\langle I(\theta) x ; x \rangle}$, da

cui risulta la tesi passando al "$\sup_{x \in \mathbb{R}^k \setminus \{0\}}$" (anzi "$\max_{x \in \mathbb{R}^k \setminus \{0\}}$") col usando un fatto elementare richiamato di algebra lineare. $\square$]

Torniamo al problema di stima di una funzione $g(\theta) : \Theta \to D$, $D \subseteq \mathbb{R}$ aperto $\neq \emptyset$ fissato, interessandoci a quegli stimatori $U : \Omega \to D$ (v.a.) che stanno in $\bigcap_{\theta \in \Theta} L^2(P^\theta)$, cercando di minimizzare il rischio quadratico $R_U(\theta) \underset{(\theta \in \Theta)}{=} \mathbb{E}^{P^\theta}[|U - g(\theta)|^2] \overset{(\text{visto})}{=} \mathrm{Var}^{P^\theta}[U] + (\mathbb{E}^{P^\theta}[U] - g(\theta))^2$

Abbiamo, grazie alla proposizione precedente, nel caso che esista $I(\theta)^{-1}$, abbiamo

$$R_U(\theta) \geq \langle I(\theta)^{-1} \nabla_\theta \mathbb{E}^{P^\theta}[U] ; \nabla_\theta \mathbb{E}^{P^\theta}[U] \rangle + (\mathbb{E}^{P^\theta}[U] - g(\theta))^2 \quad (\forall \theta \in \Theta)$$
$$\left(\geq \langle I(\theta)^{-1} \nabla_\theta \mathbb{E}^{P^\theta}[U] ; \nabla_\theta \mathbb{E}^{P^\theta}[U] \rangle \right).$$

DISUGUAGLIANZA DI CRAMER - ~~[illegible]~~ - RAO: se $I(\theta)$ è invertibile e se U è corretto per g (cioè $\mathbb{E}^{P^\theta}[U] = g(\theta) \; \forall \theta \in \Theta$), allora $\forall \theta \in \Theta$ vale

$$\underset{(\mathrm{Var}^{P^\theta}[U])}{R_U(\theta)} \geq \langle I(\theta)^{-1} \nabla_\theta g(\theta) ; \nabla_\theta g(\theta) \rangle .$$

→ Alla luce di tutto questo, considereremo "STIMATORE EFFICACE DI g" uno stimatore U di g che sia in $\bigcap_{\theta\in\Theta} L^2(P_\theta)$, che sia corretto per g, e che abbia esattamente $(R_U(\theta) =)\ \mathrm{Var}^{P_\theta}[U] = \langle I(\theta)^{-1}\nabla_\theta g(\theta); \nabla_\theta g(\theta)\rangle$ $\forall\theta\in\Theta$.

$\boxed{\text{NOTA}}$ In ogni caso, sarebbe affermato $\underset{(\mathrm{Var}^{P_\theta}[U])}{R_U(\theta)} \geq \langle I(\theta)^{-1}\nabla_\theta g(\theta); \nabla_\theta g(\theta)\rangle \overset{(\text{visto})}{\geq}$

$\geq \left(\frac{\langle x; \nabla_\theta g(\theta)\rangle^2}{\|x\|^2}\right)\cdot \frac{1}{\|I(\theta)\|_2}$ $\forall x\in\mathbb{R}^k\setminus\{0\}$ fissato, da cui osserviamo che più "$I(\theta)$ è grande", o cioè più sofisticato del modello, e meno saranno variabili gli stimatori di quadrato integrabili corretti di $g(\theta)$ (rispetto alla loro P-media, che è proprio $g(\theta)$), a meno di una qualche compensazione che derivi dalle derivate di $g(\theta)$ stesse. ✓

Terminiamo la sezione mostrando che esistono stimatori efficaci di una qualche $g(\theta)$, ma anche che esistono $g(\theta)$ che non ammettono stimatori efficaci.

ES. Sia $(\Omega,\mathcal{A})$ mis. tale che esistano $k\in\mathbb{N}_*$, $T : (\Omega,\mathcal{A})\to(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$ mis. e μ misura σ-finita su $\mathcal{A}$ tale che $T(\mu)$ sia σ-finita su $(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$ e tale che esiste un aperto convesso $(\neq\emptyset)$ $\Theta\subseteq\mathbb{R}^k$ tale che, $\forall\theta\in\Theta$, risulti $0 < \int_\Omega \exp(\langle\theta; T(\omega)\rangle)\,d\mu(\omega) < +\infty$. Posta quindi $f(\theta) := \log\left(\int_\Omega \exp\langle\theta; T(\omega)\rangle\,d\mu(\omega)\right)$ $\forall\theta\in\Theta$, consideriamo il modello stat. regolare $(\Omega,\mathcal{A},\{P_\theta\,|\,\theta\in\Theta\},\mu)$ avente $L^\theta(\omega) := \frac{dP_\theta}{d\mu}(\omega) :=$
$:= \exp(\langle\theta; T(\omega)\rangle - f(\theta))$ $\forall(\theta,\omega)\in\Theta\times\Omega$, o brevemente "il modello esponenziale associato a T e a μ". Allora T è esaustiva e completa per il modello, ed inoltre ammette momenti di ogni ordine: precisamente, (in particolare,) f è $C^\infty(\Theta)$ e vale $\frac{\partial f}{\partial\theta_i}(\theta) = E^{P_\theta}[T_i]$ e

$\frac{\partial^2 f}{\partial\theta_i\partial\theta_j}(\theta) = Cov^{P^\theta}[T_i, T_j]$ per ogni $\theta\in\Theta$ e $i,j=1,\dots,k$. Per ciò, resta ben definita (ed invertibile) la matrice d'informazione secondo Fisher $(I(\theta)_{i,j})_{i,j=1,\dots,k}$ del modello in ogni $\theta\in\Theta$, ed è proprio $I(\theta)_{i,j} = Cov^{P^\theta}[T_i,T_j] = \frac{\partial^2 f}{\partial\theta_i\partial\theta_j}(\theta)$. ✓ Ebbene, volendo stimare proprio $g_h(\theta) = \frac{\partial}{\partial\theta_h}f(\theta)$, $h=1,\dots,k$, abbiamo chiaramente che T_h ne è stimatore corretto (di quadrato integrabile): verifichiamo che è pure efficiente, cioè che $Var^{P^\theta}[T_h] = \langle I(\theta)^{-1}\nabla_\theta\frac{\partial}{\partial\theta_h}f(\theta) ; \nabla_\theta\frac{\partial}{\partial\theta_h}f(\theta)\rangle$ $\forall\theta\in\Theta$, $h=1,\dots,k$. Ora, infatti, $Var^{P^\theta}[T_h] = Cov^{P^\theta}[T_h,T_h] = \frac{\partial^2}{\partial(\theta_h)^2}f(\theta)$, e $\nabla_\theta\frac{\partial}{\partial\theta_h}f(\theta) \overset{(imm.)}{=} I(\theta)\cdot e_h$ $\left(\underset{\text{(h-esimo posto)}}{[0\dots 0\,1\,0\dots 0]^t}\right)$ implica che pure

$$\langle I(\theta)^{-1}\nabla_\theta\frac{\partial}{\partial\theta_h}f(\theta) ; \nabla_\theta\frac{\partial}{\partial\theta_h}f(\theta)\rangle = \langle e_h ; I(\theta)\cdot e_h\rangle = I(\theta)_{h,h} =$$
$$= \frac{\partial^2}{\partial(\theta_h)^2}f(\theta) \quad . \quad //$$

(ES.) Mettiamoci su un modello stat. $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\},\mu)$ dominato per un campione $(X_1,\dots,X_n)$ di taglia $n\in\mathbb{N}_+$ (scelta) ($n\geq 3$) e di legge μ^θ che sia esponenziale "$\mathcal{E}(\theta)$" di parametro $\theta>0$: dunque,

$$(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\},\mu) = \left((0,+\infty)^n, \mathcal{B}((0,+\infty)^n), \{(\mu^\theta)^{\otimes n}|\theta\in(0,+\infty)\}, \nu^{\otimes n}\right),$$

dove ν è di Lebesgue su $\mathcal{B}((0,+\infty))$, dove $\frac{d\mu^\theta}{d\nu}(x) = \theta e^{-\theta x}$ $\forall\theta,x>0$, e quindi $X_i(x_1,\dots,x_n) = x_i$. ✓ Ebbene, il modello è esponenziale "rispetto" a μ e a $T := -\sum_{i=1}^n X_i$, statistica anche completa, essendo $\underset{\text{(per indip.)}}{T\sim}$ "$-\Gamma(n,\theta)$", con $f(\theta) = -n\log\theta$ e con "matrice" dell'informazione secondo Fisher $I(\theta) = \frac{n}{\theta^2}$ (>0). ✓ Adesso, è chiaro che la statistica $-T = \sum_{i=1}^{n} X_i \sim \Gamma(n,\theta)$ resta esaustiva e completa per il modello, col secondo che perché la densità Gamma è pure chiaro che la statistica $\frac{1}{-T}$ sta in $\bigcap_{\theta>0} L^2(P^\theta)$: verifichiamo che, $\forall \theta \in \Theta$,

$E^{P^\theta}[\frac{n-1}{-T}] = \theta$ e $Var^{P^\theta}[\frac{n-1}{-T}] = \frac{\theta^2}{n-2}$, e deduciamo che lo stimatore corretto $U := \frac{n-1}{-T}$ (di quadrato integrabile) della funzione $g(\theta) := \theta$ (in $(0,+\infty)$) non è per ora efficace. Alle, infatti, finiti dei calcoli, osserviamo subito che $U = \frac{n-1}{-T}$ sarebbe ottimale fra tutti gli stimatori V di quadrato integrabili corretti per $g(\theta)=\theta$, essendo $-T$ esaustiva e completa (col teorema Blackwell-Rao), col allora per questi varrebbe $R_V(\theta) \geq R_U(\theta) = Var^{P^\theta}[U] = \frac{\theta^2}{n-2} \overset{(n>2)}{>} \frac{\theta^2}{n} =$

$= \frac{(g'(\theta))^2}{I(\theta)}$, da cui effetto non può esistere alcun stimatore di quadrato integrabile e corretto per g che sia anche efficace per g. ✓

Non resta che verificare i conti: $\forall \theta \in \Theta$, $E^{P^\theta}[\frac{1}{-T}] =$

$$= \int_0^{+\infty} \frac{1}{t} \frac{1}{(n-1)!} \theta^n t^{n-1} e^{-\theta t}\, dt \overset{(-T\sim\Gamma(n,\theta))}{=} \frac{\theta}{n-1} \int_0^{+\infty} \frac{1}{(n-2)!} \theta^{(n-1)} t^{(n-1)-1} e^{-\theta t}\, dt =$$

$= \frac{\theta}{n-1}$ ✓. Analogamente, $E^{P^\theta}[\frac{1}{T^2}] = \frac{\theta^2}{(n-1)(n-2)}$ $(\forall \theta \in \Theta)$ e allora $Var^{P^\theta}[\frac{n-1}{-T}] = (n-1)^2 \left\{ E^{P^\theta}[\frac{1}{T^2}] - E^{P^\theta}[\frac{1}{T}]^2 \right\} =$

$$= (n-1)^2 \underbrace{\left\{ \frac{\theta^2}{(n-1)(n-2)} - \frac{\theta^2}{(n-1)^2} \right\}}_{\left(\frac{\theta^2 \overbrace{[(n-1)-(n-2)]}^{(=1)}}{(n-1)^2(n-2)} \right)} = \frac{\theta^2}{n-2} \quad (\forall \theta \in \Theta)$$

, e quindi. □ ✓

Vogliamo, ormai, raffinare perfino la teoria delle stime, e farlo mediante il concetto di "stimatore di massima verosimiglianza", ma non prima di qualche osservazione conclusiva sull'informazione secondo Fischer di un modello stat. . L' $\boxed{\text{IDEA}}$ alla base di tali, ultime osservazioni è molto naturale: associare al modello stat. un "opportuno" concetto di "entropia" del modello stesso rispetto ai suoi parametri $\vartheta \in \Theta$, e cogliere un "legame stretto" fra l'informazione secondo Fischer del modello e la "variabilità" di questa entropia rispetto ai parametri $\vartheta \in \Theta$! Per ciò, prendiamola un po' alla larga cominciando col ragionare sul "grado di caoticità" di un qualsiasi modello stat. regolare rispetto ai suoi parametri $\vartheta \in \Theta$.

Sia pertanto $(\Omega, \mathcal{A}, \{P^\vartheta | \vartheta \in \Theta\}), \mu$ un generico modello stat. regolare, Θ infinito, dunque di verosimiglianza $L^\vartheta(\omega) := \frac{dP^\vartheta}{d\mu}(\omega) > 0$ per ogni $(\vartheta, \omega) \in \Theta \times \Omega$. In particolare, $\forall \vartheta_1, \vartheta_2 \in \Theta$, è $\frac{dP^{\vartheta_1}}{dP^{\vartheta_2}}(\omega) = \frac{L^{\vartheta_1}(\omega)}{L^{\vartheta_2}(\omega)}$. La questione è questa: in base al generico esito $\omega \in \Omega$ dell'esperimento statistico "di sottofondo", quanto riusciamo a discriminare due diversi parametri ϑ_1 e ϑ_2 in Θ nel senso di ritenere che uno sia più "giusto" dell'altro? In effetti, quanto più il modello è caotico, confuso, tanto più sarà difficile scegliere esattamente un parametro come "migliore" rispetto ogni altro, o avverrà per un modello facilmente riconoscibile o altrimenti riconoscibile). Per fornire delle nostre idee ad una definizione matematica tecnica effettiva, possiamo ragionare così: immaginando che $\{\omega\} \in \mathcal{A}$ $\forall \omega \in \Omega$, abbiamo

Chiaramente che $\mathcal{L}^{\theta}(\omega)\cdot\mu[\{\omega\}] = P^{\theta}[\{\omega\}]$ $\forall(\theta,\omega)\in\Theta\times\Omega$, da cui

$$\frac{\mathcal{L}^{\theta_1}(\omega)}{\mathcal{L}^{\theta_2}(\omega)} = \frac{P^{\theta_1}[\{\omega\}]}{P^{\theta_2}[\{\omega\}]} \quad \forall\,\theta_1,\theta_2\in\Theta \text{ e } \forall\omega\in\Omega,$$

e naturalmente che un tale

in effetti possiamo riconoscere tutte le possibili relazioni d'ordine : (in "$[0,+\infty]$")

$\forall\,\theta_1,\theta_2\in\Theta$ e $\forall\omega\in\Omega$,

$$\frac{\mathcal{L}^{\theta_1}(\omega)}{\mathcal{L}^{\theta_2}(\omega)} \text{ è } \begin{cases} = 1 \\ < 1 \\ \text{“}\downarrow 0\text{”} \\ > 1 \\ \text{“}\uparrow +\infty\text{”} \end{cases} \overset{\text{(ovvio)}}{\Longleftrightarrow} \text{ risp. } \begin{cases} P^{\theta_1}[\omega] = P^{\theta_2}[\omega] \\ P^{\theta_1}[\omega] < P^{\theta_2}[\omega] \\ \text{“}P^{\theta_1}[\omega]\downarrow 0\text{”} \quad (\text{variando } \theta_1) \\ P^{\theta_1}[\omega] > P^{\theta_2}[\omega] \\ \text{“}P^{\theta_2}[\omega]\downarrow 0\text{”} \quad (\text{variando } \theta_2) \end{cases},$$

da cui facilmente in base all'esito ω abbiamo una discriminazione fra θ_1 e θ_2 (associata a ω stesso). Precisamente, se $\frac{\mathcal{L}^{\theta_1}(\omega)}{\mathcal{L}^{\theta_2}(\omega)}$ è

$$\begin{cases} = 1 \\ < 1 \\ \text{“}= 0\text{”} \\ > 1 \\ \text{“}= +\infty\text{”} \end{cases}, \text{ allora, risp., } \boxed{\omega \text{ è tale che}} \begin{cases} \theta_1 \text{ e } \theta_2 \text{ non si distinguono} \\ \theta_2 \text{ sembra più “giusto” di } \theta_1 \\ \theta_2 \text{ è giusto rispetto a } \theta_1 \\ \theta_1 \text{ sembra più “giusto” di } \theta_2 \\ \theta_1 \text{ è giusto contro } \theta_2 \end{cases}.$$

Da tutto questo, l'idea "giusta" e naturale di entropia per il modello in questione : $\forall\omega\in\Omega$ e $\forall\theta_1,\theta_2\in\Theta$, "IL POTERE DISCRIMINANTE DI ω TRA (DI) θ_1 CONTRO θ_2" è il numero reale

$$\boxed{i(\theta_1/\theta_2)(\omega) := \log\frac{\mathcal{L}^{\theta_1}(\omega)}{\mathcal{L}^{\theta_2}(\omega)}} \quad \left(= \log\mathcal{L}^{\theta_1}(\omega) - \log\mathcal{L}^{\theta_2}(\omega)\right)$$

(mentre $i(\theta_1/\theta_2)(\cdot)$ è una v.a.), mentre "L'INFORMAZIONE DI KULLBACK, O ENTROPIA, DI θ_1 CONTRO θ_2 PER IL MODELLO" è il numero (in $[0,+\infty]$) dato calcolando

$$\boxed{I(\theta_1/\theta_2) := \mathbb{E}^{P^{\theta_1}}[i(\theta_1/\theta_2)] = \mathbb{E}^{P^{\theta_1}}\left[\log\frac{\mathcal{L}^{\theta_1}}{\mathcal{L}^{\theta_2}}\right]}.$$

In effetti, osservando questo fatto, se $p^{\theta_1}[\omega] \approx p^{\theta_2}[\omega]$ allora ω ha potere ~~discriminante~~ pressoché nullo nel discriminare θ_1 da θ_2, e viceversa; invece, quanto più $p^{\theta_1}[\omega] < p^{\theta_2}[\omega]$, tanto più ω ha potere negativo nel discriminare θ_1 contro θ_2, e viceversa, come a dire che θ_1 è tanto più peggiore rispetto a θ_2; infine, al contrario, quanto più $p^{\theta_1}[\omega] > p^{\theta_2}[\omega]$ tanto più ω discrimina favorevolmente θ_1 da θ_2, e viceversa.

NOTA Nessuna sorpresa a riguardo della quasi-simmetria delle situazioni (in θ), essendo $\log \frac{L^{\theta_1}(\omega)}{L^{\theta_2}(\omega)} = -\log \frac{L^{\theta_2}(\omega)}{L^{\theta_1}(\omega)}$, e così (per definizione)

$$i(\theta_1/\theta_2)(\omega) = -i(\theta_2/\theta_1)(\omega).$$

A questo punto, per una misura di "bontà" del modello nei parametri, una media "opportuna" su tutto Ω di tali poteri discriminanti, come è naturale provare a fare.

OSS. Dunque, $I(\theta_1/\theta_2) = \mathbb{E}^{p^{\theta_1}}[\log L^{\theta_1}] - \mathbb{E}^{p^{\theta_1}}[\log L^{\theta_2}] =$

$$= -\text{"}\mathbb{E}^{\mu}\text{"}[(\log L^{\theta_2}) L^{\theta_1}] + \mathbb{E}^{p^{\theta_1}}[\log L^{\theta_1}]$$

: come forse più adesso può risultare intuitivo, risulta interessante per l'analisi la funzione $\boxed{\theta \mapsto I(\theta_1/\theta)}$, $\Theta \to [0,+\infty]$, per ogni $\theta_1 \in \Theta$ fissato, funzione "intera", più che ogni scalare $I(\theta_1/\theta_2)$, $\theta_2 \in \Theta$. ($\blacktriangleright$ Potremmo intendere "entropia del modello in θ_1" proprio come la funzione

$$\theta \mapsto I(\theta_1/\theta) = \mathbb{E}^{p^{\theta_1}}\left[\log \frac{L^{\theta_1}}{L^{\theta}}\right].)$$

! NON vale $I(\theta_1/\theta_2) = -I(\theta_2/\theta_1)$, perché $I(\theta_2/\theta_1) = \mathbb{E}^{p^{\theta_2}}\left[\log \frac{L^{\theta_2}}{L^{\theta_1}}\right] =$

$$= -\mathbb{E}^{p^{\theta_2}}\left[\log \frac{L^{\theta_1}}{L^{\theta_2}}\right] \neq -\mathbb{E}^{p^{\theta_1}}\left[\log \frac{L^{\theta_1}}{L^{\theta_2}}\right]!$$

Comunque, (per tutta una serie di motivi) saremo davvero soddisfatti da queste definizioni, e allora senza ulteriori indugi procediamo con la tecnica, cominciando con alcune col cominciare che, $\forall \theta_1,\theta_2 \in \Theta$, $I(\theta_1/\theta_2) \doteq E^{P^{\theta_1}}[\log \frac{L^{\theta_1}}{L^{\theta_2}}]$ sia davvero un numero ≥ 0 ed eventualmente $+\infty$.

Breve digressione. Sia $(\Omega,\mathcal{A},P)$ uno sp. di prob., e sia Y una v.a. reale su di esso. Allora $E^P[Y] > -\infty$ se, e solo se, $E^P[Y^-] < +\infty$.

[Questo perché $Y = Y^+ - Y^-$, con sia Y^+ sia $Y^- \geq 0$ su Ω (P-q.c.), e allora, in caso di sommabilità, $E^P[Y] = \underbrace{E^P[Y^+]}_{(\geq 0)} - \underbrace{E^P[Y^-]}_{(\geq 0)}$. ✓]

Sia quindi X una v.a. in $\mathcal{L}^1(P)$, a valori in un intervallo aperto $I \subseteq \mathbb{R}$, per cui $X:\Omega\to I$, e sia $\phi: I \to \mathbb{R}$ una qualsiasi funzione convessa. Allora $E^P[\phi(X)^-] < +\infty$ (o cioè $E^P[\phi(X)] > -\infty$).

[Infatti, per convessità, per ogni $x_0 \in I$ esiste $k(x_0) \in \mathbb{R}$ tale che sia

$\phi(X) - \phi(x_0) \geq k(x_0)(X - x_0)$, ossia $\phi(X) \geq k(x_0)X + (\phi(x_0) - x_0 k(x_0))$.

Allora, adesso, osservando/ricordando che, $\forall a,b \in \mathbb{R}$, vale che $a \leq b \Rightarrow b^- \leq a^-$, e che $(a+b)^- \leq a^- + b^-$ (essendo $a^- = \max\{0,-a\}$, e b^- analogo), deduciamo subito che, $\forall x_0 \in I$, vale

$$\phi(X)^- \leq (k(x_0)X)^- + (\phi(x_0) - x_0k(x_0))^- \overset{(a^- \leq |a|)}{\leq}$$
$$\leq |k(x_0)||X| + |\text{"costante in } x_0\text{"}|$$

da cui subito la tesi. □]

DUNQUE, ⇒ Essendo $\phi(x) := -\log(x) : I := (0,+\infty) \to \mathbb{R}$ convessa, ed essendo ($\forall \theta_1,\theta_2 \in \Theta$ fissati) $X(\omega) := \frac{L^{\theta_2}(\omega)}{L^{\theta_1}(\omega)} : \Omega \to I = (0,+\infty)$ una v.a. in $\mathcal{L}^1(P^{\theta_1})$, con esattamente $E^{P^{\theta_1}}[X] = E^{P^{\theta_1}}[\frac{L^{\theta_2}}{L^{\theta_1}}] = E^{P^{\theta_2}}[1] = 1$ (visto che "$dP^{\theta_1}(\omega) = \frac{L^{\theta_1}(\omega)}{L^{\theta_2}(\omega)} \cdot dP^{\theta_2}(\omega)$"), otteniamo intanto che $E^{P^{\theta_1}}[(-\log\frac{L^{\theta_2}}{L^{\theta_1}})^-] = E^{P^{\theta_1}}[(\log\frac{L^{\theta_1}}{L^{\theta_2}})^-] < +\infty$, cioè che $I(\theta_1/\theta_2) = \mathbb{E}^{p^{\theta_1}}\left[\log \frac{L^{\theta_1}}{L^{\theta_2}}\right] > -\infty$. ✓ (Rimarchiamo che ovviamente senato (bisogno di speranza per $i(\theta_1/\theta_2)$!) A questo punto che ha senso $I(\theta_1/\theta_2)$, verifichiamo che è ≥ 0 : in realtà è sostanzialmente già dimostrato, perché potendo ora usare Jensen abbiamo

$$I(\theta_1/\theta_2) = \mathbb{E}^{p^{\theta_1}}\left[-\log \frac{L^{\theta_2}}{L^{\theta_1}}\right] \overset{(J.)}{\geq} -\log\Big(\underbrace{\mathbb{E}^{p^{\theta_1}}\left[\frac{L^{\theta_2}}{L^{\theta_1}}\right]}_{=(1)}\Big) = 0 \quad . ✓$$

$\boxed{\text{OSS.}}$ Da appunto a questo, $I(\theta_1/\theta_2) = 0$ se, e solo se, $p^{\theta_1} = p^{\theta_2}$ (cioè $L^{\theta_1} = L^{\theta_2}$ p^{θ_1}-q.c.). (Infatti, guardando proprio Jensen di sopra, se $I(\theta_1/\theta_2) = 0$ allora $\frac{L^{\theta_2}}{L^{\theta_1}}$ è costante p^{θ_1}-q.c., ossia

$$\frac{L^{\theta_2}}{L^{\theta_1}} \overset{(p^{\theta_1}\text{-q.c.})}{=} \mathbb{E}^{p^{\theta_1}}\left[\frac{L^{\theta_2}}{L^{\theta_1}}\right] \overset{(\text{visto})}{=} 1 \quad . ✓)$$

(ES.) Supponiamo che il modello nel quale stiamo lavorando sia il "solito" modello esponenziale "completo", nel senso che fissato $k \in \mathbb{N}_+$, e $T : \Omega \to \mathbb{R}^k$ mis., tale che $\Theta \subseteq \mathbb{R}^k$ sia un aperto convesso ($\neq \emptyset$) tale che, $\forall \theta \in \Theta$, sia

$0 < \int_\Omega \exp(\langle \theta ; T(\omega) \rangle) d\mu(\omega) < +\infty$, e tale che $T(\mu)$ sia completa su $\mathcal{B}(\mathbb{R}^k)$: posto allora $f(\theta) = \log\left(\int_\Omega \exp(\langle \theta ; T(\omega) \rangle) d\mu(\omega)\right)$ $\forall \theta \in \Theta$, supponiamo che $L^\theta(\omega) = \exp\left(\langle \theta ; T(\omega) \rangle - f(\theta)\right)$ $\forall (\theta, \omega) \in \Theta \times \Omega$. In particolare, il modello è regolare e come tale hanno senso il potere discriminante e, quindi, l'entropia: precisamente, $\forall \omega \in \Omega$ e $\forall \theta_1, \theta_2 \in \Theta$, è $i(\theta_1/\theta_2)(\omega) = \log \frac{L^{\theta_1}(\omega)}{L^{\theta_2}(\omega)} = f(\theta_2) - f(\theta_1) - \langle T(\omega) ; \theta_2 - \theta_1 \rangle$ e quindi, ricordando che $\mathbb{E}^{p^{\theta_1}}[T_h] = \frac{\partial}{\partial \theta_h} f(\theta_1)$ $\forall h = 1, \dots, k$,

$$I(\theta_1/\theta_2) = \mathbb{E}^{p^{\theta_1}}[i(\theta_1/\theta_2)] = f(\theta_2) - f(\theta_1) - \langle \nabla_\theta f(\theta_1) ; \theta_2 - \theta_1 \rangle .$$

✓

Inoltre, un tale modello ammette la matrice d'informazione secondo Fischer in ogni $\theta\in\Theta$, che coincide con $I(\theta) = \mathcal{D}^2 f(\theta)$ $\forall\theta\in\Theta$, cioè con la matrice hessiana di f calcolata in θ: insomma, $\forall\theta\in\Theta$ e $\forall i,j=1,\dots,k$, è $I(\theta)_{i,j} = \frac{\partial^2}{\partial\theta_i\partial\theta_j} f(\theta)$. Ebbene, osserviamo meglio $\theta \mapsto I(\theta_1/\theta)$ per ogni $\theta_1\in\Theta$: $\forall\theta\in\Theta$, $I(\theta_1/\theta) =$

$= f(\theta) - f(\theta_1) - \langle\nabla_\theta f(\theta_1); \theta-\theta_1\rangle = f(\theta) +$ «un'applicazione lineare affine $\Theta\to\mathbb{R}$» $(\Rightarrow C^\infty \text{ in } \theta)$. Allora è palese che $\frac{\partial^2}{\partial\theta_i\partial\theta_j} I(\theta_1/\theta) = I(\theta)_{i,j}$ $\forall\theta\in\Theta$ e $\forall i,j=1,\dots,k$, cioè che $I(\theta) = \mathcal{D}^2 I(\theta_1/\theta)$ $\forall\theta\in\Theta$. /

Sarà vero in generale ?? — Sì!

Teorema. Supponiamo che per $(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in\Theta\}, \mu)$ (regolare) sia ben definita la matrice d'informazione secondo Fischer in ogni $\theta\in\Theta$: dunque, a parte $L^\theta(\omega)>0$ "sempre", richiediamo che $\theta\mapsto L^\theta(\omega)$ sia in $C^2(\Theta)$ $\forall\omega\in\Omega$, che L^θ, $\frac{\partial}{\partial\theta_i}L^\theta$, $\log L^\theta$, $\frac{\partial}{\partial\theta_i}\log L^\theta$ stiano in $L^2(P^\theta)$ $\forall\theta\in\Theta$ e $\forall i=1,\dots,k$, e che $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L^\theta$, $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L^\theta$ stiano in $L^1(P^\theta)$ $\forall\theta\in\Theta$ e $\forall i,j=1,\dots,k$. / Se quindi $I(\theta) := \mathrm{Cov}^{P^\theta}[\nabla_\theta \log L^\theta]$, $\theta\in\Theta$, è la matrice d'informazione secondo Fischer, allora vale che $\theta\mapsto I(\theta_1/\theta)$ sia in $C^2(\Theta)$ per ogni $\theta_1\in\Theta$, e che ($\forall\theta_1\in\Theta$ e $\forall i,j=1,\dots,k$)

$$\boxed{I(\theta_1)_{i,j} = \frac{\partial^2}{\partial\theta_i\partial\theta_j} I(\theta_1/\theta)\Big|_{\theta=\theta_1}} \quad \left(\text{cioè } I(\cdot) = \mathcal{D}^2 I(\theta_1/\cdot)\right).$$

Dim. Abbiamo che $I(\theta_1/\theta) = -\mathbb{E}^{P^{\theta_1}}[\log L^\theta] + \mathbb{E}^{P^{\theta_1}}[\log L^{\theta_1}]$ $\forall\theta_1,\theta\in\Theta$, dove solo l'addendo $-\mathbb{E}^{P^{\theta_1}}[\log L^\theta] = -\int_\Omega (\log L^\theta)\cdot L^{\theta_1}\,d\mu$ dipende da $\theta\in\Theta$, e di questo sappiamo ormai derivarne due volte (grazie alle ipotesi) $\forall i,j=1,\dots,m$: $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(-\int_\Omega (\log L^\theta) L^{\theta_1} d\mu\right) = -\int_\Omega \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L^\theta \cdot L^{\theta_1} d\mu$,

e per cui, chiaramente, $\frac{\partial^2}{\partial\theta_i\partial\theta_j} I(\theta_1/\theta) = -\mathbb{E}^{P^{\theta_1}}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L^\theta\right]$.

Dunque la tesi, perché allora $\frac{\partial^2}{\partial\theta_i\partial\theta_j} I(\theta_1/\theta)\Big|_{\theta=\theta_1} =$

$$= -\mathbb{E}^{P^{\theta_1}}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L^\theta\Big|_{\theta=\theta_1}\right] \overset{(\text{visto})}{=} I(\theta_1)_{ij}, \quad \forall i,j=1,\dots,m. \quad \square$$

OSS Analogamente, è evidente che $\nabla_\theta I(\theta_1/\theta)\Big|_{\theta=\theta_1} = 0_{\mathbb{R}^m}$, e comunque è coerente col fatto che $\theta=\theta_1$ è di minimo globale per $I(\theta_1/\theta)$ (essendo $I(\theta_1/\theta_1)=0$), ed infatti è anche $D^2 I(\theta_1/\theta)\Big|_{\theta=\theta_1} \overset{(\text{Teo.})}{=}$ $= I(\theta_1)$ che è semi-definita positiva. Sviluppiamo con Taylor al 2° ordine, e con resto di Peano, $I(\theta_1/\theta)$ per $\theta\approx\theta_1$:

$$I(\theta_1/\theta) \underset{(\theta=\theta_1+(\theta-\theta_1))}{=} \underbrace{I(\theta_1/\theta_1)}_{(0)} + \underbrace{\langle\nabla_\theta I(\theta_1/\theta)\Big|_{\theta=\theta_1}}_{(0)}; \theta-\theta_1\rangle +$$

$$+\frac{1}{2}\langle \underbrace{D^2 I(\theta_1/\theta)\Big|_{\theta=\theta_1}}_{I(\theta_1)}\cdot(\theta-\theta_1); \theta-\theta_1\rangle + \underset{\theta\to\theta_1}{o}(\|\theta-\theta_1\|^2),$$

da cui insomma $\boxed{I(\theta_1/\theta) \underset{\theta\to\theta_1}{\sim} \frac{1}{2}\langle I(\theta_1)\cdot(\theta-\theta_1); \theta-\theta_1\rangle}$, il che dà legami più intuitivi fra $I(\theta_1)$ e $I(\theta_1/\theta)$, $\theta\approx\theta_1$.

NOTA Come per l'informazione di Fisher, anche per l'entropia si ha la "somma delle entropie nel caso di modello prodotto "indipendente"". Più precisamente, mettiamoci nel modello $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\},\mu)$ per un campione di uso $(X_1,\dots,X_n)$ di legge μ^θ su $\mathcal{B}(\mathbb{R})$, $\theta\in\Theta$, in modo che sia tutto regolare: dunque $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\},\mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mu^{\theta\otimes n}|\theta\in\Theta\}, \nu^{\otimes n})$ dove ν è di Lebesgue 1-dim., $\frac{d\mu^\theta}{d\nu}(x) > 0$ per ogni $\theta\in\Theta$ e $x\in\mathbb{R}$, e

$X_i(x_1,\dots,x_n) = x_i$. / Allora $L^\theta(x_1,\dots,x_n) = \prod_{i=1}^{n} \frac{d\mu^\theta}{d\nu}(x_i)\ (>0)$, per cui

$\log L^\theta(x_1,\dots,x_n) = \sum_{i=1}^{n} \log \frac{d\mu^\theta}{d\nu}(x_i)$ e quindi, $\forall \theta_1,\theta_2\in\Theta$ (e $x_i\in\mathbb{R}$),

$$\log \frac{L^{\theta_1}(x_1,\dots,x_n)}{L^{\theta_2}(x_1,\dots,x_n)} = \log L^{\theta_1}(\omega) - \log L^{\theta_2}(\omega) = \sum_{i=1}^{n}\left[\log \frac{d\mu^{\theta_1}}{d\nu}(x_i) - \log \frac{d\mu^{\theta_2}}{d\nu}(x_i)\right],$$

da cui subito $I(\theta_1/\theta_2) := E^{P^{\theta_1}}\left[\log \frac{L^{\theta_1}}{L^{\theta_2}}\right] = \sum_{i=1}^{n} E^{P^{\theta_1}}\left[\log \frac{d\mu^{\theta_1}}{d\nu}(X_i) - \log \frac{d\mu^{\theta_2}}{d\nu}(X_i)\right] =$

$$= \sum_{i=1}^{n} \underbrace{E^{\mu^{\theta_1}}\left[\log \frac{d\mu^{\theta_1}}{d\nu} - \log \frac{d\mu^{\theta_2}}{d\nu}\right]}_{=:\, I_0(\theta_1/\theta_2)} = n\, I_0(\theta_1/\theta_2) \quad . \quad \surd$$

(analoga del modello "fattore")

Veniamo finalmente al concetto di stimatore di massima verosimiglianza, concetto perfettamente intuitivo e comprensibile dopo quelli di potere discriminante ed entropia ragionato poc'anzi: l'idea base, infatti, è quella di trovare (se possibile) per ogni $\omega\in\Omega$ un parametro $\hat\theta = \hat\theta(\omega)\in\Theta$ (possibilmente unico) tale che $\boxed{L^{\hat\theta}(\omega) = \max_{\theta\in\Theta} L^\theta(\omega)}$, o (cioè) tale che $L^{\hat\theta}(\omega) \geq L^\theta(\omega)$ per ogni $\theta\in\Theta$, in virtù dell'ovvio fatto che in tal caso, se ω davvero si verifica, allora "dal punto di vista di ω" il parametro $\hat\theta$ è più "giusto" rispetto ad ogni altro $\theta\in\Theta$. / In effetti, questa relazione di massima verosimiglianza equivale alla relazione $\log L^{\hat\theta}(\omega) = \max_{\theta\in\Theta,\ \text{t.c. } L^\theta(\omega)>0} \log L^\theta(\omega)$, o cioè a $\log L^{\hat\theta}(\omega) - \log L^\theta(\omega) \geq 0$ $\forall\theta$, o cioè a $i(\hat\theta/\theta)(\omega) \geq 0$ $\forall\theta$: $\hat\theta$ ragionevolmente più adatto di ogni altro θ, e tanto più adatto quanto più $L^{\hat\theta}(\omega) \gg L^\theta(\omega)$, ossia $i(\hat\theta/\theta)(\omega) \gg 0$. / Generalizziamo:

dato un modello stat. dominato $(\Omega, \mathcal{A}, \{P^\vartheta | \vartheta \in \Theta\}, \mu)$, ovvio regolare, dove $\Theta \subseteq \mathbb{R}$ è un intervallo aperto $(\neq \emptyset)$, e considerato uno stimatore $U : \Omega \to \Theta$ del parametro ϑ (che come sempre stia in $\bigcap_{\vartheta \in \Theta} L^2(P^\vartheta)$), tale U è "UNO STIMATORE DI MASSIMA VEROSIMIGLIANZA (DI ϑ)" se, $\forall \omega \in \Omega$, $L^{U(\omega)}(\omega) = \max_{\vartheta \in \Theta} L^\vartheta(\omega)$ $(U(\omega) \in \Theta)$, ed in tal caso lo indichiamo con $\hat{\vartheta}(\omega) \doteq U(\omega)$. Dunque si tratta di trovare una v.a. $\hat{\vartheta}(\cdot)$ a valori in Θ (e di quadrato integrabile) tale che, per ogni $\omega \in \Omega$, risulti $L^{\hat{\vartheta}(\omega)}(\omega) = \max_{\vartheta \in \Theta} L^\vartheta(\omega)$, o equivalentemente

$$\log L^{\hat{\vartheta}(\omega)}(\omega) = \max_{\vartheta \in \Theta} \log L^\vartheta(\omega) .$$

$\boxed{\text{SITUAZIONE IDEALE}}$: che $\vartheta \mapsto L^\vartheta(\omega)$ sia in $C^2(\Theta)$ per ogni $\omega \in \Omega$, o cioè che $\vartheta \mapsto \log L^\vartheta(\omega)$ sia in $C^2(\Theta)$ $\forall \omega \in \Omega$. In questo caso, infatti, $\forall \omega \in \Omega$, cerchiamo fra quegli $\hat{\vartheta} = \hat{\vartheta}(\omega)$ di Θ tali che

$$\frac{d}{d\vartheta} L^\vartheta(\omega) \Big|_{\vartheta=\hat{\vartheta}} = 0 \quad \text{e} \quad \frac{d^2}{d\vartheta^2} L^\vartheta(\omega) \Big|_{\vartheta=\hat{\vartheta}} \leq 0 \quad , \text{o}$$

equivalentemente $\frac{d}{d\vartheta} \log L^\vartheta(\omega) \Big|_{\vartheta=\hat{\vartheta}} = 0$ e $\frac{d^2}{d\vartheta^2} \log L^\vartheta(\omega) \Big|_{\vartheta=\hat{\vartheta}} \leq 0$,

e "poi vediamo". Ancora più ideale sarebbe che $\vartheta \mapsto L^\vartheta(\omega)$, oppure $\vartheta \mapsto \log L^\vartheta(\omega)$, sia CONCAVA $\forall \omega \in \Omega$, così che un eventuale punto di massimo locale $\hat{\vartheta} = \hat{\vartheta}(\omega)$ sia automaticamente punto di massimo globale, cioè quello che vogliamo!

$\boxed{\text{OSS.}}$ Tendiamo a preferire la scrittura logaritmica sempre pensando ai modelli per i campioni. Infatti, previamente, se scegliamo mod.,

$\Theta \subseteq \mathbb{R}$ intervallo aperto ($\neq \emptyset$), e $\forall \theta \in \Theta$ una legge μ^θ su $\mathcal{B}(\mathbb{R})$ che sia equivalente alla misura di Lebesgue λ 1-dim., allora ponendo che sia $(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in\Theta\}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{(\mu^\theta)^{\otimes n} \mid \theta \in \Theta\}, \lambda^{\otimes n})$, e se $X_i(x_1,\dots,x_n) = x_i$ ($\forall i=1,\dots,n$ e $\forall x_1,\dots,x_n\in\mathbb{R}$), allora il nostro modello è per il campione $(X_1,\dots,X_n)$ di legge μ^θ, $\theta\in\Theta$, ed ha

$L^\theta(x_1,\dots,x_n) = \prod_{i=1}^n \frac{d\mu^\theta}{d\lambda}(x_i)$, per cui $\log L^\theta(x_1,\dots,x_n) = \sum_{i=1}^n \log\frac{d\mu^\theta}{d\lambda}(x_i)$ ed il problema della massima verosimiglianza diventa, $\forall x_1,\dots,x_n\in\mathbb{R}$, si cerchi di trovare $\hat\theta = \hat\theta(x_1,\dots,x_n)\in\Theta$ tale che

$$\frac{d}{d\theta}\log L^\theta(x_1,\dots,x_n)\Big|_{\theta=\hat\theta} = \sum_{i=1}^n \frac{d}{d\theta}\log\frac{d\mu^\theta}{d\lambda}(x_i)\Big|_{\theta=\hat\theta} = 0 \quad \text{e}$$

(analogamente) $$\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\frac{d\mu^\theta}{d\lambda}(x_i)\Big|_{\theta=\hat\theta} \le 0 .$$ //

ES. Mettiamoci in un modello del esponenziale "completo" (ma con $k=1$): supponiamo che μ e $\Theta\subseteq\mathbb{R}$ siano tali che esiste $T:(\Omega,\mathcal{A})\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mis. con $T(\mu)$ σ-finita su $\mathcal{B}(\mathbb{R})$ in modo che risulti, $\forall\theta\in\Theta$,

$0<\int_\Omega e^{\theta T(\omega)}\,d\mu(\omega)<+\infty$, e quindi, se poniamo $f(\theta) := \log\left(\int_\Omega e^{\theta T(\omega)}\, d\mu(\omega)\right)$ $\forall\theta\in\Theta$, allora supponiamo che $L^\theta(\omega) = \exp(\theta T(\omega) - f(\theta))$ per ogni $(\theta,\omega)\in\Theta\times\Omega$. Dunque T ammette momenti di ogni ordine, e f è in $C^\infty(\Theta)$, ed anche $f'(\theta) = E^{P^\theta}[T]$ e $f''(\theta) = Var^{P^\theta}[T]$ (≥ 0) per ogni $\theta\in\Theta$. In particolare, f è convessa, ossia $-f$ è concava, e allora $\log L^\theta(\omega) = \theta T(\omega) - f(\theta)$ è in $C^\infty(\Theta)$ e concava su Θ, $\forall\omega\in\Omega$. Pertanto, in questa situazione, il problema della massima verosimiglianza equivale al seguente: per ogni $\omega\in\Omega$, trovare

$\hat{\vartheta} = \hat{\vartheta}(\omega)$ in Θ tale che $\frac{d}{d\vartheta} \log \mathcal{L}^{\vartheta}(\omega)\big|_{\vartheta=\hat{\vartheta}} = 0$, ed in caso di riuscita dimostrare che $\omega \mapsto \hat{\vartheta}(\omega)$ è stimatore. Dunque l'equazione è $\boxed{f'(\hat{\vartheta}(\omega)) = T(\omega)}$, e per ciò: se in realtà T è a valori in $f'(\Theta)$, cioè se $T : \Omega \to \varphi'(\Theta)$, e se $f' : \Theta \to \mathbb{R}$ è iniettiva, e dunque invertibile considerando $f'(\Theta)$ come codominio (automaticamente aperto, fra l'altro, poiché φ' è continua ("TEO. DELLA MAPPA APERTA")), allora esiste uno ed un solo $\hat{\vartheta}$ ed è $\hat{\vartheta}(\omega) = (f')^{-1}(T(\omega))$ ($\Rightarrow$ è v.a.!)

Naturalmente, gli stimatori di massima verosimiglianza potrebbero benissimo non esistere per un dato modello stat., ed inoltre, anche ammesso di trovarne uno, potrebbero benissimo non esser corretti (come stimatori di ϑ).

Tuttavia, noi ci concentreremo solo sui modelli esponenziali "completi" per i quali "tutto torna" ($\rightarrow$ vedi ES. precedente $\leftarrow$), ed anzi sui modelli esponenziali completi che siano prodotti di modelli esponenziali completi, studiando la correttezza degli stimatori di massima verosimiglianza.

(ES.) Riprendiamo il modello per un campione $(X_1, \dots, X_n)$ di legge $\mu^{\vartheta} = \mathcal{E}(\vartheta)$, $\vartheta > 0$: dunque $(\Omega, \mathcal{A}, \{P^{\vartheta} | \vartheta \in \Theta\}, \mu) = ((0,+\infty)^n, \mathcal{B}((0,+\infty)^n), \{(\mu^{\vartheta})^{\otimes n} | \vartheta \in \Theta\}, \nu^{\otimes n})$ con ν di Lebesgue su $(0,+\infty)$, con $\frac{d\mu^{\vartheta}}{d\nu}(x) = \vartheta e^{-\vartheta x}$ $\forall \vartheta > 0, x > 0$, e quindi $X_i(x_1, \dots, x_n) = x_i$ e $\mathcal{L}^{\vartheta}(x_1, \dots, x_n) = \vartheta^n e^{-\vartheta \sum_{i=1}^{n} x_i} =$
$= \exp(-\vartheta \sum_{i=1}^{n} x_i + n \log \vartheta)$. Per ciò, nelle notazioni dell'ES. precedente, $T(x_1, \dots, x_n) = \sum_{i=1}^{n} X_i(x_1, \dots, x_n)$ e $f(\vartheta) = -n \log \vartheta$, quindi

$\psi'(\theta) = -\frac{n}{\theta}$: $(0,+\infty) \to (-\infty,0)$ è bigettiva, e $T(\omega)$ è a valori <0, per cui l'equazione di massima verosimiglianza $-\frac{n}{\theta} = T$ è risolubile in uno ed un sol modo, con $\boxed{\hat{\theta} = \frac{n}{-T}}$. Questo stimatore però non è corretto (per θ) perché (come calcolato ultimamente) $-T \sim \Gamma(n,\theta)$ e allora è $\frac{n-1}{-T}$ (ad essere) corretto per θ, per cui $E^{P^\theta}[\hat{\theta}] = \frac{n}{n-1}\theta \quad \forall \theta \in \Theta$.

Tuttavia, ed allora dico che $\hat{\theta}$ "tende" a diventare corretto, nel senso che è "asintoticamente corretto", appunto perché $\forall\theta\in\Theta$ $E^{P^\theta}[\hat{\theta}] \xrightarrow[n\to+\infty]{} \theta$.
Questo non è nemmeno troppo sorprendente ... La speranza è che questo converga più in generale, motivo per il quale più avanti terremo a sottolineare la dipendenza di $\hat{\theta}$ da n, scrivendo "$\hat{\theta}_n(\omega)$".

NOTA ! Supponiamo di riconsiderare l'Es. precedente sostituendo $\Theta = (0,+\infty)$ con, ad esempio, $\Theta = (0,1)$: allora, $\forall\omega\in\Omega$ scelto, NON è affatto detto che $\frac{n}{-T(\omega)} \in (0,1)$ ...! Però $-T(\omega) = \sum_{i=1}^{n} X_i(\omega)$, dove $X_1,\dots,X_n$ sono tutte i.i.d. (ovviamente) in $L^1(P^\theta)$, $\forall\theta\in\Theta$, per cui (per la legge forte dei grandi numeri di Kolmogorov) è $\frac{1}{n}(-T) \xrightarrow[n\to+\infty]{P^\theta\text{-q.c.}} E^{P^\theta}[X_1] \overset{(\text{legge Exp.})}{=} \frac{1}{\theta}$ $\forall\theta\in\Theta$, ossia chiaramente $\frac{n}{-T} \xrightarrow[n\to+\infty]{P^\theta\text{-q.c.}} \theta \quad \forall\theta\in\Theta$, e θ sta nell'aperto $(0,1)$: di conseguenza, $\forall\theta\in\Theta$ e per P^θ-quasi ogni $\omega\in\Omega$, per $n \geq$ "$n(\theta,\omega)$" è $\frac{n}{-T(\omega)} \in (0,1)$, per cui (per tali ω e tali n) $\hat{\theta}_n(\omega) = \frac{n}{-T(\omega)}$! Riassumendo : $\forall\theta\in\Theta$, $\hat{\theta}_n(\cdot)$ resta ben definito P^θ-q.c. e solo da un certo n in poi, e per tali resta uguale a $\frac{n}{-T}$ e allora certamente $\xrightarrow[n\to+\infty]{P^\theta\text{-q.c.}} \theta$.

Facciamo qualche altro semplice esercizio (come di attinenza con la teoria).

(ES.) Mettiamoci nel modello stat. canonico per un campione $(X_1,\dots,X_n)$ di legge $\mu^\theta = \Pi_\theta$, $\theta>0$: dunque $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\},\mu) =$

$= (\mathbb{N}^n, \mathcal{B}(\mathbb{N}^n), \{\mu^{\theta\otimes n}|\theta\in(0,+\infty)\}, \nu^{\otimes n})$, dove ν conta i punti di $\mathbb{N}$, e dove $\frac{d\mu^\theta}{d\nu}(k) = e^{-\theta}\frac{\theta^k}{k!}$ $\forall\theta>0, \forall k\in\mathbb{N}$, per cui $X_i(k_1,\dots,k_n)=k_i$ e

$$L^\theta(k_1,\dots,k_n) = \frac{1}{k_1!\cdots k_n!}\, e^{-n\theta}\,\theta^{\sum_{i=1}^n k_i} = \frac{1}{k_1!\cdots k_n!}\exp\Big(\log(\theta)\cdot\sum_{i=1}^n k_i - n\theta\Big).$$

Per cui, posto $T \doteq \sum_{i=1}^n X_i$, è chiaro che il problema della massima verosimiglianza è quello di trovare, $\forall\omega=(k_1,\dots,k_n)\in\Omega$, un $\hat\theta=\hat\theta(\omega)>0$ tale che $\frac{1}{\hat\theta(\omega)}T(\omega) - n = 0$: cioè $\hat\theta(\omega) = \frac{T}{n} \left(= \overline{X}\right)$.

In questo caso, per di più, $\hat\theta$ è uno stimatore anche corretto di $\theta\in\Theta$, in quanto $E^{\mu^\theta}[X_1] \overset{(\text{Poisson})}{=} \theta$ $\forall\theta\in\Theta$. ✓

(ES.) Mettiamoci nel modello stat. canonico per un campione $(X_1,\dots,X_n)$ di legge μ^θ su $\mathcal{B}(\mathbb{R})$ avente densità $f^\theta(x)\doteq(\theta+1)x^\theta 1_{(0,1)}(x)$, $x\in\mathbb{R}$, $\theta>-1$, rispetto alla Lebesgue: dunque, $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\},\mu)=$

$= ((0,1)^n, \mathcal{B}((0,1)^n), \{\mu^{\theta\otimes n}|\theta\in(-1,+\infty)\}, \nu^{\otimes n})$ dove ν è di Lebesgue su $(0,1)$ e dove $\frac{d\mu^\theta}{d\nu}(x) = f^\theta(x) = (\theta+1)x^\theta$ $\forall\theta>-1, \forall x\in(0,1)$, per cui $X_i(x_1,\dots,x_n)=x_i$ e $L^\theta(x_1,\dots,x_n) = (\theta+1)^n\left(\prod_{i=1}^n x_i\right)^\theta =$

$= \exp\Big(\theta\cdot\sum_{i=1}^n \log(x_i) + n\log(\theta+1)\Big)$. Per cui, se $T \doteq \sum_{i=1}^n \log(X_i)$ (<0), allora, $\forall\omega=(x_1,\dots,x_n)\in\Omega$, l'equazione di massima verosimiglianza è

$T(\omega) + \frac{n}{\hat\theta(\omega)+1} = 0$, da cui $\hat\theta(\omega) = \frac{n}{-T(\omega)} - 1$ (>-1).

Ricordando adesso che, come visto, $\forall\theta\in\Theta$ si ha

$\mathbb{E}^{\mu^\theta}\left[-\frac{T}{m}\right] = \frac{1}{\theta+1}$, allora non sarà $\mathbb{E}^{\mu^\theta}\left[-\frac{m}{T}\right] = \theta+1$, onde $\hat{\theta}$ non sarà corretto per θ (quanto, tuttavia, asintoticamente corretto (grazie poi a Kolmogorov)). ✓ ✓

ES. Mettiamoci, infine, nel modello stat. canonico per un campione $(X_1,\dots,X_m)$ di legge "$\mu^\theta = N(m,\sigma^2)$", prima con $\theta = m \in \mathbb{R}$ e poi con $\theta = \sigma^2 > 0$: dunque, $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mu^{\theta\otimes n} \mid \theta \in \Theta\}, \nu^{\otimes n})$ dove ν è la Lebesgue 1-dim., e $\frac{d\mu^\theta}{d\nu}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$ per ogni $\theta \in \Theta = \begin{cases} \mathbb{R}_m \\ (0,+\infty)_{\sigma^2} \end{cases}$ e $x \in \mathbb{R}$, quindi $X_i(x_1,\dots,x_n) = x_i$ e

$$L^\theta(x_1,\dots,x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2\right) =$$

$$= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2 - \frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi\right) =$$

$$= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{m}{\sigma^2}\sum_{i=1}^{n}x_i - \left(\frac{n m^2}{2\sigma^2} + \frac{n}{2}\log\sigma^2 + \frac{n}{2}\log 2\pi\right)\right).$$

Per ciò, $\boxed{\text{caso } \theta = m \in \mathbb{R}}$: $\forall x_1,\dots,x_n \in \mathbb{R}$, l'equazione di massima verosimiglianza è $\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i - \frac{n}{\sigma^2}\hat{m}(x_1,\dots,x_n) = 0$, cioè $\hat{m}(x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{X}(x_1,\dots,x_n)$, e $\hat{m}$ è quindi anche corretto per $\theta = m$. //

Poi, $\boxed{\text{caso } \theta = \sigma^2 \in (0,+\infty)}$: $\forall x_1,\dots,x_n \in \mathbb{R}$, l'equazione di massima verosimiglianza è

$$\frac{1}{2\left(\widehat{\sigma^2}(x_1,\dots,x_n)\right)^2}\sum_{i=1}^{n}(x_i-m)^2 - \frac{n}{2\,\widehat{\sigma^2}(x_1,\dots,x_n)} = 0$$

(unitamente alla disequazione

$$-\frac{1}{(\widehat{\sigma^2})^3}\sum_{i=1}^{n}(x_i-m)^2 + \frac{n}{2(\widehat{\sigma^2})^2} \leq 0,$$

cioè $\widehat{\sigma^2}(x_1,\dots,x_n) \leq \frac{2}{n}\sum_{i=1}^{n}(x_i-m)^2$),

(MASSIMIZZANDO ANCHE m) ([NOTA] σ^2 si lib. "corretto"!)

cioè $\widehat{\sigma^2}(x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i-m)^2 \;,\; = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{X}(x_1,\dots,x_n)\right)^2 = \frac{n-1}{n}S^2$,

stimatore non corretto di $\theta = \sigma^2$, ma asintoticamente corretto. //

Passiamo finalmente a modelli stat. esponenziali „completi" che siano di campione, e concentriamoci sul problema dell'<u>asintotica correttezza</u> dello stimatore di massima verosimiglianza (per il parametro). Siano per questo $n\in\mathbb{N}_*$, ν σ-finita su $\mathcal{B}(\mathbb{R})$, $T:\mathbb{R}\to\mathbb{R}$ mis. (NON cost.) e $\Theta\subseteq\mathbb{R}$ int. aperto ($\neq\emptyset$) tali che risulti, $\forall\theta\in\Theta$, $0<\int_{\mathbb{R}} e^{\theta T(x)}\,d\nu(x)<+\infty$: definita quindi $f(\theta)\equiv\log\left(\int_{\mathbb{R}} e^{\theta T(x)}\,d\nu(x)\right)$ $\forall\theta\in\Theta$, sia μ^θ legge su $\mathcal{B}(\mathbb{R})$ equivalente a ν e con $\frac{d\mu^\theta}{d\nu}(x)=\exp(\theta T(x)-f(\theta))$ per ogni $\theta\in\Theta$ e $x\in\mathbb{R}$.

Supponiamo dunque che sia $(\Omega,\mathcal{A},\{P^\theta|\theta\in\Theta\},\mu)=(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\{\mu^{\theta\otimes n}|\theta\in\Theta\},\nu^{\otimes n})$: per ciò, se $X_i(x_1,\dots,x_n)\equiv x_i$ $\forall i=1,\dots,n$ e $\forall x_1,\dots,x_n\in\mathbb{R}$, allora si tratta del modello stat. canonico per il campione $(X_1,\dots,X_n)$ di legge μ^θ. Quindi

$$L^\theta(x_1,\dots,x_n)=\exp\left(\theta\sum_{i=1}^n T(x_i)-n f(\theta)\right)$$

, e, supponendo $T(\nu)$ σ-finita su $\mathcal{B}(\mathbb{R})$, abbiamo automaticamente f in $C^\infty(\Theta)$ (e convessa su Θ) : l'equazione di massima verosimiglianza è, $\forall x_1,\dots,x_n\in\mathbb{R}$,

$$f'(\hat\theta(x_1,\dots,x_n))=\frac{1}{n}\sum_{i=1}^n T(x_i)$$

, o scritta altrimenti

$$\forall\omega\in\Omega,\ \hat\theta_n(\omega)\ \text{tale che}\ f'(\hat\theta_n(\omega))=\frac{1}{n}\sum_{i=1}^n T(X_i(\omega)).$$

Teorema (di <u>forte consistenza</u>). All'interno del modello stat. appena formulato, e nelle notazioni usate a riguardo, supponiamo che $f':\Theta\to\mathbb{R}$ sia <u>iniettiva</u> : allora, $\forall\theta\in\Theta$, e per P^θ-quasi ogni $\omega\in\Omega$, è ben definito ed unico lo stimatore di massima verosimiglianza $\hat\theta_n(\omega)$ (per θ) per $n\geq n(\theta,\omega)$, ossia definitivamente in n, e vale che

$$\forall\theta\in\Theta,\ \hat\theta_n(\cdot)\xrightarrow[n\to+\infty]{P^\theta\text{-q.c.}}\theta$$

(per cui si ha l'asintotica correttezza).

Inoltre, $\forall \theta \in \Theta$, $\sqrt{n}\left(\hat{\theta}_n(\cdot) - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}^{P^\theta}} N\left(0, \frac{1}{\psi''(\theta)}\right)$, nel senso di convergenza in "P^θ-legge" ad una v.a. di legge $N(0, (\psi''(\theta))^{-1})$.

[Dim. La prima convergenza è sostanzialmente già dimostrata, mentre la seconda richiede qualche richiamo di "probabilità intermedia". Allora, vista la situazione e le ipotesi iniettività della ψ' (che è pure continua), abbiamo automaticamente che $\psi'(\Theta)$ resta un $\overset{(\text{int.})}{}$ aperto di $\mathbb{R}$ (teorema della mappa aperta). Inoltre, dato che $T(X_1), \dots, T(X_n)$ sono i.i.d. e P^θ-integrabili $\forall \theta \in \Theta$ (anzi, hanno ogni momento finito), per Kolmogorov si ha che, $\forall \theta \in \Theta$,

$$\frac{1}{n}\sum_{i=1}^{n} T(X_i) \xrightarrow[n\to+\infty]{P^\theta\text{-q.c.}} \mathbb{E}^{P^\theta}[T(X_1)] = \mathbb{E}^{\mu_\theta}[T] \overset{(\text{noto})}{=} \psi'(\theta) \in \psi'(\Theta) :$$

segue subito che, $\forall \theta \in \Theta$ e per P^θ-quasi ogni $\omega \in \Omega$, per ogni $n \geq n(\theta, \omega)$ si ha che $\frac{1}{n}\sum_{i=1}^{n} T(X_i(\omega)) \in \psi'(\Theta)$, ω e n per i quali dunque possiamo risolvere l'equazione di massima verosimiglianza, ottenendo $\hat{\theta}_n(\omega) = (\psi')^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i(\omega))\right)$ il quale $\hat{\theta}_n(\omega) \xrightarrow[n\to+\infty]{} (\psi')^{-1}(\psi'(\theta)) = \theta$. // Adesso,

PARENTESI SULLA CONVERGENZA IN LEGGE: su uno sp. di prob. $(\Omega, \mathcal{A}, P)$, sia Y una v.a. e sia $(Y_n)_{n\geq 1}$ una succ. di v.a.. Allora $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$ se le leggi $Y_n(P) \xrightarrow[n\to+\infty]{} Y(P)$ strettamente (in quanto misure finite su $\mathcal{B}(\mathbb{R})$): cioè se, per ogni $f: \mathbb{R} \to \mathbb{R}$ continua e limitata, si ha $\int_{\mathbb{R}} f(x)\, dY_n(P)(x) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}} f(x)\, dY(P)(x)$, ossia se

$\mathbb{E}^P[f(Y_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}^P[f(Y)]$. Fra tutti i (numerosi) risultati (più o meno di base) che riguardano la convergenza in legge, ricordiamo (solo) i seguenti:

(1) Se $Y_n \xrightarrow[n\to+\infty]{P} Y$, allora $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$; se $Y \equiv \text{cost.}$ P-q.c., allora vale il viceversa.

(2) Se $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}^P} Y$ e se $U_n \xrightarrow[n\to+\infty]{P} 0$, allora $U_n Y_n \xrightarrow[n\to+\infty]{P} 0$.

(3) Se $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}^P} Y$ e se $V_n \xrightarrow[n\to+\infty]{P} c\ (\in\mathbb{R})$, allora $Y_n + V_n \xrightarrow[n\to+\infty]{\mathcal{L}^P} Y + c$.

(4) LEMMA ("metodo delta"). (CHE USIAMO ORA) Se $Y_n \xrightarrow[n\to+\infty]{P-q.c.} a\ (\in\mathbb{R})$ e se

$\sqrt{n}\,(Y_n - a) \xrightarrow[n\to+\infty]{\mathcal{L}^P} N(0,\sigma^2)$ (per un'opportuna $\sigma^2 \geq 0$), allora, per ogni intervallino aperto $I \subseteq \mathbb{R}$ con $a \in I$ e per ogni $g : I \to \mathbb{R}$ in $\mathcal{C}^2(I)$, vale

$$\sqrt{n}\,(g(Y_n) - g(a)) \xrightarrow[n\to+\infty]{\mathcal{L}^P} N(0, \sigma^2 \cdot (g'(a))^2).$$

(5) $\boxed{\text{TEOREMA LIMITE CENTRALE (DI PAUL LÉVY)}}$ ("TLC"). Se le $(Y_n)_{n\geq 1}$ sono in $\mathcal{L}^2(P)$, non costanti, e sono i.i.d. con media $m \doteq \mathbb{E}^P[Y_1]$ e varianza $\sigma^2 \doteq \mathrm{Var}^P[Y_1]_{(>0)}$, allora vale

$$\frac{\sum_{i=1}^n Y_i - n\cdot m}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}^P} N(0,1), \text{ ossia } \frac{\sum_{i=1}^n Y_i - nm}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}^P} N(0,\sigma^2).$$

Detto altrimenti, usando $S_n \doteq \sum_{i=1}^n Y_i$ e $\overline{Y}_n \doteq \frac{1}{n} S_n$, dal momento che

$$\frac{\overbrace{\sum_{i=1}^n Y_i}^{S_n} - nm}{\sigma\sqrt{n}} = \sqrt{n}\,\frac{\overline{Y}_n - m}{\sigma},$$ vale

$$\sqrt{n}\,\frac{\overline{Y}_n - m}{\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}^P} N(0,1), \text{ ossia } \boxed{\sqrt{n}\,(\overline{Y}_n - m) \xrightarrow[n\to+\infty]{\mathcal{L}^P} N(0,\sigma^2)}.$$

→ Ultimissimo richiamo (ma super-elementare): dato $I \subseteq \mathbb{R}$ intervallo aperto $(\neq\emptyset)$ e data $f : I \to \mathbb{R}$ (INIETTIVA e) derivabile su I, allora $f^{-1} : f(I) \to I$ è derivabile su tutto $f(I)$ e vale che, $\forall x_0 \in I$,

$$\boxed{(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}}$$ (nel senso che $\left.\frac{df^{-1}(y)}{dy}\right|_{y=f(x_0)} = \frac{1}{\left.\frac{df(x)}{dx}\right|_{x=x_0}}$).

Concludiamo finalmente la dimostrazione del teorema: come dimostrato, $\forall \theta\in\Theta$ e per P^θ-q.o. $\omega\in\Omega$, per ogni $n\geq n(\theta,\omega)$ è $\hat{\vartheta}_n(\omega)=(f')^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}T(X_i(\omega))\right)$, facendo in modo che $\underbrace{\frac{1}{n}\sum_{i=1}^{n}T(X_i(\omega))}_{(Y_n)} \xrightarrow[n\to+\infty]{} \underbrace{f'(\theta)}_{(a)} \in f'(\Theta)$ (che è aperto).

Allora ora, sempre grazie al fatto che $T(X_1),\dots,T(X_n)$ sono i.i.d., e con varianza $Var^{P^\theta}[T(X_1)] \overset{(noto)}{=} \underset{(\sigma^2)}{f''(\theta)}$ $\forall\theta\in\Theta$, per il TLC abbiamo anche

$$\sqrt{n}\Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n}T(X_i)}_{(Y_n)} - \underbrace{f'(\theta)}_{(a)}\Big) \xrightarrow[n\to+\infty]{\mathcal{L}^\theta} N(0, \underbrace{f''(\theta)}_{(\sigma^2)}) \quad \forall\theta\in\Theta$$

: pertanto, considerata $\underbrace{(f')^{-1}}_{(g)} \underset{(C^1)}{:} f'(\Theta)\to\Theta$, grazie al metodo delle derivazioni risulta che

$$\sqrt{n}(\hat{\vartheta}_n-\theta) \xrightarrow[n\to+\infty]{\mathcal{L}^\theta} N\left(0, f''(\theta)\cdot\left(((f')^{-1})'(f'(\theta))\right)^2\right) \quad \forall\theta\in\Theta .$$

Per concludere, basta osservare che $((f')^{-1})'(f'(\theta)) \underset{(\text{"}f'=f\text{"})}{=} \frac{1}{f''(\theta)}$ $\forall\theta\in\Theta$. $\square$

$\boxed{\text{NOTA}}$ In riferimento al precedente teorema, capiamo che i modelli stat. di campione "tendono" a diventare non regolari, in quanto le P^θ tendono a diventare mutuamente singolari, per $n\to+\infty$. "Infatti", immaginando che $(\Omega,\mathcal{A},(P^\theta|\theta\in\Theta))$ sia per un campione infinito $(X_n)_{n\geq1}$ (una successione), per ogni $\theta_1\neq\theta_2$ in Θ (fissati) abbiamo che $A_1 \doteq \{\omega\in\Omega \mid \hat{\vartheta}_n(\omega)\xrightarrow[n\to+\infty]{}\theta_1\}$ e $A_2 \doteq \{\omega\in\Omega \mid \hat{\vartheta}_n(\omega)\xrightarrow[n\to+\infty]{}\theta_2\}$ stanno in $\mathcal{A}$ ed hanno $P^{\theta_1}[A_1] = 1 = P^{\theta_2}[A_2]$, ma $A_1\cap A_2=\emptyset$ ovviamente. ✓

Veniamo ora a concludere la nostra teoria della stima studiando una certa classe di stimatori "lineari corretti" per i parametri di una certa classe di modelli statistici (parametrici), quelli "di regressione lineare", capendo non solo (quanto) tali stimatori risultano interessanti essendo interessante già le categorie dei modelli affetto lineari ove questi sono definiti: ma capendo anche come tali stimatori siano calcolabili esplicitamente (→ "metodo dei minimi quadrati"!).

Per questo, dedichiamo di dare (calcolato) tutta una serie di preliminari sia di algebra lineare sia, ancor prima, di "probabilità intermedia", avendo in mente (in realtà) anche un loro utilizzo più futuro.

$\longmapsto$ Sia $(\Omega,\mathcal{F},P)$ uno sp. di prob. Per ogni v.a.r. X su Ω, resta definita "la funzione caratteristica di X" $\varphi_X(t): \mathbb{R}_t \to \overline{B_1(0)}^{\mathbb{C}} = \{z\in\mathbb{C} \mid |z|\le 1\}$, $\varphi_X(t) := \mathbb{E}^P[e^{itX}]$ per ogni $t\in\mathbb{R}$. Dunque, $\varphi_X(t) = \mathbb{E}^P[\cos(tX)] + i\,\mathbb{E}^P[\sin(tX)] = \int_{\mathbb{R}}\cos(tx)\,dX(P)(x) + i\int_{\mathbb{R}}\sin(tx)\,dX(P)(x)$, $t\in\mathbb{R}$, e, ad esempio, φ_X è uniformemente continua su $\mathbb{R}$ e vale che, $\forall a,b\in\mathbb{R}$, $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$, $t\in\mathbb{R}$. Più in generale, se $d\in\mathbb{N}^*$, e se $X = (X_1,\dots,X_d)^t$ è un vettore aleatorio su Ω a valori in $\mathbb{R}^d$, allora, per ogni $u = (u_1,\dots,u_d)^t\in\mathbb{R}^d$, $\varphi_X(u) := \mathbb{E}^P[e^{i\langle u;X\rangle_{\mathbb{R}^d}}] = \mathbb{E}^P[\prod_{j=1}^d e^{iu_jX_j}]$. Dunque, ad esempio, se $X_1,\dots,X_d$ sono indipendenti, allora $\varphi_X(u) = \prod_{j=1}^d\varphi_{X_j}(u_j)$.

Di per sé, la teoria delle funzioni caratteristiche non è decisamente delle più immediate, ma senza dubbio i risultati più importanti sono i due seguenti:

[1] due v.a.r. X,Y su Ω hanno la stessa legge se, e solo se, vale
$$\varphi_X(t) = \varphi_Y(t) \quad \text{per ogni } t\in\mathbb{R};$$

[2] due v.a.r. X,Y su Ω sono indipendenti se, e solo se, vale
$$\varphi_{(X,Y)}(u,v) = \varphi_X(u)\cdot\varphi_Y(v) \quad \text{per ogni } u,v\in\mathbb{R}.$$

Questi due teoremi permettono in particolare una trattazione più agile dei "vettori gaussiani" che adesso richiamiamo, ma sempre senza calcoli.

Prima, però, ricordiamo che una v.a.r. $Z \sim N(0,1)$ su Ω ha funzione caratteristica $\varphi_Z(t) = e^{-t^2/2}$, $t \in \mathbb{R}$ (la quale è reale), oppure che una v.a.r. $X \sim N(m, \sigma^2)$ $(m \in \mathbb{R})$, $m \in \mathbb{R}$ e $\sigma^2 \geq 0$, ha f.z. caratteristica $\varphi_X(t) = \exp(imt - \sigma^2 t^2/2)$, $t \in \mathbb{R}$ $(X = \sigma Z + m)$; e ricordiamo che una v.a.r. $Y \sim \Gamma(r, \lambda)$ su Ω, $r, \lambda > 0$, ha f.z. caratteristica $\varphi_Y(t) = \left(\frac{\lambda}{\lambda - it}\right)^r$, $t \in \mathbb{R}$.

Sempre su $(\Omega, \mathcal{F}, P)$ e con $d \in \mathbb{N}_*$, un vettore aleatorio $X = (X_1, \dots, X_d)^t$ in $\mathbb{R}^d$ è un "<u>vettore gaussiano</u>" se, per ogni $u = (u_1, \dots, u_d)^t \in \mathbb{R}^d$, la v.a.r. $\langle u; X \rangle_{\mathbb{R}^d} = \sum_{i=1}^d u_i X_i$ è gaussiana ($\Rightarrow$ componenti gaussiane). In particolare, un <u>campione</u> di legge gaussiana è chiaramente un vettore gaussiano. Comunque, per un tale vettore gaussiano X, abbiamo che il vettore (in $\mathbb{R}^d$) delle medie $m \doteq \mathbb{E}^P[X] = (\mathbb{E}^P[X_1], \dots, \mathbb{E}^P[X_d])^t$ e la matrice ($d \times d$) delle covarianze $Q \doteq \mathrm{Cov}^P[X] = (\mathrm{Cov}^P[X_i, X_j])_{i,j=1,\dots,d}$ <u>identificano</u> la legge di X, perché risulta, per ogni $u \in \mathbb{R}^d$,

$$\varphi_X(u) \doteq \mathbb{E}^P[e^{i\langle u; X\rangle}] = \exp\left(i\langle m; u\rangle - \tfrac{1}{2}\langle Qu; u\rangle\right)$$

: scriviamo per ciò $X \sim N_d(m, Q)$. A questo punto, è possibile dimostrare che:

$\longmapsto$ se $X \sim N_d(m, Q)$, allora le componenti $X_1, \dots, X_d$ di X sono indipendenti se, e solo se, sono scorrelate (cioè se Q è diagonale);

$\longmapsto$ se $X \sim N_d(m, Q)$, allora per ogni $k \in \mathbb{N}_*$, $A \in \mathbb{R}^{k \times d}$ e $b \in \mathbb{R}^k$ il vettore aleatorio $Y \doteq AX + b$, su Ω e valori in $\mathbb{R}^k$, resta gaussiano ma con legge $N_k(Am + b, AQA^t)$.

(<u>Ricordiamo</u> subito la formula generale (nota) $\mathrm{Cov}^P[AX+b] = A\,\mathrm{Cov}^P[X]A^t$.)

In particolare, un campione $Z \doteq (Z_1, \dots, Z_d)$ su Ω di legge $N(0,1)$ è un vettore gaussiano di legge $N_d(0, I_d)$, e per ogni vettore ($d \in \mathbb{N}_*$) $m \in \mathbb{R}^d$ e matrice $Q \in \mathbb{R}^{d \times d}$ simmetrica e semi-definita positiva il vettore aleatorio $X \doteq \sqrt{Q}Z + m$

è gaussiano e di legge $N_d(m,Q)$. ✓ Infine, si dimostra che:

▷ se $X \sim N_d(m,Q)$ e se Q è invertibile (ossia definita positiva), allora X ammette densità $f(x) : \mathbb{R}^d \to [0,+\infty)$ (rispetto alla Lebesgue d-dim.) data da $f(x) \doteq \frac{1}{(2\pi)^{d/2}\sqrt{\det Q}} \exp\left(-\frac{1}{2}\langle Q^{-1}(x-m) ; x-m\rangle_{\mathbb{R}^d}\right)$, $x\in\mathbb{R}^d$.

Ad esempio, un campione X su Ω di legge $N(m,\sigma^2)$ con $m\in\mathbb{R}$ e $\sigma^2>0$ è un vettore gaussiano di legge $N_d(\underbrace{\begin{bmatrix} m \\ m \\ \vdots \\ m \end{bmatrix}}_{(=:\mathbf{m})}, \sigma^2 I_d)$ e, quindi, di densità

$$f(x) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\Big(-\frac{1}{2\sigma^2}\underbrace{\|x-\mathbf{m}\|^2}_{(=\sum_{i=1}^{d}(x_i-m)^2)}\Big), \quad x\in\mathbb{R}^d,$$

come già sapevamo. ✓

$\boxed{\text{OSS.}}$ Risulta adesso immediato riconoscere che, se $X \doteq (X_1,\dots,X_d)^t$ è un campione su Ω di legge $N(0,\sigma^2)$, $\sigma^2>0$, e se $U\in\mathbb{R}^{d\times d}$ è ortogonale, allora $Y \doteq UX$ resta un campione su Ω di taglia d e di legge $N(0,\sigma^2)$.

[Infatti, X è un vettore gaussiano di legge $N_d(0,\sigma^2 I_d)$, e la tesi è esattamente che pure UX sia un vettore gaussiano di legge $N_d(0,\sigma^2 I_d)$. Questo è vero perché UX è un vettore gaussiano in $\mathbb{R}^d$, e ha legge $N_d(0, U(\sigma^2 I_d)U^t)$, ed ove $U\cdot U^t = U^t \cdot U = I_d$ per ortogonalità di U. ✓]

Dati di nuovo $(\Omega,\mathcal{A},P)$ sp. di prob. e X v.a.r. su Ω, e dato $n\in\mathbb{N}_*$, X ha "la legge del chi-quadro a n gradi di libertà", "$\chi^2(n)$", se X ha legge $\Gamma(\frac{n}{2},\frac{1}{2})$, per cui $\chi^2(n) = \Gamma(\frac{n}{2},\frac{1}{2})$ e ha densità $f(x) \doteq \mathbb{1}_{(0,+\infty)}(x)\frac{1}{\Gamma(\frac{n}{2})2^{n/2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$, $x\in\mathbb{R}$. Invece "$X \sim q\,\chi^2(n)$", con $q\in\mathbb{R}\setminus\{0\}$, $(\Rightarrow$ è $>0)$ se $\frac{X}{q} \sim \chi^2(n)$. Comunque, in particolare, $\chi^2(1) = \Gamma(\frac{1}{2},\frac{1}{2})$ ed è elementare trovare che una v.a.r. $Z\sim N(0,1)$ su Ω ha $Z^2 \sim \chi^2(1) = \Gamma(\frac{1}{2},\frac{1}{2})$. ✓ Di conseguenza, è evidente

che, se $Z \doteq (Z_1,\dots,Z_m)$ è un campione su Ω di legge $N(0,1)$, allora
$\|Z\|^2 = Z_1^2+\dots+Z_m^2 \sim \Gamma(\frac{m}{2},\frac{1}{2}) = \chi^2(m)$ (per indipendenza delle Z_i),
o equivalentemente che, se $X \doteq (X_1,\dots,X_m)$ è un campione su Ω di legge $N(0,\sigma^2)$, $\sigma^2>0$, allora $\|X\|^2 = X_1^2+\dots+X_m^2 \sim \sigma^2\chi^2(m)$.
($X_j=\sigma Z_j$)

Infine, una v.a.r. Y su Ω che "ha legge di Student a m gradi di libertà" "$t(m)$", se esistono due v.a.r. (indipendenti) $Z\sim N(0,1)$ e $X\sim\chi^2(m)=\Gamma(\frac{m}{2},\frac{1}{2})$ tali per cui valga $\boxed{Y=\sqrt{m}\,\frac{Z}{\sqrt{X}}}$ (P-q.c.) .

$\boxed{\text{NOTA}}$ Si può dimostrare che $t(m)$ ammette densità però rispetto alla Lebesgue 1-dim., e che $t(m) \xrightarrow[m\to+\infty]{\mathcal{L}^P} N(0,1)$.

[Verifichiamo, per esercizio, solo che consegue la legge enunciata. Abbiamo, preso un campione $(Z_1,\dots,Z_m)$ di legge $N(0,1)$ indipendente da Z, cosicché $(Z,Z_1,\dots,Z_m)$ resta un campione di legge $N(0,1)$, abbiamo che $\|Z\|^2 = Z_1^2+\dots+Z_m^2 \sim \chi^2(m) \sim$
$\sim X$, e pure che $\frac{1}{m}\sum_{j=1}^{m} Z_j^2 \xrightarrow[m\to+\infty]{P\text{-q.c.}} \mathbb{E}^P[Z_1^2] = 1$ (Var!), da cui deriviamo

$$\frac{1}{\sqrt{\frac{1}{m}\sum_{j=1}^{m} Z_j^2}} = \frac{\sqrt{m}}{\sqrt{\sum_{j=1}^{m} Z_j^2}} \xrightarrow[m\to+\infty]{P} 1 \quad , \text{ e pertanto } Y=\sqrt{m}\,\frac{Z}{\sqrt{X}} \sim$$

$$\sim \frac{\sqrt{m}}{\sqrt{\sum_{j=1}^{m} Z_j^2}}\, Z = \underbrace{\Big(\underbrace{\frac{\sqrt{m}}{\sqrt{\sum_{j=1}^{m} Z_j^2}}-1}_{\downarrow P\ 0}\Big)\underbrace{Z}_{\downarrow\mathcal{L}\ Z}}_{\downarrow P\ 0} + \underbrace{Z}_{\downarrow\mathcal{L}\ Z} \xrightarrow[m\to+\infty]{\mathcal{L}^P} Z \quad . \quad]$$

Teorema (di COCHRAN). Su uno sp. di prob. $(\Omega,\mathcal{S},P)$, sia $X\doteq(X_1,\dots,X_m)$ un campione di taglia $m\in\mathbb{N}_*$, $m\geq 2$, e di legge $N(0,\sigma^2)$, $\sigma^2>0$. Siano quindi $k\in\mathbb{N}_*$ con $2\leq k\leq m$, e siano $\{0\}\subsetneq E_1,E_2,\dots,E_k\subsetneq\mathbb{R}^m$ k $\mathbb{R}$-sottosp. propri ortogonali di $\mathbb{R}^m$ tali che $E_1\oplus E_2\oplus\dots\oplus E_k=\mathbb{R}^m$, ed aventi dim $E_j =: r_j$ $(\in \{1,\dots,n-1\})$, $j=1,\dots,k$ (per cui $r_1+r_2+\dots+r_k = n$).

Allora, se "X_{E_j}" $: \Omega \to E_j < \mathbb{R}^n$ è il vettore aleatorio in $\mathbb{R}^n$ che coincide con la proiezione ortogonale di X su E_j, $j=1,\dots,k$, allora i vettori $X_{E_1}, X_{E_2}, \dots, X_{E_k}$ sono fra loro indipendenti, ed inoltre hanno $\|X_{E_j}\|^2_{\mathbb{R}^n} \sim \sigma^2 \chi^2(r_j)$ per ogni $j=1,\dots,k$.

Dim. Fissiamo una base ortonormale di $\mathbb{R}^n$ "mettendo insieme ordinatamente" k basi ortonormali di $E_1, E_2, \dots, E_k$, una per ogni E_j; più precisamente, scegliamo: $u_1,\dots,u_{r_1}$ base ortonormale di E_1, $u_{r_1+1},\dots,u_{r_1+r_2}$ base ortonormale di E_2, $\dots$, $u_{r_1+\dots+r_{k-1}+1},\dots,u_n$ base ortonormale di E_k, per cui (per ortogonalità degli E_j) $u_1,u_2,\dots,u_n$ è una base ortonormale di $\mathbb{R}^n$. Per cui, considerata la matrice $U \in \mathbb{R}^{n\times n}$ di righe gli u_i, ossia $U := \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, U è ortogonale e ~~per~~ quindi, come già osservato, il vettore aleatorio UX in $\mathbb{R}^n$ resta un campione su Ω di legge $N(0,\sigma^2)$. Allora, se indichiamo con "$U_{(i)}$" le trasposte delle i-esime righe di U (che in questo caso è $U_{(i)} = u_i$) per ogni $i=1,\dots,n$, allora le componenti di UX sono chiaramente $\langle U_{(i)}; X\rangle = \langle u_i ; X\rangle$ per ogni $i=1,\dots,n$: riassumendo, le v.a. su Ω $\langle u_1;X\rangle, \langle u_2;X\rangle, \dots, \langle u_n;X\rangle$ sono indipendenti e di legge $N(0,\sigma^2)$. Questo sostanzialmente conclude la dimostrazione, perché

$$X_{E_1} = \sum_{i=1}^{r_1} \langle X; u_i\rangle u_i, \quad X_{E_2} = \sum_{i=r_1+1}^{r_1+r_2} \langle X; u_i\rangle u_i, \quad \dots, \quad X_{E_k} = \sum_{i=r_1+\dots+r_{k-1}+1}^{n} \langle X; u_i\rangle u_i$$

risultano di conseguenza indipendenti, e con $\|X_{E_1}\|^2 = \sum_{i=1}^{r_1} \langle X; u_i\rangle^2 \sim \sigma^2\chi^2(r_1)$, $\|X_{E_2}\|^2 = \sum_{i=r_1+1}^{r_1+r_2} \langle X; u_i\rangle^2 \sim \sigma^2\chi^2(r_2)$, $\dots$, $\|X_{E_k}\|^2 \sim \sigma^2\chi^2(r_k)$, come ben sappiamo. $\square$

COROLLARIO (di Cochran). Su uno sp. di prob. $(\Omega,\mathcal{F},P)$, sia $X \doteq (X_1,\dots,X_n)$ un campione di taglia $n\in\mathbb{N}^*$, $n\geq 2$, e di legge $N(\mu,\sigma^2)$, $\mu\in\mathbb{R}$ e $\sigma^2>0$, ed introduciamo come di consueto $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ e $S^2 \doteq \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$. Allora vale quanto segue:

(1) $\sqrt{n}\,\frac{\overline{X}-\mu}{\sigma} \sim N(0,1)$.

(2) $\frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2(n-1)$.

(3) $\overline{X}$ e S^2 sono indipendenti.

(4) $\sqrt{n}\,\frac{\overline{X}-\mu}{\sqrt{S^2}} \sim t(n-1)$.

[Dim. (1) Essendo, per indip., $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$, è evidente che $(\overline{X}-\mu)\cdot\frac{\sqrt{n}}{\sigma}$ sia $N(0,1)$. //

(4) È una immediata conseguenza di (1), di (2) e di (3), dandoli per buoni per il momento: infatti, $\frac{\overline{X}-\mu}{\sqrt{S^2}} =$

$$= \frac{\sqrt{n-1}\,\frac{\overline{X}-\mu}{\sigma}}{\sqrt{\frac{n-1}{\sigma^2}S^2}}\,, \quad \text{da cui} \quad \sqrt{n}\,\frac{\overline{X}-\mu}{\sqrt{S^2}} = \sqrt{n-1}\,\frac{\sqrt{n}\,\frac{\overline{X}-\mu}{\sigma}}{\sqrt{\frac{n-1}{\sigma^2}S^2}} \sim t(n-1)$$

appunto perché $\sqrt{n}\,\frac{\overline{X}-\mu}{\sigma}\overset{(1)}{\sim}N(0,1)$, $\frac{n-1}{\sigma^2}S^2\overset{(2)}{\sim}\chi^2(n-1)$ e tali sono indipendenti$_{(3)}$. //

(2)+(3) Sia $Z\doteq(Z_1,\dots,Z_n)$ il campione su Ω di legge $N(0,1)$ avente componenti $Z_i \doteq \frac{X_i-\mu}{\sigma}$ per ogni $i=1,\dots,n$, per cui $\overline{Z} \doteq \frac{1}{n}\sum_{i=1}^{n} Z_i \overset{(\text{lin.})}{=} \frac{\overline{X}-\mu}{\sigma}$.

Dunque, se $e\doteq[1,1,\dots,1]^t\in\mathbb{R}^n$, allora $\frac{1}{\sqrt{n}}e\in\mathbb{R}^n$ è unitario: consideriamo $E_1 \doteq \mathrm{Span}_{\mathbb{R}}\langle\frac{1}{\sqrt{n}}e\rangle$ ed $E_2\doteq E_1^{\perp}$, cosicché $\{0\}\subsetneq E_1,E_2\subsetneq\mathbb{R}^n$ sono due $\mathbb{R}$-sottosp. ortogonali propri di $\mathbb{R}^n$ tali che $E_1\oplus E_2=\mathbb{R}^n$, ed aventi $\dim E_1=1$ e $\dim E_2=n-1$. Allora, grazie al teorema di Cochran, la proiezione ortogonale Z_{E_1} di Z su E_1 e la proiezione ortogonale Z_{E_2} di Z su E_2 sono indipendenti e hanno $\|Z_{E_1}\|^2 \sim \chi^2(1)$ e $\|Z_{E_2}\|^2 \sim \chi^2(n-1)$. Osserviamo in particolare che $Z_{E_1} =$

$$= \langle Z ; \tfrac{1}{\sqrt{n}} e \rangle \tfrac{1}{\sqrt{n}} e = \tfrac{1}{n} \langle Z ; e \rangle e = \overline{Z} e = [\overline{Z}, \overline{Z}, \dots, \overline{Z}]^t$$

(da cui notiamo che, in effetti, $\|Z_{E_1}\|^2 = n \overline{Z}^2 = (\sqrt{n} \frac{\overline{X} - \mu}{\sigma})^2 \overset{(1)}{\sim} \chi^2(1)$), da cui anche d'altro $\overline{Z} \overset{(iii)}{=} \langle Z_{E_1} ; \frac{1}{\sqrt{n}} e \rangle \frac{1}{\sqrt{n}}$. Adesso, il punto chiave è che vale $\|Z\|^2 = \|Z_{E_1}\|^2 + \|Z_{E_2}\|^2$ (Pitagora), per cui:

$$\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \overset{(iii)}{=} \sum_{i=1}^{n} \Big(\underbrace{\frac{X_i - \mu}{\sigma}}_{(Z_i)} - \underbrace{\frac{\overline{X} - \mu}{\sigma}}_{(\overline{Z})}\Big)^2 = \sum_{i=1}^{n} \underbrace{(Z_i - \overline{Z})^2}_{= (Z_i^2 + \overline{Z}^2 - 2\overline{Z} Z_i)} =$$

$$= \sum_{i=1}^{n} Z_i^2 - n \overline{Z}^2 = \|Z\|^2 - \|Z_{E_1}\|^2 = \|Z_{E_2}\|^2 \sim \chi^2(n-1),$$

e ciò prova la (2). // Per la (3), infine, osservato appunto che

$$S^2 = \frac{\sigma^2}{n-1} \|Z_{E_2}\|^2 \quad \text{e che} \quad \overline{X} = \sigma \overline{Z} + \mu = \frac{\sigma}{n} \langle Z_{E_1} ; e \rangle + \mu,$$

dall'indipendenza di Z_{E_1}, Z_{E_2} segue subito quella di $\overline{X}, S^2$. $\square$

Partiamo finalmente coi modelli stat. "lineari" presentando l'IDEA base che ci sta sotto: modellizzare un "quasi-campione" $X \doteq$ $= (X_1, \dots, X_n)$ di taglia $n \in \mathbb{N}_*$, $n \geq 2$, e di legge naturalmente da indagare, sotto l'assunzione che esistano

- $\to$ $x_1, x_2, \dots, x_n \in \mathbb{R}$ quantità date,
- $\to$ $f : \mathbb{R} \to \mathbb{R}$ fz. deterministica sconosciuta,
- $\to$ $\xi_1, \xi_2, \dots, \xi_n$ v.a. sconosciute ma scorrelate, centrate ed aventi la medesima varianza $\sigma^2 > 0$

tali che $\boxed{X_i = f(x_i) + \xi_i}$ per ogni $i = 1, \dots, n$.

f facile, vorremmo modellizzare un „quasi-campione" che dipende in un modo da individuare da quantità ben note, a meno di un rumore bianco più o meno ragionevole. Dunque il fatto chiave sarebbe quello di „filtrare il rumore", nel senso di dedurre f nonostante il rumore stesso. ✓

$\boxed{\text{NOTA}}$ La (stessa) locuzione „quasi-campione" dovrebbe ora esser più chiara: ammesso anche che le $X_i(/\xi_i)$ siano fra loro indipendenti, e ammesso anche che $\xi_1,\dots,\xi_n$ abbiano la stessa legge, tuttavia la presenza di „$\dots + f(x_i)$" implica che, in generale, X_i ha una legge dipendente (proprio da $f(x_i)$! ✓)

Ebbene, un modello stat. sul quale si basi una tale situazione lo intendiamo "lineare" se, sostanzialmente, un problema di filtraggio da rumore come quello presentato si riconduce ad un „opportuno" problema lineare.

Più precisamente, se assumiamo in più che esiste pure un certo $k\in\mathbb{N}_*$ con $k<n$ tale che $f(x)=\sum_{j=1}^{k}\theta_j x^{j-1}$, $x\in\mathbb{R}$, per opportuni $\theta_1,\dots,\theta_k\in\mathbb{R}$ (sconosciuti), allora il problema diventa quello di trovare $k+1$ ($\leq n$) parametri $\theta_1,\dots,\theta_k,\sigma^2$ ($\sigma^2>0$) basandosi sulle n osservazioni $X_1,\dots,X_n$ supponendo appunto che $X_i=f(x_i)+\xi_i=\sum_{j=1}^{k}\theta_j(x_i)^{j-1}+\xi_i$, $i=1,\dots,n$: scritto altrimenti, se $a_{i,j}\doteq x_i^{j-1}$ per ogni $i=1,\dots,n$ e $j=1,\dots,k$, e se $Z_i\doteq\frac{\xi_i}{\sigma}$ per ogni $i=1,\dots,n$ ($Z_1,\dots,Z_n$ standard, centrate e di varianza 1), allora

$$\boxed{X_i=\sum_{j=1}^{k}a_{i,j}\theta_j+\sigma Z_i}\quad\text{per ogni } i=1,\dots,n.$$

Più compattamente, se $X\doteq(X_1,\dots,X_n)^t$, $Z\doteq(Z_1,\dots,Z_n)^t$, $\theta\doteq(\theta_1,\dots,\theta_k)^t$ e

$$A\doteq(a_{i,j})_{\substack{i=1,\dots,n\\ j=1,\dots,k}}=\begin{bmatrix}1 & x_1 & \dots & x_1^{k-1}\\ 1 & x_2 & \dots & x_2^{k-1}\\ \vdots & \vdots & & \vdots\\ 1 & x_n & \dots & x_n^{k-1}\end{bmatrix}\ (\in\mathbb{R}^{n\times k}),\quad\text{allora}\quad \boxed{X=A\theta+\sigma Z}.$$

$\boxed{\text{NOTA}}$ Tra l'altro, questa matrice $A\in\mathbb{R}^{n\times k}$ è una matrice di Vandermonde, per cui essere eventualmente che $x_1, x_2, \dots, x_n$ non fra loro tutti diversi equivarrebbe al essere che A sia di rango massimo (uguale a k), ossia che $A: \mathbb{R}^k \to \mathbb{R}^n$ sia iniettiva, e cioè $\dim A(\mathbb{R}^k) = k$. ✓

Definizione rigorosa: un "MODELLO STAT. DI REGRESSIONE LINEARE"

è un modello stat. parametrico $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\})$ tale per cui esistono $k, n \in \mathbb{N}_+$ con $k < n$ ed esiste una matrice $A \in \mathbb{R}^{n \times k}$ di rango massimo (uguale a k) tale che

(i) l'insieme dei parametri Θ sia l'aperto convesso di $\mathbb{R}^{k+1}$ $(\subset \mathbb{R}^n)$

$$\Theta \doteq \mathbb{R}^k_\theta \times (0,+\infty)_{\sigma^2} = \{(\theta, \sigma^2) = (\theta_1, \dots, \theta_k, \sigma^2) \in \mathbb{R}^k \times (0,+\infty)\}$$

(ii) questo modello $(\Omega, \mathcal{F}, \{P^{(\theta,\sigma^2)} \mid (\theta,\sigma^2) \in \Theta\})$ supporti l'esistenza di un vettore aleatorio $X \doteq (X_1, \dots, X_n)^t$ in $\mathbb{R}^n$ della forma (in $P^{(\theta,\sigma^2)}$-legge)

$\boxed{X = A \cdot \theta + \sigma Z}$ dove $Z \doteq (Z_1, \dots, Z_n)^t$ è un campione su Ω di taglia n, di media 0 e di varianza 1, e dove $\sigma/\sigma^2 > 0$.

In particolare, se "A_i" è la i-esima riga (trasposta) di A, $i=1,\dots,n$, allora $X_i = \langle A_i ; \theta \rangle + \sigma Z_i$ per ogni $i=1,\dots,n$ e quindi $X_1, \dots, X_n$ sono indipendenti di varianza σ^2 "ma" di medie $\langle A_i ; \theta \rangle$ (non uguale per tutte): più in breve, $E^{P^{(\theta,\sigma^2)}}[Z] = 0$ e $Cov^{P^{(\theta,\sigma^2)}}[Z] = I_n$, mentre

$E^{P^{(\theta,\sigma^2)}}[X] = A \cdot \theta$ e $Cov^{P^{(\theta,\sigma^2)}}[X] = \sigma^2 I_n$, $\forall (\theta, \sigma^2) \in \Theta$.

⟼ La costruzione esplicita è quella di un modello stat. prodotto: fissati $k, n \in \mathbb{N}_+$ con $k < n$, $A \in \mathbb{R}^{n \times k}$ di rango k, $\Theta \doteq \mathbb{R}^k_\theta \times (0,+\infty)_{\sigma^2}$, e $\sigma^{(2)} > 0$, poniamo $(\Omega, \mathcal{F}, \{P^{(\theta,\sigma^2)} \mid (\theta,\sigma^2) \in \Theta\}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\bigotimes_{i=1}^n \mu_i^{(\theta,\sigma^2)} \mid (\theta,\sigma^2) \in \Theta\})$

dove $\mu_i^{(\theta,\sigma^2)}$ è la legge su $\mathcal{B}(\mathbb{R})$ immagine mediante la v.a. $z \mapsto \langle A_{(i)};\theta\rangle + \sigma z$, $z \in \mathbb{R}$, di una fissata legge μ su $\mathcal{B}(\mathbb{R})$ avente $\int_{\mathbb{R}} x\, d\mu(x) = 0$ e $\int_{\mathbb{R}} x^2\, d\mu(x) = 1$; infine, ponevamo $X_i(x_1,\dots,x_n) \equiv x_i$ ($\forall i=1,\dots,n$, $\forall x_i \in \mathbb{R}$). ✓

$\boxed{\text{CASO SPECIALE}}$: quello gaussiano! Nel senso che, per un modello stat. come quello ora descritto, $Z = (Z_1,\dots,Z_n)$ sia un campione di legge $N(0,1)$ su Ω, ossia un vettore gaussiano su Ω di legge $N_n(0, I_n)$ ("$\mu = N(0,1)$"): allora $X = A\cdot\theta + \sigma Z$ è a sua volta un vettore gaussiano su Ω in $\mathbb{R}^n$, ma di legge $N_n(A\cdot\theta, \sigma^2 I_n)$ ($\mu_i^{(\theta,\sigma^2)} = N(\langle A_{(i)};\theta\rangle, \sigma^2)$ e $P^{(\theta,\sigma^2)} = N_n(A\cdot\theta, \sigma^2 I_n)$) e, quindi, di densità

$f^{(\theta,\sigma^2)}(x) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\|x - A\cdot\theta\|^2\right)$, $x \in \mathbb{R}^n$, che poi è la verosimiglianza del modello rispetto alla Lebesgue n-dim.

Guardiamo meglio i vettori $x - A\theta$ di $\mathbb{R}^n$, al variare di $x \in \mathbb{R}^n$: abbiamo $\|x - A\theta\|^2 = \|x\|^2 + \|A\theta\|^2 - 2\langle x; A\theta\rangle$; dunque, se consideriamo il sottospazio immagine $A(\mathbb{R}^k) \lneq \mathbb{R}^n$ (avente dim $A(\mathbb{R}^k) = k$) e quindi la "proiezione ortogonale" $\Pi: \mathbb{R}^n \to \underset{(\cong \mathbb{R}^k)}{A(\mathbb{R}^k)}$ di $\mathbb{R}^n$ su $A(\mathbb{R}^k)$, allora dalla scrittura additiva $x = \underset{A(\mathbb{R}^k)}{\underset{\cap}{\underline{\Pi x}}} + \underset{A(\mathbb{R}^k)^{\perp}}{\underset{\cap}{\underline{(x - \Pi x)}}}$, $x \in \mathbb{R}^n$, deduciamo subito che $\langle x; A\theta\rangle = \langle \Pi x; A\theta\rangle$, cosicché $\|x - A\theta\|^2 = \|x\|^2 + \|A\theta\|^2 - 2\langle \Pi x; A\theta\rangle$. Pertanto, la verosimiglianza è

$$f^{(\theta,\sigma^2)}(x) = \exp\left(-\frac{1}{2\sigma^2}\left(\|x\|^2 + \|A\theta\|^2 - 2\langle \Pi x; A\theta\rangle\right) - \frac{n}{2}\log(2\pi\sigma^2)\right) =$$

$$= \exp\Big(\underbrace{\frac{1}{\sigma^2}\langle \Pi x; A\theta\rangle - \frac{1}{2\sigma^2}\|x\|^2}_{\shortparallel ?} - \left(\frac{1}{2\sigma^2}\|A\theta\|^2 + \frac{n}{2}\log(2\pi\sigma^2)\right)\Big)$$

$$\left\langle \frac{A\cdot\theta}{\sigma^2} ; \Phi x \right\rangle - \frac{1}{2\sigma^2}\|x\|^2 =$$

$$= \Big\langle \Big(\underbrace{\frac{A\cdot\theta}{\sigma^2}}, \underbrace{\frac{1}{2\sigma^2}}_{(>0)}\Big) ; \Big(\underbrace{\Phi x ; -\|x\|^2}\Big) \Big\rangle_{\mathbb{R}^{n+1}}$$

$\left(\in A(\mathbb{R}^n) \; (\cong \mathbb{R}^n)\right)$

, e pertanto il modello è tipo esponenziale con statistica sufficiente e completa $T \doteq$ (vd. pensando ai fattori...)

$$\doteq \Big(\underbrace{\Phi X}_{\text{(stimatori di } \theta\in\mathbb{R}^n)} ; \underbrace{-\|X\|^2}_{\text{(stimatore di } -\sigma^2)}\Big) .$$

Di un modello stat. di regressione lineare, c'interessano particolarmente gli "stimatori lineari vettoriali" del parametro $\theta \in \mathbb{R}^k$: più precisamente, i vettori aleatori su Ω e valori in $\mathbb{R}^k$ del tipo $V\cdot X$, con $V \in \mathbb{R}^{k\times n}$. Cosa possiamo intendere per stimatori vettoriali: volendo stimare $g(\theta) \doteq (g_1(\theta),\dots,g_k(\theta))^t : \Theta \to \mathbb{R}^k$ mediante una statistica $U \doteq (U_1,\dots,U_k)^t : \Omega \to \mathbb{R}^k$ che stia in $\bigcap_{\theta\in\Theta} \mathcal{L}^2(P^\theta)$ (stimando $g_i(\theta)$ con U_i) $\forall i$, avremo che $\mathbb{E}^{P^\theta}[U] = (\mathbb{E}^{P^\theta}[U_1],\dots,\mathbb{E}^{P^\theta}[U_k])^t$ e che $\boxed{R_U(\theta) = \sum_{i=1}^k R_{U_i}(\theta)}$ per ogni $\theta\in\Theta$, allora intenderemo che U è corretto per $g(\theta)$ $\Leftrightarrow$ U_i è corretto per $g_i(\theta)$ ($\forall i=1,\dots,k$), ottenendo in tal caso che $R_U(\theta) = \sum_{i=1}^k \mathrm{Var}^{P^\theta}[U_i]$ per ogni $\theta\in\Theta$. [Verifichiamo che $R_U(\theta) = \sum_{i=1}^k R_{U_i}(\theta)$, $\forall\theta$: $\forall\theta$, $R_U(\theta) \doteq$

$$\doteq \mathbb{E}^{P^\theta}[\|U - g(\theta)\|^2] \overset{(\pm\mathbb{E}^{P^\theta}[U])}{=} \mathbb{E}^{P^\theta}[\underbrace{\|U - \mathbb{E}^{P^\theta}[U]\|^2}_{(=\sum_{i=1}^k (U_i - \mathbb{E}^{P^\theta}[U_i])^2)}] + \underbrace{\|\mathbb{E}^{P^\theta}[U] - g(\theta)\|^2}_{(=\sum_{i=1}^k (\mathbb{E}^{P^\theta}[U_i] - g_i(\theta))^2)} =$$

$$= \sum_{i=1}^k \left(\mathrm{Var}^{P^\theta}[U_i] + (\mathbb{E}^{P^\theta}[U_i] - g_i(\theta))^2\right) = \sum_{i=1}^k R_{U_i}(\theta) . \quad \checkmark]$$

(61)

NOTA Se un tale U è della forma $U=VZ$ con $V\in\mathbb{R}^{k\times n}$ e $Z=(Z_1,\dots,Z_n)^t$ è una stat. $\Omega\to\mathbb{R}^n$, allora $U_i=\langle V_{(i)},Z\rangle$ per ogni $i=1,\dots,k$, ossia $U_i=\sum_{j=1}^n V_{ij}Z_j$, e quindi, se U è corretto per $g(\theta)$, allora $\forall\theta\in\Theta$ $R_U(\theta)=\sum_{i=1}^k \mathrm{Var}^{P_\theta}[U_i]=\sum_{i=1}^k \mathrm{Var}^{P_\theta}\left[\sum_{j=1}^n V_{ij}Z_j\right]$: per cui, se $Z_1,\dots,Z_n$ sono scorrelate, allora $R_U(\theta)=\sum_{i=1}^k\sum_{j=1}^n V_{ij}^2\,\mathrm{Var}^{P_\theta}[Z_j]$, e se $Z_1,\dots,Z_n$ hanno tutte la stessa varianza allora $R_U(\theta)\equiv\|V\|^2\,\mathrm{Var}^{P_\theta}[Z_1]$.

Riassumendo, uno stimatore lineare vettoriale VZ, con $V\in\mathbb{R}^{k\times n}$ e $Z\in\mathbb{R}^n$ (campione), che sia corretto per una $g(\theta)$ in $\mathbb{R}^k$, e con $\sigma^2=\mathrm{Var}^{P_\theta}[Z_i]$ $\forall i=1,\dots,n$, ha rischio $\boxed{R_{VZ}(\theta)\equiv\sigma^2\|V\|^2}$ per ogni $\theta\in\Theta$ (calcolato rispetto a $g(\theta)$).

Siamo pronti per il teorema principale della sezione, e manca solo un risultato elementare di algebra lineare che è bene ricordare.

Lemma (di proiezione). Dati $k,n\in\mathbb{N}_*$ con $k<n$, sia $A\in\mathbb{R}^{n\times k}$ una matrice avente rango massimo (uguale a k), per cui $A(\mathbb{R}^k)<\mathbb{R}^n$ ha $\dim A(\mathbb{R}^k)=k$. Allora la matrice simmetrica $A^t\cdot A\in\mathbb{R}^{k^2}$ è invertibile ed è tale che, $\forall x\in\mathbb{R}^n$, risulti

$$\|A((A^t\cdot A)^{-1}A^t x)-x\|^{(2)}_{(\mathbb{R}^n)}\le\|Ay-x\|_{(\mathbb{R}^n)}\quad\text{per ogni } y\in\mathbb{R}^k,$$

ed in particolare la proiezione ortogonale $\Pi:\mathbb{R}^n\to A(\mathbb{R}^k)$ di $\mathbb{R}^n$ su $A(\mathbb{R}^k)$ "è" data da $\Pi=A\cdot(A^t\cdot A)^{-1}\cdot A^t$.

[Dim. Il problema del lemma è quello di minimizzare la distanza di x dal convesso chiuso ($\ne\emptyset$) $A(\mathbb{R}^k)$ di $\mathbb{R}^n$, per ogni $x\in\mathbb{R}^n$; sappiamo che questo si può fare in uno ed un solo modo, e che la funzione $\mathcal{U}:\mathbb{R}^n\to\mathbb{R}^k$ tale che $\|A\,\mathcal{U}x-x\|\le\|Ay-x\|$ per ogni $y\in\mathbb{R}^k$ è lineare e tale che, ovviamente, sia $A\,\mathcal{U}=\Pi$. Allora, dando per buono che

$A^t \cdot A$ sia invertibile, dall'ovvio fatto che $\Pi A = A$ deduciamo subito $A^t = A^t \Pi$ (ogni proiezione ortogonale è simmetrica) $\overset{(\Pi = A\mathcal{U})}{=} A^t \cdot A\mathcal{U}$ e cioè affatto $\mathcal{U} = (A^t \cdot A)^{-1} A^t$ e $\Pi = A\mathcal{U}$. / Concludiamo verificando quanto che, $\forall y \in \mathbb{R}^k$, se $A^t \cdot A \cdot y = 0_{\mathbb{R}^k}$ allora $y = 0_{\mathbb{R}^k}$: $0 = \langle (A^t \cdot A) y ; y \rangle_{\mathbb{R}^k} = \|Ay\|^2_{\mathbb{R}^m} \Rightarrow Ay = 0_{\mathbb{R}^m}$, cioè $y = 0_{\mathbb{R}^k}$ per iniettività di A. $\square$

Teorema (di GAUSS – MARKOV). Siano $k, m \in \mathbb{N}_*$ con $k < m$, $A \in \mathbb{R}^{m \times k}$ di rango k, $\mathcal{U} \doteq (A^t \cdot A)^{-1} A^t$ $(\in \mathbb{R}^{k \times m})$, $\Theta \doteq \mathbb{R}^k_\theta \times (0, +\infty)_{\sigma^2}$, $\sigma > 0$, e sia $(\Omega, \mathcal{A}, \{P^{\theta,\sigma^2} \mid (\theta, \sigma^2) \in \Theta\})$ un modello stat. di regressione lineare sul quale esistono un campione $Z \doteq (Z_1, \dots, Z_m)^t$ di media 0 e di varianza 1 ed un vettore aleatorio $X \doteq (X_1, \dots, X_m)^t$ tale che $X = A\theta + \sigma Z$.
Calendoci nel problema (sopraddetto) di stima per $\theta \in \mathbb{R}^k$ per mezzo di stimatori lineari della forma $V \cdot X$, $V \in \mathbb{R}^{k \times m}$, vale che:

(1) $\mathcal{U} \cdot X$ è corretto per θ, ed è ottimale fra tutti gli stimatori corretti di θ della forma VX, $V \in \mathbb{R}^{k \times m}$;

(2) $\frac{1}{m-k} \| A\mathcal{U}X - X \|^2$ è uno stimatore corretto di σ^2;

(3) nel caso gaussiano, o cioè con Z di legge $N(0,1)$, questi due stimatori $\mathcal{U}X$ e $\frac{1}{m-k} \|A\mathcal{U}X - X\|^2$ sono ottimali fra tutti gli stimatori corretti di θ e di σ^2 rispettivamente, lineari e non.

Dim. (1) Cominciamo osservando che, $\forall (\theta, \sigma^2) \in \Theta$ e $\forall V \in \mathbb{R}^{k \times m}$, è

$\mathbb{E}^{p(\theta,\sigma)}[VX] = \mathbb{E}^{p(\theta,\sigma)}[VA\theta + \sigma VZ] = VA\theta$ (perché Z è centrato),

per cui VX è corretto per θ se, e solo se, $\boxed{V\cdot A = I_k}$, il che è palese per $V = \mathcal{U} = (A^t\cdot A)^{-1}\cdot A^t$: dunque $\mathcal{U}X$ è corretto per θ. ✓

Inoltre, per ogni $V \in \mathbb{R}^{k\times n}$ tale che $V\cdot X$ sia corretto per θ, o ossia tale che $V\cdot A = I_k$, è $R_{\mathcal{U}X}(\theta) \le R_{V\cdot X}(\theta)$ $\forall\theta\in\mathbb{R}^k$: infatti, come già osservato, $R_{VX}(\theta) \equiv \sigma^2\|V\|$, ed ora $\|\mathcal{U}\| \le \|V\|$ perché $\Pi = A\mathcal{U}$ ed allora $\mathcal{U} = I_k\cdot\mathcal{U} = VA\mathcal{U} = V\Pi$, da cui $\|\mathcal{U}\| = \|V\Pi\| \overset{(sub. molt.)}{\le} \|V\|\underbrace{\|\Pi\|}_{(=1)} = \|V\|$ e ciò che fine l'obbiettivo. ✓

(2) $\sigma^2\|A\mathcal{U}X - X\| \overset{(ok)}{=} \|\Pi X - X\| = \|\Pi(A\theta+\sigma Z) - (A\theta + \sigma Z)\| =$
$= \sigma\|\Pi Z - Z\|$, per cui la tesi è $\mathbb{E}^{p(\theta,\sigma)}[\|\Pi Z - Z\|^2] = n-k$.

Ma, infatti, se $Q\in\mathbb{R}^{n\times n}$ è una matrice ortogonale tale che $Q(\Pi Z) = \Pi(QZ) = (Z_1,\dots,Z_k,\underbrace{0,\dots,0}_{(n-k\ \text{zeri})})$, allora $\|\Pi Z - Z\|^2 \overset{(ortog.)}{=}$
$= \|Q(\Pi Z - Z)\|^2 = \|\Pi(QZ) - (QZ)\|^2$, e inoltre per ortogonalità $QZ \sim N_n(0, I_n)$ $(\sim Z)$, da cui $\mathbb{E}^{p(\theta,\sigma)}[\|\Pi Z - Z\|^2] = \mathbb{E}^{p(\theta,\sigma)}[\|\Pi(QZ) - (QZ)\|^2] =$
$= \mathbb{E}^{p(\theta,\sigma)}[\|\Pi(QZ) - Z\|^2] = \sum_{i=k+1}^{n} \underbrace{\mathbb{E}^{p(\theta,\sigma)}[Z_i^2]}_{(1)} = n-k$. ✓

(3) A questo punto, basta verificare che $\mathcal{U}X$ e $\|A\mathcal{U}X - X\|^2$ siano funzioni (misurabili) di ΠX e $\|X\|^2$. Ma ciò è immediato: da una parte, infatti, $\mathcal{U}\Pi = \mathcal{U}$ (perché $\mathcal{U} = (A^t\cdot A)^{-1}A^t$ e $\Pi = A\mathcal{U}$) da cui $\mathcal{U}X = \mathcal{U}(\Pi X)$; dall'altra parte, infine, visto che $X = \Pi X + (X - \Pi X)$, $\|A\mathcal{U}X - X\|^2 = \|\Pi X - X\|^2 \overset{(Pitagora)}{=} \|X\|^2 - \|\Pi X\|^2$. □

$\Omega \to$ Volendo stimare $\vartheta \in \mathbb{R}^k$ per un modello di regressione lineare (come quello descritto nel teorema), il punto è il calcolo del vettore aleatorio (in $\mathbb{R}^k$) $\mathcal{U}X$: senza voler calcolare $A^t \cdot A$, $(A^t \cdot A)^{-1}$ e $(A^t \cdot A)^{-1} A^t = \mathcal{U}$, possiamo usare il fatto fondamentale che, $\forall \omega \in \Omega$, $\|A(\mathcal{U}X(\omega)) - X(\omega)\| \leq \|A y - X(\omega)\|$ per ogni $y \in \mathbb{R}^k$, o cioè il fatto che $\mathcal{U}X(\omega)$ minimizzi $\begin{array}{l} \mathbb{R}^k \longrightarrow [0,+\infty) \\ y \longmapsto \|Ay - X(\omega)\|^2 \end{array}$ (C^∞!) $= \sum_{i=1}^{m} \left[X_i(\omega) - \sum_{j=1}^{k} a_{i,j} y_j \right]^2$. Dunque, $\forall \omega \in \Omega$, cerchiamo il $\vartheta_0(\omega) \in \mathbb{R}^k$ che minimizza $\vartheta \mapsto \|X(\omega) - A\vartheta\|^2$, $\vartheta \in \mathbb{R}^k$, e poniamo $\mathcal{U}X(\omega) = \vartheta_0(\omega)$: è un problema ai minimi quadrati, chiaramente. ✓

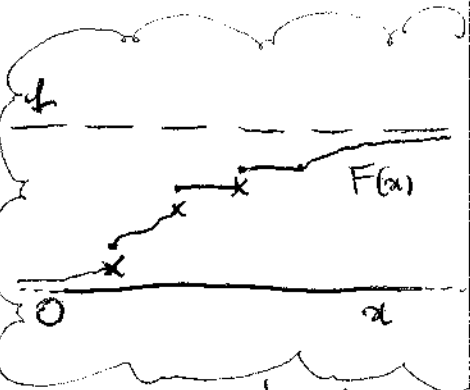
Procediamo finalmente verso la "Teoria dei test (statistici)" premettendo però una piccola digressione sulle funzioni di ripartizione 1-dim. e sui quantili (quando) di una v.a.r., ed in realtà premettendo anche il concetto di "regione di rifiuto" (di un certo livello) legato ad un modello stat., concetto che alla fine si risolve e elementari calcoli probabilistici strettamente legati al (corollario del) teorema di Cochran.

(Iniziamo!)

(Digressione sui QUANTILI) : sia $(\Omega, \mathcal{F}, P)$ uno sp. di prob., sia X una v.a.r.

su Ω, e se $F(x) \doteq F_X(x) : \mathbb{R}_x \to [0,1]_t$ la funzione di ripartizione della X, ovvero $F(x) \doteq P[X \le x] = P[X^{-1}((-\infty,x])]$, $x\in\mathbb{R}$.

Dunque F è una funzione non-decrescente, avente $\lim\limits_{x\to-\infty} F(x)=0$ e $\lim\limits_{x\to+\infty} F(x)=1$, ed identifica la legge $X(P) \doteq P_X$ di X, immagine di P mediante X. Inoltre, F è continua se, e solo se, F è suriettiva, ed in tal caso F è invertibile se, e solo se, è strettam. crescente. In più, equivalentemente, F è continua se, e solo se, $X(P)$ è diffusa (nel senso che, $\forall x\in\mathbb{R}$, $X(P)[\{x\}] = P[X=x] = 0$), ed in tal caso X è simmetrica (nel senso che $X(P)=(-X)(P)$) se, e solo se, $\forall x\in\mathbb{R}$, $\boxed{F(-x)=1-F(x)}$.



[X è simmetrica $\Leftrightarrow$ $\forall x\in\mathbb{R}$, $F(x)=P[X\le x] = P[-X \le x] \overset{(X\ge -x)}{,} = 1-P[X<-x] =$
$= 1-F(-x)$ (perché X è diffusa). $\surd$]

In ogni caso, restano definite le due pseudo-inverse $G^-(t), G^+(t) : (0,1)_t \to \mathbb{R}_x$ della F: la pseudo-inversa sinistra di F $G^-(t) \doteq$
$\doteq \inf\{x\in\mathbb{R} \mid F(x)\ge t\} = \min\{x\in\mathbb{R} \mid F(x)\ge t\}$, $t\in(0,1)$, e la pseudo-inversa destra di F $G^+(t) \doteq \inf\{x\in\mathbb{R} \mid F(x) > t\}$, $t\in(0,1)$.

La X, la F, la G^- e la G^+ sono legate da tutta una serie di proprietà classiche, tra le quali vogliamo ricordare le seguenti.

(1) $\forall t\in(0,1)$, $G^-(t)\le G^+(t)$ e vale $G^-(t)<G^+(t)$ se, e solo se, è $F\equiv t$ su tutto un intervallo (limitato); in particolare, G^- e G^+ sono diverse su un sottoinsieme al più numerabile di $(0,1)$.

(2) $\forall t\in(0,1)$, $t\le F(G^-(t))$ $(\le F(G^+(t)))$ ed inoltre, $\forall x\in\mathbb{R}$, è $G^-(t)\le x$ se, e solo se, $t\le F(x)$; inoltre, se F è continua,

allora vale che $F(G^+(t)) = F(G^-(t)) = t$ per ogni $t \in (0,1)$.

(3) Valgono altre disuguaglianze/uguaglianze come le precedenti, ed anche proprietà "analitiche" di base di G^- e G^+ $(\ldots)$, fra le quali che, se F è continua e strettamente crescente, allora $G^- \equiv G^+$ è continua e strettamente crescente e vale $G^{-/+} = F^{-1}$ (la "vera" inversa di F).

(4) Vale che se $U : \Omega \to \mathbb{R}$ v.a.r. uniforme su $[0,1]$, ossia con $P[U \leq u] = (u \vee 0) \wedge 1$ $\forall u \in \mathbb{R}$, vale che $G^-(U)$ ha f.d.r. F, o cioè che $G^-(U) \overset{\mathcal{L}}{=} X$, per cui (G1) anche $G^+(U) \overset{\mathcal{L}}{=} X$.

(5) Se F è continua, allora $F(X)$ ha legge uniforme su $[0,1]$, e quindi $P[F(X) \leq F(x)] = F(x) = P[X \leq x]$.

Notiamo, tipicamente la $G^+(t) : (0,1)_t \to \mathbb{R}_x$ viene rappresentata simbolicamente come "$q_p(X) : (0,1)_p \to \mathbb{R}_x$", cioè $q_p(X) := G^+(p) =$ $= \inf\{x \in \mathbb{R} \mid F(x) > p\}$, $p \in (0,1)$, e viene quindi chiamata come "il p-quantile di X". In particolare, se F è continua e strettam. crescente (quindi invertibile), allora semplicemente $q_p(X) = F^{-1}(p)$ per ogni $p \in (0,1)$. $\to$ $\boxed{\text{EX.}}$ Sia X diffusa e simmetrica, per cui F è ovunque continua. Se F è anche strettam. crescente, e quindi invertibile, allora

$\forall p \in (0,1)$, $\boxed{q_{1-p}(X) = -q_p(X)}$.

[Per quanto compreso, la tesi è che risulti $F^{-1}(1-p) = -F^{-1}(p)$ per ogni $p \in (0,1)$, o anche $F(-F^{-1}(p)) = 1-p$. Ma X è simmetrica, e allora (come visto) $F(-F^{-1}(p)) = 1 - \underbrace{F(F^{-1}(p))}_{(p)} = 1-p$. //]

Almeno per quanto riguarda il nostro prossimo futuro, saremo interessati a v.a.r. X su Ω che abbiano legge gaussiana, o del chi-quadro, o di Student, per cui fissiamo ora ed una volta per tutte alcune

NOTAZIONI : riserveremo il simbolo "$\Phi(x)$" $:= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$ alla f.d.r. di una v.a.r. $Z \sim N(0,1)$ su Ω, ed il simbolo "q_p" $= \Phi^{-1}(p)$ per il relativo p-quantile ($p \in (0,1)$); invece, per ogni $n \in \mathbb{N}_*$, il simbolo "$t_{p,n}$" per il p-quantile di una Student a n gradi di libertà "$t(n)$", mentre "$\chi^2_{p,n}$" per il p-quantile di una $\chi^2(n)$. (Rimarchiamo subito che stiamo parlando del valore in p della ~~...~~ inversa della f.d.r. relativa, visto che una v.a.r. che sia o gaussiana o $\chi^2(n)$ o $t(n)$ ha comunque f.d.r. continua e strettam. crescente; tuttavia, la $\chi^2(n)$ NON è simmetrica, ovviamente ("è > 0"), mentre per la $N(0,1)$ e la $t(n)$ possiamo tranquillamente scrivere le formule $q_p = -q_{1-p}$ e $t_{p,n} = -t_{1-p,n}$ per ogni $p \in (0,1)$ e $n \in \mathbb{N}_*$, come visto.)

Torniamo ad un modello stat. $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\})$. L'IDEA di "regione di fiducia", o "di confidenza", è semplicemente quella di restringere l'insieme dei parametri Θ ad un sottoinsieme proprio $S \subsetneq \Theta$ tale per cui "P^θ-molto-probabilmente" giace θ, per ogni $\theta \in \Theta$, ma con tale P^θ-probabilità non ~~...~~ dipendente in modo essenziale da Θ : più precisamente, scelto un $\alpha \in (0,1)$ con $\alpha \ll 1$ (come potrebbe essere $\alpha = 0.05$), dovremo trovare $S_\alpha \subsetneq \Theta$ (non vuoto) tale che, $\forall \theta \in \Theta$, $P^\theta[\theta \in S_\alpha] \geq 1 - \alpha$, o cioè

$P[\theta \notin S_\alpha] \le \alpha$. ✓ Formulazione ben più rigorosamente, ma poi calchiamo in concreto: scelto $\alpha \in (0,1)$ con $\alpha \ll 1$, una funzione $S(\omega) = S_\alpha(\omega) : \Omega \to \mathcal{P}(\Theta)$ tale che, $\forall \theta \in \Theta$,

"$\{\theta \in S\}$" $\doteq \{\omega \in \Omega \mid \theta \in S(\omega)\}$ $(\in \mathcal{F})$ è (DA) "UNA

REGIONE DI FIDUCIA (I) DI LIVELLO $1-\alpha$ (PER IL MODELLO)" se,

$\forall \theta \in \Theta$, $\boxed{P^\theta[\theta \in S] \ge 1-\alpha}$. ✓ (Avendo una regione di fiducia di livello $1-\alpha$ in questo senso, l'idea è che, in seguito all'esito $\omega \in \Omega$ dell'esperimento, per ogni $\theta \in \Theta$ vi è P^θ-probabilità $1-\alpha \simeq 1$ che tale ω sia uno di quelli tali che $\boxed{\theta \in S(\omega)}$ $(\subsetneq \Theta)$! ✓)

$\longmapsto$ Per $\Theta \subseteq \mathbb{R}$, le regioni di fiducia saranno tipicamente intervalli di fiducia, che forse sono quelli di cui ci ne occuperemo adesso, ed inoltre tipicamente si fa in modo che $P^\theta[\theta \in S] = 1-\alpha$, $\theta \in \Theta$. ✓

ES. Mettiamoci nel modello stat. per una v.a.r. X di legge $\mu^\theta = \mathcal{E}(\theta)$, $\theta > 0$, cioè in $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\}) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \{\mu^\theta \mid \theta > 0\})$ con $\frac{d\mu^\theta}{d\lambda}(x) = \theta e^{-\theta x}$ per ogni $\theta, x > 0$, se λ è di Lebesgue su $(0,+\infty)$, e per così $X = id_\Omega = id_{\mathbb{R}_+}$. Dunque, equivalentemente, X ha f.d.r. $P[X \le x] = [-e^{-\theta t}]_{t=0}^{t=x} = 1 - e^{-\theta x}$ per ogni $x \underset{(-)}{>} 0$, o cioè la v.a.r. $Y \doteq \theta X$ ha legge $\mathcal{E}(1)$ indipendente da Θ. Per questo, fissiamo $a, b \ge 0$ con $a < b$ e facciamo il seguente calcolo:

$\forall \theta \in \Theta$, avendo appunto che $P^\theta[a < Y < b] = (1-e^{-b}) - (1-e^{-a}) =$
$= e^{-a} - e^{-b}$, abbiamo che $e^{-a} - e^{-b} = P^\theta[a < Y < b] =$
$= P^\theta[a < \theta X < b] = P^\theta[\frac{a}{X} < \theta < \frac{b}{X}]$ per ogni $\theta \in \Theta$. ✓

Pertanto, comunque scelti $\alpha \in (0,1)$, $\alpha \ll 1$, e $a, b \geq 0$ con $a < b$ tali che $e^{-a} - e^{-b} = 1 - \alpha$, la funzione $S(\omega) : \Omega \to \mathcal{P}((0,+\infty))$ data da $S(\omega) \doteq (\frac{a}{X(\omega)}, \frac{b}{X(\omega)})$ è una regione, anzi un intervallo, di fiducia di livello $1-\alpha$ per il modello, perché per ogni $\theta \in \Theta$ è ovvio che $\{\omega \in \Omega \mid \theta \in S(\omega)\} = \{a < Y(\omega) < b\} \in \mathcal{F}$, ed inoltre come calcolato

$P^\theta[\{\omega \in \Omega \mid \theta \in S(\omega)\}] = e^{-a} - e^{-b} = 1 - \alpha$ per ogni $\theta \in \Theta$. ✓

Attenzione, $\forall \theta \in \Theta$, $P^\theta[\{\omega \in \Omega \mid \frac{a}{X(\omega)} < \theta < \frac{b}{X(\omega)}\}] = 1 - \alpha$: se ω è l'esito dell'esperimento, allora il generico parametro $\theta > 0$ "ha" la P^θ-probabilità di $1-\alpha$ di stare in $(\frac{a}{X(\omega)}, \frac{b}{X(\omega)})$, essendo appunto che ω che ha la P^θ-probabilità di $1-\alpha$ di essere tale che $\theta \in (\frac{a}{X(\omega)}, \frac{b}{X(\omega)})$!

> In generale, per gli intervalli reali di fiducia, cerchiamo di "stringere" ogni $\theta \in \Theta$ fra due "estremi" (reali) indipendenti da θ e $\mathcal{F}$-misurabili (ad esempio, uno dei due costante), in modo comunque da avere una P^θ-probabilità dell'intervallo stesso di $1-\alpha$ desiderato.
($\alpha \rightarrow$ "METODO DELLA QUANTITÀ PIVOTALE")

$\boxed{\text{NOTA}}$ Tornando all'ES. precedente (ma con $a=0$), se $\alpha \in (0,1)$ è $\alpha \ll 1$ e se $b > 0$ è tale che $1 - e^{-b} = 1 - \alpha$, e cioè $b = -\log\alpha$, allora è intervallo di fiducia di livello $1-\alpha$ l'intervallo $\left(0, \frac{-\log(\alpha)}{\overline{X}}\right)$. ✓

CASO GAUSSIANO : intervalli bilaterale, unilaterale destro ed unilaterale sinistro. (Poi test statistico!)

Mettiamoci nel modello stat. per un campione $X \doteq (X_1, \dots, X_n)$ di una scelta taglia $n \in \mathbb{N}_*$, $n \geq 2$, e di legge $\mu^{(m,\sigma^2)} = N(m, \sigma^2)$, $m \in \mathbb{R}$ e $\sigma^2 > 0$: dunque $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{(\mu^{(m,\sigma^2)})^{\otimes n} | (m,\sigma^2) \in \mathbb{R} \times (0,+\infty)\})$, con appunto $\mu^{(m,\sigma^2)} = N(m, \sigma^2)$, per cui $X_i(x_1, \dots, x_n) = x_i$. Vogliamo studiare vari, possibili intervalli di fiducia per questo modello MA in modo separato per m e per σ^2, nel senso: prima per m, che σ^2 sia nota O NO ; e poi per σ^2, che m sia nota O NO. Per il caso del "O NO" c'entrerà in aiuto Cochran! Procediamo.

INTERVALLI PER m CON σ^2 NOTA

Avendo in mente l'idea generale del metodo delle quantità pivotale, troviamo ad usare che $\sqrt{n}\frac{\overline{X}-m}{\sigma} \sim N(0,1)$ (indip. da m!) $\forall (m,\sigma^2) \in \Theta$, dove al solito $\overline{X} \doteq \frac{1}{n}\sum_{i=1}^n X_i$.

BILATERALE Per ogni $a, b \in \mathbb{R}$ con $a < b$, $\forall (m,\sigma^2) \in \Theta$, e posto $Z \doteq \sqrt{n}\frac{\overline{X}-m}{\sigma} \sim N(0,1)$, abbiamo $\Phi(b) - \Phi(a) = P^m[a < Z < b] \overset{(\text{imm.})}{=} P^m\left[\overline{X} - \frac{\sigma}{\sqrt{n}} b < m < \overline{X} + \frac{\sigma}{\sqrt{n}} a\right]$, da cui, immediatamente: se $\alpha \in (0,1)$, $\alpha \ll 1$, e se $1-\alpha = \Phi(b) - \Phi(a)$, allora $\left(\overline{X} - \frac{\sigma}{\sqrt{n}} b, \overline{X} + \frac{\sigma}{\sqrt{n}} a\right)$ è un intervallo di fiducia "bilaterale" di livello $1-\alpha$. ✓ Concentrandoci dunque sull'equazione

$\Phi(b) - \Phi(a) = 1-\alpha$, se $b = q_q$ $(= \Phi^{-1}(q))$ e $a = q_p$ $(= \Phi^{-1}(p))$ per

(66)

due (unici) $p, q \in (0,1)$ $(p<q)$, allora è equivalente ad avere $q-p = 1-\alpha$:

ad esempio, $\begin{cases} p \doteq \frac{\alpha}{2} \\ q \doteq 1-\frac{\alpha}{2} \end{cases}$ risolvono ($\alpha \ll 1$!), ed in loro corrispondenza

$$\begin{cases} a = q_p = q_{\frac{\alpha}{2}} \overset{\text{(visto)}}{=} -q_{1-\frac{\alpha}{2}} \\ b = q_q = q_{1-\frac{\alpha}{2}} \end{cases} \quad (\text{per cui, } b>0 \text{ e } a=-b), \text{ e}$$

per così $\left(\overline{X} - \frac{\sigma}{\sqrt{m}} q_{1-\frac{\alpha}{2}} , \overline{X} + \frac{\sigma}{\sqrt{m}} q_{1-\frac{\alpha}{2}}\right)$ è un bilaterale di livello $1-\alpha$. ("centrato")

$\boxed{\text{UNILATERALE DESTRO}}$ Riguardando questo fatto, è evidente che $\forall b \in \mathbb{R}$, $\forall (m,\sigma^2) \in \Theta$,

$$\Phi(b) = P^m\left[\overline{X} - \frac{\sigma}{\sqrt{m}} b < m\right]$$, e visto che, $\forall \alpha \in (0,1)$

con $\alpha \ll 1$, $\Phi(b) = 1-\alpha$ se, e solo se, $b = \Phi^{-1}(1-\alpha) = q_{1-\alpha}$,

l'intervallo unilaterale destro di livello $1-\alpha$ è $\left(\overline{X} - \frac{\sigma}{\sqrt{m}} q_{1-\alpha} , +\infty\right)$ ✓

$\boxed{\text{UNILATERALE SINISTRO}}$ Di nuovo, $\forall a \in \mathbb{R}$, $\forall (m,\sigma^2) \in \Theta$ e $\forall \alpha \in (0,1)$ con $\alpha \ll 1$, da

$$1-\alpha = 1 - \Phi(a) = P^m\left[m < \overline{X} \cancel{+} \frac{\sigma}{\sqrt{m}} a\right]$$, e da

$q_\alpha = -q_{1-\alpha}$, otteniamo il "$(1-\alpha)$-sinistro" $\left(-\infty, \overline{X} + \frac{\sigma}{\sqrt{m}} q_{1-\alpha}\right)$. ✓

($\boxed{\text{OSS.}}$ È vero che il $(1-\alpha)$-bilaterale "centrato" che abbiamo trovato fornisce, all'uscita, sia una stima dal basso che una dall'alto di $m \in \mathbb{R}$, mentre gli $(1-\alpha)$-unilaterali forniscono solo una delle due ; tuttavia, questi altri, nel fornire solo una stima la forniscono "meglio" delle due del bilaterale, perché sono più strette essendo ($\alpha \ll 1$ e, quindi,) $q_{1-\alpha} < q_{1-\frac{\alpha}{2}}$. ✓)

$\boxed{\text{INTERVALLI PER } m \text{ CON } \sigma^2 \text{ SCONOSCIUTA}}$ Non potendo, ovviamente, procedere come sopra, ricordiamo che $\sqrt{n}\,\frac{\overline{X}-\mu}{\sqrt{S^2}} \sim t(n-1)$ per Cochran, dove al solito

$S^2 \doteq \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$: indicando $S \doteq \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$,

possiamo insomma „sostituire" σ con S nei calcoli eseguiti (come ! (→ non cambia μ!)

Infatti, effettivamente, anche $\sqrt{n}\,\frac{\overline{X}-\mu}{S} \overset{(C.)}{\sim} t(n-1)$ che legge indipendente

da μ con f.d.r. invertibile e „simmetrica", per cui analogamente

otteniamo l'$(1-\alpha)$-bilaterale centrato $\left(\overline{X}-\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2},n-1},\ \overline{X}+\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2},n-1}\right)$,

l'$(1-\alpha)$-destro $\left(\overline{X}-\frac{S}{\sqrt{n}}t_{1-\alpha,n-1},\ +\infty\right)$ e l'$(1-\alpha)$-sinistro

$\left(-\infty,\ \overline{X}+\frac{S}{\sqrt{n}}t_{1-\alpha,n-1}\right)$. ✓

INTERVALLI PER σ^2 CON μ → NOTA / → NON NOTA

Se μ è nota, allora $Y \doteq \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$ (indip. σ^2)

(perché $\frac{X_i-\mu}{\sigma} \sim N(0,1)\ \forall i=1,\dots,n$), ed inoltre se

$U^2 \doteq \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ (indip. σ^2) allora $Y = \frac{n}{\sigma^2}U^2$, e $\forall a,b>0$ con $a<b$ si ha

$$P^{\sigma^2}[a<Y<b] = P^{\sigma^2}\left[\frac{nU^2}{b}<\sigma^2<\frac{nU^2}{a}\right],$$ da cui i

(= ~~...~~ q − p ...)

tre $(1-\alpha)$-intervalli $\left(\frac{nU^2}{\chi^2_{1-\frac{\alpha}{2},n}},\ \frac{nU^2}{\chi^2_{\frac{\alpha}{2},n}}\right)$, $\left(\frac{nU^2}{\chi^2_{1-\alpha,n}},\ +\infty\right)$ e

$\left(0,\ \frac{nU^2}{\chi^2_{\alpha,n}}\right)$. ✓ Se invece μ NON è nota, allora usiamo che

$\frac{n-1}{\sigma^2}S^2 \overset{(C.)}{\sim} \chi^2(n-1)$, da cui analogamente i tre $(1-\alpha)$-intervalli

$\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}},\ \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right)$, $\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}},\ +\infty\right)$ e $\left(0,\ \frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}\right)$. //

La statistica matematica NON può avere una parte teorica astratta tanto generale ed ampia da poter essere considerata come „completa", come „unificante": se questo ci è già intuitivo in principio al nostro percorso, per generalità e complessità di un modello stat. $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$, dovremo fare davvero riavolgere dopo tutti i nostri obiettivi studi più o meno in astratto. Così, anche la teoria dei test (statistici) dovrà necessariamente essere „limitata", e fermarsi quindi ad un certo punto per procedere invece con ragionamenti in contesti specifici, concreti. Detto questo, partiamo comunque in astratto per costruire (le basi di) una tale teoria, ovvero partiamo dalle idee essenziali e riguardo.

Abbiamo dunque un generico modello stat. $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\})$, possibilmente dominato o addirittura regolare, con insieme dei parametri Θ avente $\#\Theta \geq 2$ tale che, „idealmente", esista uno ed un solo parametro $\theta^* \in \Theta$ „giusto", nel senso che, per il nostro esperimento aleatorio, il „miglior" modello probabilistico tra gli $(\Omega, \mathcal{F}, P^\theta)$, al variare di $\theta \in \Theta$, sia appunto $(\Omega, \mathcal{F}, P^{\theta^*})$. Per questo,
(modello che comunque stiamo supponendo „giusto tutto", che P^{θ^*} sia noto a noi o meno!)
abbiamo costruito tutto un complesso macchinario matematico che ci potesse aiutare a stimare proprio θ^*, o piuttosto una sua funzione $g(\theta^*)$:

$\longmapsto$ mettersi in un modello stat. dominato, meglio se regolare e se per un campione di taglia „elevata";

$\longmapsto$ riassumere / semplificare tale modello mediante una sua statistica esaustiva, meglio se completa, cosicché si conosce la forma di tutta una classe di stimatori ottimali; viceversa, conosceremo anche tutte le statistiche che mai potrebbero essere stimatori: le statistiche ancillari;

2) ▶ tutto questo funziona tipicamente bene per un modello esponenziale "completo", magari per un campione ;

→ altrimenti, o in aggiunta, in situazioni di "opportuna" regolarità per la verosimiglianza, ricorriamo/semplificare il modello per mezzo della relativa matrice d'informazione secondo Fisher, V (sperando risulti invertibile) : sia, volendo, per valutare quanto una statistica sia esaustiva piuttosto che ancillare (tra le due, stanno gli stimatori), sia per discutere l'efficacia di un dato stimatore (corretto) ;

→ sempre in una tale situazione, per la quale sappiamo essere ben definita anche l'entropia del modello, discutere se uno stimatore del parametro è di massima verosimiglianza ;

b) ▶ di nuovo, per modelli esponenziali completi di campione riusciamo ad andare avanti ;

→ in ogni caso, se il modello è di regressione lineare, ad esempio gaussiano (quindi esponenziale), allora abbiamo un vero e proprio algoritmo per il calcolo di "buoni" stimatori di una certa tipologia.

Adesso, riguardando questo sommario riflesso, torniamo ad esprimere diversamente ogni risultato teorico estratto che ci indichi la forma dei "migliori" stimatori tra quelli di una classe piuttosto ampia, in quanto in generale dobbiamo essere noi a fare una scelta sulla precisa tipologia degli "buoni" stimatori per risolvere effettivamente il problema della stima, e questo è in linea di principio estremamente complicato.

Comunque, in presenza di uno stimatore sommabile e corretto UMVU di $g(\theta)$, abbiamo direttamente $\theta \in \Theta$ (e allora $\Theta \subseteq \mathbb{R}^k$), com'è che stimiamo θ^* ?. Sostituendo $U(\omega^*)$ a θ^* stesso, se ω^* è l'esito dell'esperimento !! È così, chiaramente, con rischio $R_U(\theta^*)$. Per questo, sceglieremo U "il migliore possibile" all'interno di una pre-scelta classe di stimatori di Θ, e saremmo sicuri di aver fatto più o meno il nostro meglio a fronte di una situazione statistica complicata.

Possiamo immaginare, ad esempio, che $U=\hat{\theta}$ sia di massima verosimiglianza, e che il modello sia disponibile completo per un campione di taglia "elevata" $n\in\mathbb{N}_*$, cosicché magari $U=\hat{\theta}_n \xrightarrow[n\to+\infty]{P^{\theta^*}\text{-q.c.}} \theta^*$ (e allora, con "P^{θ^*}-tutte probabilità" (e P^{θ^*} è la probabilità vera su $\mathcal{F}$), ω^* è fra quegli $\omega\in\Omega$ tali che $U(\omega)=\hat{\theta}_n(\omega) \xrightarrow[n\to+\infty]{} \theta^*$, per cui, per $n\to+\infty$, $U(\omega^*)\approx\theta^*$).

Ora, l'$\boxed{\text{IDEA}}$ è eseguire un'opportuna serie di ragionamenti e di calcolo che ci possano aiutare a "migliorare" la nostra stima

$$U(\omega^*)\approx\theta^*$$

, almeno con un certo grado di fiducia.

Per esempio, se in qualche modo riuscissimo a trovare una regione di fiducia $S_\alpha(\omega):\Omega\to\mathcal{P}(\Theta)$ di livello $1-\alpha$ per il modello, per un pre-fissato $\alpha\in(0,1)$ con $\alpha\ll 1$, di modo che risulti $\{\omega\in\Omega \mid \theta^*\in S(\omega)\}\in\mathcal{F}$ e $P^{\theta^*}[\{\omega\in\Omega\mid\theta^*\in S(\omega)\}]\gtrsim 1-\alpha$, allora potremmo affermare che, con "P^{θ^*}-quasi tutte probabilità" ($\geq 1-\alpha$), ω^* è fra quegli $\omega\in\Omega$ tali che $\theta^*\in S(\omega)$, per cui appunto

$\theta^*\in S(\omega^*)$ con "P^{θ^*}-alta probabilità", ed in definitiva avremmo migliorato la nostra stima del desiderato θ^* ottenendo

$\begin{cases} U(\omega^*) \approx \vartheta^* & \text{„PO*– da sicuro"} \\ \vartheta^* \in S(\omega^*) & \text{„PO*– molto probabilmente"} \end{cases}$. Ecco, finalmente, che la teoria dei test entra in gioco proprio per migliorare il lavoro fatto dalla teoria delle stime; e ciò accade in un modo abbastanza analogo al caso dell'osservazione data da una regione di fiducia:

un test si propone infatti di risolvere l'eventualità che ϑ^* stia in un pre-scelto sottoinsieme proprio Θ_0 di Θ anziché nel relativo complementare $\Theta_1 \doteq \Theta_0^c$ $(= \Theta \setminus \Theta_0)$, portandoci ai due possibili esiti

$$\begin{cases} U(\omega^*) \approx \vartheta^* \\ \dots \begin{cases} \to \vartheta^* \in \Theta_0 \\ \to \vartheta^* \notin \Theta_0 \quad (\text{cioè } \vartheta^* \in \Theta_1) \end{cases} \end{cases}$$

Per esempio, potrebbe essere $\#\Theta_0 = 1$, e cioè $\Theta_0 = \{\vartheta_0\}$ per un certo $\vartheta_0 \in \Theta$, caso in cui la questione è se $\vartheta^* = \vartheta_0$, oppure no. Se $\Theta \subseteq \mathbb{R}$, altrimenti, potrebbe essere $\Theta_0 = (-\infty, \vartheta_0] \cap \Theta$ per un certo $\vartheta_0 \in \Theta$, caso in cui ci stiamo chiedendo se $\vartheta^* \leq \vartheta_0$, oppure no.

Comunque possa essere fatto Θ e comunque si scelga dunque Θ_0, dobbiamo scegliere come risolvere un test in modo rigoroso, e forse ancor prima cosa intendere per "risolvere" un test, perché in ogni caso si avranno delle probabilità legate ai possibili risultati dei calcoli che si possono svolgere e ripercorrerlo. Dato, ovviamente, che noi lavoriamo con la tecnica e NON a parole, scegliere un modo matematico per risolvere un test è

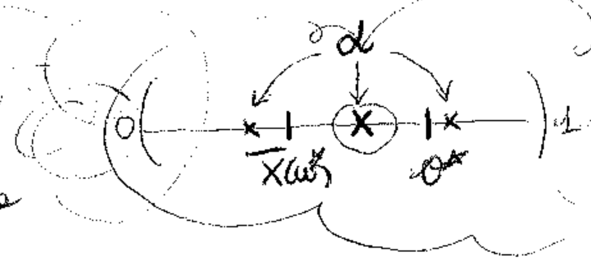
(69)

scegliere il test, nel senso che potremmo confondere il test stesso col ~~xxx~~ modo pratico di risolverlo, una volta fatte le scelte in generale, che astratte: per compiere una che ci soddisfi il più possibile, dobbiamo appunto **pensare**, partire dalle idee concrete del problema, e fatta una scelta specifica così procedendo dovremo accontentarci sempre perché all'interno della scelta possibile del modello stat. saremo sempre costretti a fare delle **scelte**, come appunto per stimare θ^* con stimatori di una precisa forma e non "quadratiche". Proponiamo tuttavia un interessante "concreto".

ES. (Controllo qualità) Mettiamoci nel modello stat. per un campione $X := (X_1, \dots, X_n)$ di una certa taglia "esatta" $n \in \mathbb{N}^*$ e di legge $\mu^\theta = B(1,\theta)$, $\theta \in (0,1)$: dunque $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\}, \mu) = (\{0,1\}^n, \mathcal{P}(\{0,1\}^n), \{(\mu^\theta)^{\otimes n} \mid \theta \in (0,1)\}, \nu^{\otimes n})$ dove ν conta i punti di $\{0,1\}$ e dove $\underset{(d\mu^\theta/d\nu)}{\mu^\theta[\{k\}]} = \theta^k (1-\theta)^{1-k}$ (per ogni $\theta \in (0,1)$ e $k \in \{0,1\}$), quindi $X_i((k_1, \dots, k_n)) = k_i$ e $\underset{(L^\theta)}{P^\theta}[\{(k_1, \dots, k_n)\}] = \theta^{\sum_{i=1}^n k_i} (1-\theta)^{1-\sum_{i=1}^n k_i}$.

Per ciò il modello è regolare, e se $\overline{X} := \frac{1}{n} \sum_{i=1}^n X_i$ come al solito, allora $\overline{X}$ è una statistica esaustiva e completa per il modello, nonché lineare, ed è uno stimatore corretto di θ, $\theta \in (0,1)$, ottimale fra gli stimatori "ammissibili" corretti di θ (lineari o no), e per di più è tale che $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \to +\infty]{P^\theta\text{-q.c.}} E^{P^\theta}[X_1] = \theta$ (K.): in questo caso, usando le notazioni della precedente introduzione euristica, possiamo scrivere che per P^{θ^*}-quasi ogni $\omega \in \Omega$, $\overline{X}(\omega) = \frac{1}{n} \sum_{i=1}^n X_i(\omega) \underset{n \to +\infty}{\simeq} \theta^*$, ed in particolare "sarà" $\overline{X}(\omega^*) \underset{n \to +\infty}{\simeq} \theta^*$, e la stima corretta nel "sostituire" le

media teorica ϑ^* con le medie campionarie o (frequenze empiriche $\overline{X}(\omega^*)$).
Ebbene, ad $\alpha \in (0,1)$ scelto, vogliamo fare un test su ϑ^* per "stabilire" se $\vartheta^* \leq \alpha$ oppure $\vartheta^* > \alpha$. Una $\boxed{\text{prima idea}}$ per risolvere tale questione potrebbe essere quella di pensare proprio per $\overline{X}$, dato che comunque vogliamo usare $\overline{X}(\omega^*) \approx \vartheta^*$, dandoci la seguente regola: posto $D := \{\omega \in \Omega \mid \overline{X}(\omega) > \alpha\} = \{\omega \in \Omega \mid \sum_{i=1}^{n} X_i(\omega) > n\alpha\}$, ("$D_n$") (ES) $\left\{ \begin{array}{l} n=100 \\ \alpha=0.1 \end{array} \right. \Rightarrow n\alpha = 10$, e $>10 \Leftrightarrow \geq 11$)
se l'esito ω^* dell'esperimento sta in D, ossia $\overline{X}(\omega^*) > \alpha$, allora stabiliremo $\vartheta^* > \alpha$; se invece $\omega^* \notin D$ ($\omega^* \in D^c$), allora stabiliremo $\vartheta^* \leq \alpha$.
(In effetti, se $\vartheta^* \not> \alpha$, allora $P_{\vartheta^*}[D] \xrightarrow[n\to+\infty]{} 0$ (Ko.), per cui (per n "elevato") se invece $\omega^* \in D$ allora "probabilmente" $P_{\vartheta^*}[D]$ non sarà così "piccola" e allora $\vartheta^* > \alpha$; è chiara l'argomentazione analoga per l'altro caso: se fosse $\vartheta^* > \alpha$, allora $P_{\vartheta^*}[D^c] \xrightarrow[n\to+\infty]{} 0$ e dunque, se invece $\omega^* \notin D$ o cioè $\omega^* \in D^c$, allora "probabilmente" $P_{\vartheta^*}[D^c]$ non sarà così piccola e quindi "probabilmente" $\vartheta^* \leq \alpha$.) Dunque si tratta sempre di fare una scelta, e per ogni possibile scelta ci saranno necessariamente vantaggi, svantaggi, caratteristiche che piacciono o caratteristiche che inducono dubbi; nel nostro esempio, volendo migliorare la stima $\overline{X}(\omega^*) \approx \vartheta^*$, possiamo immaginare che pure α sia $\alpha \approx \vartheta^*$, ovviamente, e quindi il problema principale di un tale test è: quanto essere "tranquilli" nello stabilire che $\vartheta^* \leq \alpha$ oppure $\vartheta^* > \alpha$, o viceversa, nell'eventualità che sia
$\overline{X}(\omega^*) \approx \alpha$??



Ecco che potrebbe sorgere una $\boxed{\text{seconda idea}}$

per risolvere la questione: ovvero fissare in modo così netto una regione $D \in \mathcal{B}$ che sia intesa come quella relativa al rifiuto dell'eventualità $\vartheta^* \leq \alpha$, scegliere una funzione $\Phi(\omega): \Omega \to [0,1]$ (misurabile e integrabile) che intendiamo come "risolvente" nel senso che, $\forall$ esito $\omega^* \in \Omega$, $p^* \doteq \Phi(\omega^*)$ è la probabilità che NON sia $\vartheta^* \leq \alpha$, ovvero che sia invece $\vartheta^* > \alpha$. In questo modo, infatti, avremmo una scelta finale accompagnata da un certo grado di "tranquillità" (non chiaro). Detto ciò che è certo, avere una regione di rifiuto $D \in \mathcal{B}$ oppure una "funzione di test" $\Phi(\omega)$ può essere inteso come quasi equivalente: conoscendo di risolvere il test grazie a D, otterremmo banalmente $\Phi(\omega) = \mathbb{1}_D(\omega)$; conoscendo di risolvere il test grazie invece ad una Φ, possiamo considerare $D = \Phi^{-1}(\{1\})$. L'ideale sarebbe ovviamente trovare "in modo ragionevole" una funzione di test più sofisticata di una indicatrice, e quindi assai molto più che il banale solo su una regione di rifiuto. Per esempio, tornando al controllo di qualità, volendo essere più "sicuri" accettando $\vartheta^* \leq \alpha$ se $\vartheta^* < \alpha$, potremmo basarci sulla funzione di test $\Phi(\omega) \doteq \mathbb{1}_{\{\overline{X} \geq \alpha\}}(\omega)$ (oppure con $D \subsetneq \Phi^{-1}(\{1\})$); oppure, in modo leggermente più sofisticato, potremmo "pesare" diversamente i due casi $\overline{X} > \alpha$ e $\overline{X} = \alpha$, scegliendo $\gamma \in (0,1)$ "non alto" e scegliendo

$\Phi(\omega) \doteq \Phi_\gamma(\omega) \doteq \mathbb{1}_{\{\overline{X} > \alpha\}}(\omega) + \gamma\, \mathbb{1}_{\{\overline{X} = \alpha\}}(\omega)$ (insomma, anche se $\overline{X}(\omega^*) = \alpha$ rigettiamo l'eventualità $\vartheta^* \leq \alpha$, però essendone meno sicuri). $\maltese$

Cominciamo, finalmente, la teoria: dato un modello stat. (param.) $(\Omega, \mathcal{B}, \{P^\vartheta | \vartheta \in \Theta\})$, $\#\Theta \geq 2$, e scelto un sottoinsieme proprio

Θ_0 di Θ, del quale indichiamo $\Theta_1 := \Theta \setminus \Theta_0$, UN TEST RELATIVO A TALE MODELLO E A TALE Θ_0 è un algoritmo all'=interno del modello stesso che permette sempre di dibattere e risolvere la questione se valga L'IPOTESI NULLA $\mathcal{H}_0: \vartheta \in \Theta_0$ contro L'ALTERNATIVA $\mathcal{H}_1: \vartheta \in \Theta_1$ (ossia $\vartheta \notin \Theta_0$), $\vartheta \in \Theta$. Più chiaramente, un test è un modo di risolvere il problema "$\mathcal{T}$ = «$\mathcal{H}_0: \vartheta \in \Theta_0$ contro $\mathcal{H}_1: \vartheta \in \Theta_1$»".

Se decidiamo di risolvere la cosa per mezzo di una regione di rigetto dell'ipotesi nulla, come nella prima parte dell'ES. precedente, allora il test è UNA REGIONE DI RIFIUTO, o DI RIGETTO, O CRITICA $D \in \mathcal{F}$ per la quale conveniamo che, $\forall \omega \in \Omega$, ω ci fa rifiutare l'ipotesi nulla $\mathcal{H}_0: \vartheta \in \Theta_0$ se, e solo se, $\omega \in D$, appunto. Dunque, al contrario, $D^c = \Omega \setminus D$ è la relativa REGIONE DI ACCETTAZIONE, ed il problema in generale è ovviamente quello di scegliere D in modo "soddisfacente".

Per un test come questo, chiamiamo FUNZIONE DI TEST la $\phi(\omega): \Omega \to \{0,1\}$ data da $\phi(\omega) = 1_D(\omega)$, cosicché $\phi \in \{0,1\}$ e $\phi(\omega) = 1 \Leftrightarrow \omega \in D$.

Se decidiamo invece di risolvere la cosa per mezzo proprio di funzioni di test, allora intendiamo il test come UNA FUNZIONE DI TEST $\phi(\omega): \Omega \to [0,1]$ in $\bigcap_{\vartheta \in \Theta} L^1(P^\vartheta)$, convenendo che, $\forall \omega \in \Omega$, ω ci induce a rigettare l'ipotesi nulla $\mathcal{H}_0: \vartheta \in \Theta_0$ con probabilità $\phi(\omega)$, per cui il concetto di regione di rigetto resta sfumato e spetta a noi precisarlo di volta in volta. Un test come questo è chiamato TEST ALEATORIO, ed include in modo stretto i test basati sulle regioni critiche.

Pertanto, se il problema è espresso nella sola forma possibile $\mathcal{T} =$ $=$ «$H_0 : \vartheta \in \Theta_0$ contro $H_1 : \vartheta \in \Theta_1$», ma per il modo di risolverlo va fatta una scelta, e come tale comporterà sicuramente un rischio: infatti, è tanto vero che o $\vartheta^* \in \Theta_0$ o $\vartheta^* \notin \Theta_0$, quant'è vero che il test scelto possa sbagliarsi nello stabilire la reale appartenenza, stabilendo dunque o un FALSO POSITIVO (POSITIVO A TORTO) (cioè che $\vartheta^* \in \Theta_0$ mentre invece $\vartheta^* \notin \Theta_0$) o un FALSO NEGATIVO (NEGATIVO A TORTO) (cioè che $\vartheta^* \notin \Theta_0$ mentre invece $\vartheta^* \in \Theta_0$).

Dunque l'idea è quella di minimizzare il rischio di risolvere $\mathcal{T}$ per mezzo della funzione di test Φ scelta ponendo proprio per la "probabilità" che il test Φ commette uno dei due possibili errori, per poi avere di conseguenza un modo canonico per considerare un test "migliore" di un altro test, proprio come abbiamo fatto e facciamo per gli stimatori.

Diamo un nome ai due possibili errori: lo stabilire un falso negativo, e cioè il rigettare a torto l'ipotesi nulla, è L'ERRORE DI PRIMA SPECIE; lo stabilire invece un falso positivo, e cioè l'accettare a torto l'ipotesi nulla, è L'ERRORE DI SECONDA SPECIE. Si tratta comunque di un errore, di uno sbaglio, ma è intuitivo (se non ovvio) che, a seconda della situazione statistica concreta in esame, uno dei due errori sia "oggettivamente peggiore" dell'altro: adottiamo la $\boxed{\text{CONVENZIONE}}$ di formulare $\mathcal{T}$ o come «$H_0 : \vartheta \in \Theta_0$ contro $H_1 : \vartheta \in \Theta_1$» o, al contrario, come «$H_0 : \vartheta \in \Theta_1$ contro $H_1 : \vartheta \in \Theta_0$», ossia eventualmente di "scambiare" Θ_0 con $\Theta_1 = \Theta \setminus \Theta_0$, affinché risulti oggettivamente "preferibile" un errore di seconda specie (falso positivo) ad un errore di prima specie (falso negativo). (Spieghiamo meglio con una semplice riflessione:

assoluto, è evidente che i due problemi «$H_0: O \in \Theta_0$ contro $H_1: O \in \Theta_1$» e «$H_0: O \in \Theta_1$ contro $H_1: O \in \Theta_0$» sono "completamente speculari", nel senso che sono entrambi totalmente determinati dalle scelte di Θ_0 o di Θ_1, ma l'accettare l'ipotesi nulla di uno è il rifiutare l'ipotesi nulla dell'altro, ed un errore nel risolvere uno equivale ad un errore nel risolvere l'altro, con falso positivo e falso negativo che però si scambiano. In particolare, sappiamo risolverli al tempo stesso, per cui fra i due problemi preferiamo enunciare quello che, appunto, ha come errore di falso negativo più deprecabile.

Ad esempio, in medicina: interrogandosi sulla presenza o meno di una certa malattia per un certo paziente, e volendo stabilirne appunto la presenza o l'assenza, i due possibili errori sono

→ stabilire che il paziente non è malato, mentre in realtà lo è;

→ stabilire che il paziente è sano, mentre invece è malato.

Si tratta comunque di errori gravi, ma sembra ragionevole ritenere peggiore il secondo, ossia il mandare a casa un paziente malato (che può morire, mentre un paziente sano al quale facciamo fare un percorso curativo inutile e magari pesante almeno resta in salute!). Per cui, nella nostra notazione, il problema preferiamo porlo nella forma «H_0: il paziente è malato contro H_1: il paziente è sano», anziché in quella "equivalente" speculare, affinché appunto lo stabilire un falso positivo (un falso malato) sia comunque "meglio" che lo stabilire un falso negativo (un falso sano).

Altro esempio, in legge: interrogandosi sull'innocenza o meno di un certo imputato, sarà meglio rimandare in libertà un falso innocente (un colpevole) piuttosto che ~~xxx~~ condannare un falso colpevole (un innocente), e così in questo caso preferiamo porre il problema nella forma «H_0: innocente contro H_1: colpevole», anziché in quella speculare, dal momento che, in caso di errore, sia preferibile un falso positivo (falso innocente) ad un falso negativo (falso colpevole).) Il piano dunque è il seguente: minimizzare la "potenzialità" che un test ϕ ha di commettere uno dei due errori, e restringersi ai soli test ϕ che "molto difficilmente" commetteranno un errore di prima specie, per poi sperare di trovare fra questi dei test che (in più) "difficilmente" commetteranno un errore di seconda specie. Dunque, come calcolare questa predisposizione all'errore ?? Posto il problema di test in forma $\vartheta =$ «$H_0: \vartheta \in \Theta_0$ contro $H_1: \vartheta \in \Theta_1$» in modo che sia preferibile un falso positivo ad un falso negativo, e scelta una funzione di test $\phi(\omega): \Omega \to [0,1]$, riferiamoci per un momento al "giusto" $\vartheta^* \in \Theta$: se $\omega^* \in \Omega$ è l'esito dell'esperimento e ha $\phi(\omega^*) =: p^* \overset{def}{=}$ «il grado di fiducia che conveniamo di associare alla scelta di stabilire che $\vartheta^* \notin \Theta_0$», allora è evidente che, se fosse $\vartheta^* \in \Theta_0$, $\phi(\omega^*)$ potrebbe essere considerata la percentuale di quanto stiamo sbagliando nello stabilire $\vartheta^* \notin \Theta$ (che è un falso negativo) per mezzo di ϕ e in base ω^*; viceversa, se fosse $\vartheta^* \notin \Theta_0$, allora è $1 - \phi(\omega^*)$ la percentuale di quanto stiamo sbagliando nello stabilire che $\vartheta^* \in \Theta_0$ (falso positivo) per mezzo di ϕ e in base dell'esito ω^*.

Definiamo perciò, in questo contesto, la funzione COSTO/ERRORE $C: \Theta \times [0,1] \to [0,1]$, come
$$C(\vartheta, a) := \begin{cases} a & \text{se } \vartheta \in \Theta_0 \\ 1-a & \text{se } \vartheta \notin \Theta_0 \end{cases} \quad \forall (\vartheta, a) \in \Theta \times [0,1]$$
, in modo che affinché, $\forall \omega \in \Omega$ e $\forall \vartheta \in \Theta$, il costo/l'errore di risoluzione il

problema $\mho$ per mezzo della funzione test ϕ e su base dell'esito ω sia

posto $C(\theta,\phi(\omega)) = \begin{cases} \phi(\omega) & \text{se } \theta\in\Theta_0 \quad \text{(rischio di falso negativo)} \\ 1-\phi(\omega) & \text{se } \theta\notin\Theta_0 \quad \text{(e rischio di falso positivo)} \end{cases}$,

da cui naturalmente <u>il rischio</u> di risolvere $\mho$ per mezzo del test

ϕ ottenuto mediando g/o errori su Ω, o così esplicitamente

$$R_\phi(\omega) \doteq \mathbb{E}^{P^\theta}[C(\theta,\phi(\omega))] = \begin{cases} \mathbb{E}^{P^\theta}[\phi] & \text{se } \theta\in\Theta_0 \\ 1-\mathbb{E}^{P^\theta}[\phi] & \text{se } \theta\notin\Theta_0 \end{cases}, \theta\in\Theta.$$

Per cui, la funzione $R_\phi(\cdot)|_{\Theta_0} = \mathbb{E}^{P}[\phi]$ può essere considerata come la predisposizione del test ϕ a commettere un errore di ~~seconda~~ prima specie (falso ~~positivo~~ negativo), mentre $R_\phi(\cdot)|_{\Theta_1} = 1-\mathbb{E}^{P}[\phi]$ può essere considerata come la predisposizione del test ϕ a commettere un errore di seconda specie (falso positivo). Pertanto, :

▶ scelto $\alpha\in(0,1)$ con $\alpha \ll 1$, ci restringiamo ai test ϕ <u>DI TAGLIA α</u>, nel senso che la predisposizione di ϕ a "commettere" un falso negativo sia al più α : $\boxed{\sup_{\theta\in\Theta_0}\mathbb{E}^{P^\theta}[\phi]\le\alpha}$. Per

tali test ϕ, richiediamo ~~anche~~ inoltre che pure la predisposizione di ϕ a "commettere" un falso positivo sia piuttosto bassa, quindi

$\boxed{\inf_{\theta\in\Theta_1}\mathbb{E}^{P^\theta}[\phi] \text{ "abbastanza alto"}}$. (⚠ Sembra essere più che ragionevole non creare troppo vincolo sull'errore di falso positivo, dopo quello forte sull'errore di falso negativo, anche perché $\Theta = \Theta_0 \dot\cup \Theta_1$ (unione disgiunta) e $\theta\mapsto\mathbb{E}^{P^\theta}[\phi]$, $\theta\in\Theta$, dovrebbe essere discontinua ...!) (Inoltre, insomma: non si sa farà essere un test perfetto!)

Ma in effetti, sempre per il significato che Φ ha rispetto a Θ_1, la funzione $\mathbb{E}^{P}[\Phi]\big|_{\Theta_1}$ può essere benissimo considerata la bontà del test Φ nel rivelare Θ, e per questo è LA POTENZA di Φ.

Riassumendo, ci aspettiamo idealmente a test Φ di taglia molto bassa e di potenza il più alta possibile, nel senso delle definizioni appena date che troviamo più che soddisfacenti. ✓ Ma più, inoltre, un modo canonico per "ordinare" due generici test Φ^* e Φ: Φ^* è PREFERIBILE a Φ se $R_{\Phi^*}(\theta) \le R_{\Phi}(\theta)$ per ogni $\theta \in \Theta$, o cioè se

$$\begin{cases} \mathbb{E}^{P^\theta}[\Phi^*] \le \mathbb{E}^{P^\theta}[\Phi] & \forall \theta \in \Theta_0 \\ \mathbb{E}^{P^\theta}[\Phi^*] \ge \mathbb{E}^{P^\theta}[\Phi] & \forall \theta \notin \Theta_0 \end{cases}$$

; in particolare, Φ^* è più potente di Φ. Invece, Φ^* è STRETTAMENTE PREFERIBILE a Φ se Φ^* è preferibile a Φ e se esiste un $\bar{\theta} \in \Theta$ tale che $R_{\Phi^*}(\bar{\theta}) < R_{\Phi}(\bar{\theta})$ (dunque, $\mathbb{E}^{P^{\bar{\theta}}}[\Phi^*] < \mathbb{E}^{P^{\bar{\theta}}}[\Phi]$ se $\bar{\theta} \in \Theta_0$, o il contrario esatto se $\bar{\theta} \notin \Theta_0$); Φ^* è AMMISSIBILE all'interno di una pre-scelta classe di test Φ se nessuno di tali Φ che sia preferibile a Φ^* sia anche strettamente preferibile a Φ^* (dunque, se $R_{\Phi}(\theta) \le R_{\Phi^*}(\theta)$ $\forall \theta \in \Theta$, allora $R_{\Phi} \equiv R_{\Phi^*}$); Φ^* è OTTIMALE rispetto a tali Φ se Φ^* è preferibile a tutti questi Φ. Se Φ^* è OTTIMALE rispetto ad una classe di Φ, allora è UNIFORMEMENTE PIÙ POTENTE, o "UPP", rispetto a tali Φ nel senso che $\mathbb{E}^{P^\theta}[\Phi^*] \ge \sup_{\Phi} \mathbb{E}^{P^\theta}[\Phi]$ per ogni $\theta \notin \Theta_0$, e ciò è ovvio. ✓

Dopo di che, ribadiamo che ci interessano i test Φ a taglia bassa e la potenza "ottima", o pure i TEST CORRETTI ϕ tali che $\mathbb{E}^{P^{\theta_0}}[\phi] \leq \mathbb{E}^{P^{\theta_1}}[\phi]$ $\forall \theta_0 \in \Theta_0$ e $\forall \theta_1 \in \Theta_1$, o cioè con

$$\sup_{\theta_0 \in \Theta_0} \mathbb{E}^{P^{\theta_0}}[\phi] \leq \inf_{\theta_1 \in \Theta_1} \mathbb{E}^{P^{\theta_1}}[\phi] .$$

Osserviamo che, in tutto questo, per un TEST SEMPLICE, ossia basato su un Θ_0 con $\#\Theta_0 = 1$ (un singoletto di Θ), la taglia del test ϕ è $\boxed{\mathbb{E}^{P^{\theta_0}}[\phi]}$, se $\theta_0 \in \Theta$ è quello tale che $\Theta_0 = \{\theta_0\}$; per ciò, ad esempio, all'interno di tutta una pre-scelta classe di test ϕ aventi taglia minore o uguale ad un certo $\alpha \in (0,1)$ ($\alpha \ll 1$) ma maggiore o uguale a $\mathbb{E}^{P^{\theta_0}}[\phi^*]$ per un certo test ϕ^* fra loro, è chiaro che le relazioni per ϕ^* di preferibilità, di stretta preferibilità, di ammissibilità e di ottimalità in questa classe di test ϕ diventano, rispettivamente: ϕ^* più potente di ϕ, ϕ^* strettamente più potente di ϕ, nessun ϕ è strettamente più potente di ϕ, e ϕ^* è più potente di ogni altro ϕ (cioè è le UPP).

$\boxed{\text{OSS.}}$ Se decidessimo di risolvere $\mho$ per mezzo di una check-region regione critica $D \in \mathcal{F}$, ossia se scegliessimo $\phi(\omega) \doteq \mathbb{1}_D(\omega)$, allora, giustamente, avremmo: funzione errore $C(\theta, \phi(\omega)) = \begin{cases} \mathbb{1}_D(\omega) & \text{se } \theta \in \Theta_0 \\ \mathbb{1}_{D^c}(\omega) & \text{se } \theta \notin \Theta_0 \end{cases}$, $(\theta,\omega) \in \Theta \times \Omega$, funzione rischio $R_D(\theta) \doteq R_{\mathbb{1}_D}(\theta) = \begin{cases} P^\theta[D] & \text{se } \theta \in \Theta_0 \\ 1 - P^\theta[D] & \text{se } \theta \notin \Theta_0 \end{cases}$, taglia $\sup_{\theta \in \Theta_0} P^\theta[D]$ e potenza $\inf_{\theta \in \Theta_1} P^\theta[D]$.

$\boxed{\text{NOTA}}$ Abbiamo usato le nozioni di costo/errore, e quella di rischio, per un test ϕ nel risolvere (*) come se ϕ stesse stimando una funzione $g(\theta)\in[0,1]$ di $\theta\in\Theta$ ... Ed infatti è proprio così, ovviamente, perché essere che ϕ "sta provando" a risolvere (*) è come che ϕ sta stimando la "funzione-esito-definibile" $\boxed{g(\theta)\doteq 1_{\Theta_1}(\theta)}$, ed ora basta osservare che, $\forall(\theta,a)\in\Theta\times[0,1]$, vale $C(\theta,a) = a(1-g(\theta)) + (1-a)g(\theta)$, che effettivamente dipende da θ solo attraverso la g. ✓

$\boxed{\text{OSS.}}$ Volendo risolvere (*) per mezzo di un test D, nel senso di una regione critica D, ossia una funzione di test $\phi=1_D$, in modo che tale test abbia taglia fissata, ebbene convinciamoci che questi test sono in "corrispondenza biunivoca" con le possibili regioni di fiducia (alte) per il modello. Come ottenere dai test D di taglia α una regione di fiducia $S(\omega)$ di livello $1-\alpha$ ($\alpha\in(0,1)$, $\alpha\ll 1$): per ogni $\theta_0\in\Theta$, risolviamo $T_0 = \ll H_0: \theta=\theta_0$ contro $H_1: \theta\neq\theta_0 \gg$ ($\Theta_0=\{\theta_0\}$) per mezzo di una regione critica $D(\theta_0)\in\mathcal{A}$ che abbia $P^{\theta_0}[D(\theta_0)]\le\alpha$; allora consideriamo $S(\omega):\Omega\to\mathcal{P}(\Theta)$, $S(\omega)\doteq\{\theta\in\Theta \mid \omega\notin D(\theta)\}$: dunque, $\forall\theta\in\Theta$, $\{\omega\in\Omega \mid \theta\in S(\omega)\} \overset{(\text{imm})}{=} \Omega\setminus D(\theta)$ ($\theta\in S(\omega) \Leftrightarrow \omega\in D^c(\theta)$) che ha $P^{\theta}[\{\omega\in\Omega\mid\theta\in S(\omega)\}] = 1-P^{\theta}[D(\theta)] \ge 1-\alpha$. ✓

Come ottenere dalle regioni di fiducia $S(\omega)$ di livello $1-\alpha$ un test D di taglia α: se $S(\omega): \Omega\to\mathcal{P}(\Theta)$ ha $\{\omega\in\Omega\mid\theta\in S(\omega)\}\in\mathcal{A}$ e $P^{\theta}[\{\omega\in\Omega\mid\theta\in S(\omega)\}]\ge 1-\alpha$ per ogni $\theta\in\Theta$, allora scegliamo un qualsiasi $\theta_0\in\Theta$, poniamoci il problema $T_0 = \ll H_0: \theta=\theta_0$ contro $H_1:\theta\neq\theta_0\gg$ ($\Theta_0\doteq\{\theta_0\}$) e decidiamo di risolverlo per mezzo di $D(\theta_0)\doteq$ $\doteq\{\omega\in\Omega \mid \theta_0\notin S(\omega)\}$: infatti, $D(\theta_0)\in\mathcal{A}$ e che ha $P^{\theta_0}[D(\theta_0)]\le\alpha$. ✓

(E3.) Mettiamoci nel modello stat. per un campione $X=(X_1,\dots,X_n)$ di una certa taglia $n\in\mathbb{N}^*$ (elevata) e di legge $\mu^m=N(m,\sigma^2)$, $m\in\mathbb{R}$, dove $\sigma^2>0$ è data: dunque $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\}) = (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\{\mu^{m,\otimes n}|m\in\mathbb{R}\})$, per cui $X_i(x_1,\dots,x_n)=x_i$. Fissiamo quindi $m_0\in\mathbb{R}$ e poniamoci il problema di ipotesi sulla semplice $\mathcal{T}=$ « $H_0: m=m_0$ contro $H_1: m\neq m_0$ » (cioè $\Theta_0=\{m_0\}$).

Posti $\bar X = \frac{1}{n}\sum_{i=1}^n X_i$ e $Z=\sqrt{n}\,\frac{\bar X - m_0}{\sigma}$ ($Z\overset{P^{m_0}}{\sim}N(0,1)$), decidiamo di risolvere $\mathcal{T}$ per mezzo di un test basato su un'opportuna regione critica D; guardando meglio $\mathcal{T}$ (ed ispirati dall'oss. precedente,) scegliamo di cercare D fra gli eventi della forma $D_M=\{|Z|>M\}$, al variare di $M\geq 0$: infatti, essendo $\bar X_{(n)} \xrightarrow[n\to+\infty]{P^m\text{-q.c.}} m$ (Kolmogorov) $\forall m\in\mathbb{R}$, è $Z_{(n)}\xrightarrow[n\to+\infty]{P^{m_0}\text{-q.c.}} 0$ e quindi $P^{m_0}[D_M]\xrightarrow[M\to+\infty]{} 0$, anzi $P^{m_0}[D_M]\downarrow 0$ $(M\to+\infty)$, o comunque semplicemente perché

$$P^{m_0}[D_M] = P^{m_0}[\{|Z|>M\}] = P^{m_0}[\{Z>M\}] + P^{m_0}[\{Z<-M\}] = (\Phi \text{ f.d.r. di } Z)$$
$$= 1-\Phi(M)+\underbrace{\Phi(-M)}_{(1-\Phi(M))} = 2(1-\underbrace{\Phi(M)}_{(\to 1,\ M\to+\infty)}) \xrightarrow[M\to+\infty]{} 0$$, e allora

→ D_M parrebbe avere la "forma giusta", in quanto, quale che sia l'esito $\omega^*\in\Omega$ dell'esperimento, per n elevato sarà "con tutta probabilità (P^{m^*})" $\bar X(\omega^*)\approx m^*$ ("giusto"), ossia $Z\approx 0$, e quindi il fatto che o $\omega^*\in D_M$ o $\omega^*\notin D_M$ parrebbe discriminare bene l'eventualità $m^*\neq m_0$ da $m^*=m_0$ (risp.) specialmente per M elevato a sua volta; ✓

→ al tempo stesso, possiamo richiedere M grande abbastanza affinché il test D_M abbia taglia piccola a piacere, appunto perché $P^{m_0}[D_M]\downarrow 0$ $(M\to+\infty)$! ✓

Fissato pertanto un $\alpha\in(0,1)$ con $\alpha\ll 1$, abbiamo che $P^{m_0}[D_M] =$ (visto ora) $= 2(1-\Phi(M))\leq\alpha \iff M\geq \Phi^{-1}(1-\frac{\alpha}{2}) = \varphi_{1-\frac{\alpha}{2}}$ (come nelle notazioni adottate per gli intervalli di fiducia), e perciò ristringiamoci ai soli test D_M che tagliano $M \ge q_{1-\frac{\alpha}{2}}$; ma, ovviamente, tutti questi sono chiaramente tali che $D_M \subseteq D_{q_{1-\frac{\alpha}{2}}}$, e come ovvia conseguenza c'è che $P^m[D_M] \le P^m[D_{q_{1-\frac{\alpha}{2}}}]$ per ogni $m \in \mathbb{R}$, $m = m_0$ o $m \neq m_0$ che sia : in conclusione, scegliamo il test di regione critica

$$D \doteq D_{q_{1-\frac{\alpha}{2}}} = \{|Z| > q_{1-\frac{\alpha}{2}}\}$$

(e $\alpha \in (0,1)$, $\alpha < 1$, pre-scelto), in quanto appunto che ha una "forma convincente", che taglia α ($P^{m_0}[D] = \alpha$) ed è UPP rispetto ad ogni altro test D_M che taglia α (per ogni $M \ge q_{1-\frac{\alpha}{2}}$, $P^m[D_M] \le P^m[D]$ per ogni $m \neq m_0$) . ✓ Dopo di che,

$\boxed{\text{NOTA}}$: il test risulta anche corretto, poiché

$$\inf_{\substack{m \in \mathbb{R},\\ m \neq m_0}} P^m[D_M] = \alpha \;!! \quad \text{Infatti, visto che } Z = \sqrt{n}\,\frac{\bar{X} - m_0}{\sigma} =$$

$$= \underbrace{\sqrt{n}\,\frac{\bar{X} - m}{\sigma}}_{(\sim N(0,1) \text{ sotto } P^m)} + \frac{\sqrt{n}}{\sigma}(m - m_0) \;, \overset{\forall M \ge 0 \text{ e } \forall m \in \mathbb{R}}{\text{è}} \; P^m[D_M] = P^m[\{|Z| > M\}] \overset{(\text{imm.})}{=}$$

$$= P^m\left[\left\{\left(\sqrt{n}\,\frac{\bar{X}-m}{\sigma}\right) > M - \frac{\sqrt{n}}{\sigma}(m-m_0)\right\}\right] + P^m\left[\left\{\sqrt{n}\,\frac{\bar{X}-m}{\sigma} < -M - \frac{\sqrt{n}}{\sigma}(m-m_0)\right\}\right] =$$

$$= 1 - \Phi\left(M - \frac{\sqrt{n}}{\sigma}(m-m_0)\right) + \Phi\left(-M - \frac{\sqrt{n}}{\sigma}(m-m_0)\right) \xrightarrow[\substack{m \to \pm\infty \\ (\Rightarrow m \neq m_0)}]{(\text{crescendo})} 1 \quad \text{mentre}$$

$$P^m[D_M] \xrightarrow[m \to m_0]{} 2(1 - \Phi(M)) \quad (\le \alpha \text{ per } M \ge q_{1-\frac{\alpha}{2}}) \;. \quad ✓$$

Cominciamo finalmente la nostra teoria vera e propria, partendo proprio dei test semplici, anzi dei test "più semplici" fra quelli semplici : quei test basati su un Θ_0 con $\#\Theta_0 = 1$ e pure con $\#\Theta_1 = \#(\Theta \setminus \Theta_0) = 1$. Detto altrimenti, $\#\Theta = 2$ e per ciò il nostro modello stat. $(\Omega, \mathcal{A}, \{P^\theta | \theta \in \Theta\})$ è "bivariato" nel senso che è $(\Omega, \mathcal{F}, \{P^{\theta_0}, P^{\theta_1}\})$, θ_0, θ_1 "simbolo". Per modelli stat. di questa specie, il problema è in ogni caso quello di decidersi o per P^{θ_0} o per P^{θ_1} come la probabilità $P^* \doteq P^{\theta^*}$ "giusta" per la situazione aleatoria che stiamo modellizzando. Notiamo (l'avrete) di avere e che fare con un modello stat. dominato, per esempio dalle misure finite su $\mathcal{F}$ $\mu \doteq P^{\theta_0} + P^{\theta_1}$, o meglio da

$\boxed{\mu_\lambda \doteq (1-\lambda) P^{\theta_0} + \lambda P^{\theta_1}}$ per un qualsiasi scelto $\lambda \in (0,1)$, per avere una probabilità dominante tale che, se fosse $P^{\theta_0} \sim P^{\theta_1}$, allora ovviamente sarebbe anche $\mu \sim P^{\theta_0} \sim P^{\theta_1}$ che è ~~tra~~ MA tre probabilità, per cui in tal caso CONVENIAMO di poter porre $\mu \doteq P^{\theta_0}$ oppure $\mu \doteq P^{\theta_1}$ a preferenza, ossia di poter prendere $\lambda = 0$ o $\lambda = 1$ rispettivamente.

Sia quindi $(\Omega, \mathcal{F}, \{P^0, P^1\})$ il nostro modello stat. di riferimento, per cui $\Theta \doteq \{0,1\}$, e poniamo il problema basato su $\Theta_0 \doteq \{0\}$

$\mathcal{T} \doteq$ « $H_0: \theta = 0$ contro $H_1: \theta = 1$ ». Il nostro primo, vero obiettivo a riguardo è quello di risolverlo scegliendo dei test aleatori $\phi(\omega)$ "soddisfacenti", ma cominciamo col celere provando con delle regioni critiche D "soddisfacenti". Per compiere una scelta a riguardo, scegliamo prima una qualsiasi misura dominante μ su $\mathcal{F}$, per esempio

$\mu_\lambda \doteq (1-\lambda) P^0 + \lambda P^1$ per un $\lambda \in [0,1]$, e consideriamo quindi le due densità (che esistono "dato" le assunzioni del modello)

$$\boxed{p^0(\omega) \doteq \frac{dP^0}{d\mu}(\omega)} \quad e \quad \boxed{p^1(\omega) \doteq \frac{dP^1}{d\mu}(\omega)}, \quad \omega \in \Omega.$$

Guardando $\mathcal{T}$, è evidente che un generico test aleatorio ϕ abbia taglia
$\mathbb{E}^{p^0}_{m}[\phi] = \int_{\Omega} \phi(\omega) p^0(\omega) d\mu(\omega)$ e potenza $\mathbb{E}^{p^1}_{m}[\phi] = \int_{\Omega} \phi(\omega) p^1(\omega) d\mu(\omega)$;
per ciò, sperando come sempre in $\begin{cases} \mathbb{E}^{p^0}[\phi] \text{ "molto basso"} \\ \mathbb{E}^{p^1}[\phi] \text{ "non basso"} \end{cases}$, o perlomeno in
$\mathbb{E}^{p^0}[\phi] \leq \mathbb{E}^{p^1}[\phi]$, vogliamo indagare sugli eventi della forma
"$\{\omega \in \Omega \mid p^1(\omega) \gg p^0(\omega)\}$": siano per questo $\boxed{D_M \doteq \{p^1 > M p^0\}}$ $(\downarrow \phi)$, per
ogni $M \geq 0$ (M "grande"). (NOTA Se fosse $p^0(\omega) > 0$ per ogni $\omega \in \Omega$, ossia $P^0 \sim \mu$ (e
$P^1 \ll P^0$), allora $D_M = \{\frac{p^1}{p^0} > M\}$ e $\frac{p^1(\omega)}{p^0(\omega)}$ è detto "rapporto di
verosimiglianza".) In modo equivalente, potremmo anche porre
$D_\delta \doteq D_{1/\delta} = \{p^1 > \frac{1}{\delta} p^0\}$, $\underset{(\text{"piccolo"})}{\delta > 0}$. In effetti, non solo
ci piace la "forma" dei D_M ("i quali, insomma, contengono tanto più quanto più
conta P^1 rispetto a P^0") : ma vale anche $\mathbb{P}^0[D_M] = \int_{D_M} p^0(\omega) d\mu(\omega) \underset{\substack{(\text{def. } D_M) \\ (\text{ed è} < \text{ se } \mu(D_M) > 0)}}{\leq}$
$\underset{(M>0)}{\leq} \frac{1}{M} \int_{D_M} p^1(\omega) d\mu(\omega) = \frac{1}{M} \underset{(\leq \frac{1}{M})}{\mathbb{P}^1[D_M]}$, da cui

$$\mathbb{P}^0[D_M] \leq \frac{1}{M} \xrightarrow[M \to +\infty]{} 0 \quad \text{mentre} \quad \mathbb{P}^1[D_M] \geq M\, \mathbb{P}^0[D_M] \geq \mathbb{P}^0[D_M]$$

per M grande ($M \geq 1$, in realtà).

NOTA Una volta compiuta una scelta sulla forma del test aleatorio ϕ,
o cioè sulla classe dei test ϕ che vogliamo usare come test risolventi
di $\mathcal{T}$, dobbiamo interessarci all'individuare quelli ottimali, o perlomeno
ammissibili, fra tutti. Dato un insieme di parametri così semplice,
come nel caso in esame, i veri "ordinamenti" tra due generici test
ϕ^* e ϕ della classe scelta diventano: ϕ^* è preferibile a ϕ

quando $\begin{cases} \mathbb{E}^{P^0}[\phi^*] \leq \mathbb{E}^{P^0}[\phi] \\ \mathbb{E}^{P^1}[\phi^*] \geq \mathbb{E}^{P^1}[\phi] \end{cases}$, ed in tal caso ϕ^* è strettamente preferibile a ϕ se $\mathbb{E}^{P^0}[\phi^*] < \mathbb{E}^{P^0}[\phi]$ oppure $\mathbb{E}^{P^1}[\phi^*] > \mathbb{E}^{P^1}[\phi]$; dunque, ϕ^* è ammissibile rispetto agli altri ϕ se nessuno di questi ϕ è strettamente preferibile a ϕ^*, mentre ϕ^* è ottimale rispetto ai ϕ se ϕ^* è preferibile ad ognuno dei ϕ (ed in particolare ϕ^* è UPP „di quei" ϕ). //

OSS. Quali che siano $M \geq 0$ e $\phi: \Omega \to [0,1]$ in $\mathcal{L}^1(P^0) \cap \mathcal{L}^1(P^1)$, visto che $D_M \doteq \{p^1 - Mp^0 > 0\}$, è immediato verificare che vale, $\forall \omega \in \Omega$, $(\phi(\omega) - \mathbb{1}_{D_M}(\omega))(p^1(\omega) - Mp^0(\omega)) \leq 0$, o cioè che

$$\boxed{(\phi - \mathbb{1}_{D_M})(p^1 - Mp^0) \leq 0}$$

, da cui di conseguenza

$$\boxed{\mathbb{E}^{P^1}[\phi] - P^1[D_M] \leq M\left(\mathbb{E}^{P^0}[\phi] - P^0[D_M]\right)}.$$

[Infatti, integrando rispetto a μ, $0 \geq \int_\Omega (\phi - \mathbb{1}_{D_M})(p^1 - Mp^0)\, d\mu \overset{(\text{lin.})}{=}$
$= \int_\Omega (\phi - \mathbb{1}_{D_M}) p^1 \, d\mu - M \int_\Omega (\phi - \mathbb{1}_{D_M}) p^0 \, d\mu = \mathbb{E}^{P^1}[\phi - \mathbb{1}_{D_M}] - M\,\mathbb{E}^{P^0}[\phi - \mathbb{1}_{D_M}]$,
ossia $\mathbb{E}^{P^1}[\phi - \mathbb{1}_{D_M}] \leq M\,\mathbb{E}^{P^0}[\phi - \mathbb{1}_{D_M}]$, appunto. ✓]

Lemma (di NEYMANN – PEARSON). Per il modello stat. dominato $(\Omega, \mathcal{A}, \{P^0, P^1\}, \mu)$, per il quale poniamo $p^0 \doteq \frac{dP^0}{d\mu}$ e $p^1 \doteq \frac{dP^1}{d\mu}$, e per il problema $\mathcal{T} \doteq$ « $H_0: \theta = 0$ contro $H_1: \theta = 1$ », siano scelte come regioni critiche $D_M \doteq \{p^1 > Mp^0\}$ con $M \geq 0$ grande abbastanza affinché la relativa taglia $\alpha \doteq P^0[D_M]$ sia $\ll 1$ a piacimento. Allora il test D_M è ammissibile ed UPP di ogni test $\underset{(=\mathbb{1}_D)}{\phi}$ di taglia α.

[Dim. Infatti, se $\mathbb{E}^{P^0}[\phi] \overset{(*)}{\le} \alpha = P^0[D_M]$, allora segue subito che

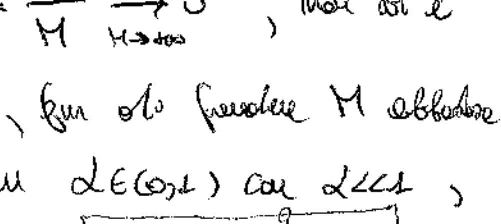
$\mathbb{E}^{P^1}[\phi] - P^1[D_M] \overset{(\text{visto})}{\le} M\left(\mathbb{E}^{P^0}[\phi] - P^0[D_M]\right) \overset{(*)}{\le} 0$, e quindi

$\mathbb{E}^{P^1}[\phi] \le P^1[D_M]$: per cui, fra tutti i test ϕ di taglia α, D_M è chiaramente UPP ed inoltre non esiste ϕ strettamente preferibile a D_M, il quale quindi è ammissibile. □] (Infatti, $\mathbb{E}^{P^1}[\phi] \le P^1[D_M]$ e, se anche fosse $\mathbb{E}^{P^0}[\phi] < \alpha$, allora $\mathbb{E}^{P^1}[\phi] < P^1[D_M]$!)

NOTA Quando $P^0[D_M] = \int_{D_M} p^0 \, d\mu \overset{(\text{def } D_M)}{\underset{(M>0)}{\le}} \frac{1}{M}\int_{D_M} p^1 \, d\mu \le \frac{1}{M} \xrightarrow[M\to+\infty]{} 0$, non vi è alcun dubbio che $P^0[D_M]$ diventi $\ll 1$ a piacere, pur di prendere M abbastanza grande. Tuttavia, scelto preliminarmente un $\alpha \in (0,1)$ con $\alpha \ll 1$, non è affatto detto che esiste $M>0$ tale per cui $\boxed{P^0[D_M] \overset{?}{=} \alpha}$! (Questo darebbe disporre (almeno) "un minimo", in quanto, se volessimo cercare $M>0$ tale che $P^0[D_M] \le \alpha$, allora mancherebbe il M "adatto" e dovremmo spingerci ad avere $P^0[D_M] < \alpha$ rendendo D_M magari non poco!)

In effetti, definita $\mathcal{A}(x) : [0,+\infty)_x \to [0,1]$ come

$$\boxed{\mathcal{A}(x) \doteq P^0[\, p^1 > x p^0 \,]} \quad , \ x \ge 0 ,$$



$\mathcal{A}$ è sì non-crescente e con $\lim_{x\to+\infty} \mathcal{A}(x) = 0$, ma, fosse anche $p^1 > 0$ P^0-q.c. o cioè $\mathcal{A}(0) = 1$, $\mathcal{A}$ è in generale NON iniettiva (anche solo su $(0,\alpha]$, $\alpha \ll 1$), o cioè NON continua, bensì soltanto continua a destra! Quello che possiamo affermare è che, per ogni $\alpha \in (0,1)$ con $\alpha \ll 1$, esiste $\bar{x} \ge 0$ tale che $\mathcal{A}(\bar{x}) \doteq P^0[\, p^1 > \bar{x} p^0 \,] \le \alpha \le P^0[\, p^1 \ge \bar{x} p^0 \,]$, per cui $\mathcal{A}(\bar{x}) = \alpha$ se $P^0[\, p^1 = \bar{x} p^0 \,] = 0$ (ad esempio quando $\mathcal{A}$ è continua), ed $\bar{x}$ è unico se $\mathcal{A}$ è strettamente decrescente.

(ES.) Dato $T>0$, nell'arco temporale $[0,T]_t$ inviamo "con continuità" un certo segnale $s(t)$ ad un certo destinatario, al quale però arriva "sporcato" e cioè da un certo rumore aleatorio "continuo" $B(t)$, per questo "piccolo" forse essere rispetto al segnale stesso; dunque, il segnale arriva a destinazione nella forma $U(t) \doteq s(t) + B(t)$. ✓ Ebbene, vogliamo ora mettere su un modello stat. affidato ad un test su di esso per "valutare" se il segnale "sia arrivato o meno", nel senso di quanto il rumore abbia sopraffatto il segnale originale. Per questo scegliamo $m \in \mathbb{N}_*$ "elevato", scegliamo m istanti $0 \le t_1 < t_2 < \dots < t_m \le T$ e CAMPIONIAMO: consideriamo cioè $s_i \doteq s(t_i)$, $B_i \doteq B(t_i)$ e $U_i \doteq U(t_i)\ (= s_i + B_i)$ per ogni $i=1,\dots,m$, e supponiamo (fare in modo) che $B_1,\dots,B_m$ siano v.a.r. indipendenti di legge $N(0,\sigma^2)$ con $\sigma^2>0$ nota, per cui anche $U_1,\dots,U_m$ sono v.a.r. indipendenti ma di legge $N(s_1,\sigma^2),\dots,N(s_m,\sigma^2)$ rispettivamente. Dunque è chiaro che il problema è il rapporto tra le due grandezze $\|s\|^2 := \sum_{i=1}^m s_i^2$ e σ^2, e che potremmo ~~impostare~~ basare il nostro test considerando il segnale come arrivato ("bene") se $(U_1,\dots,U_m)^t \sim N_m\left(\begin{bmatrix} s_1 \\ \vdots \\ s_m\end{bmatrix}, \sigma^2 I_m\right)$, mentre il segnale come NON arrivato ("bene") se $(U_1,\dots,U_m)^t \sim N_m(\underline{0}, \sigma^2 I_m)$.

Mettiamoci, pertanto, nel modello stat. prodotto $(\Omega, \mathcal{F}, \{P^0, P^1\}) \doteq \left(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \left\{ \bigotimes_{i=1}^m \mu_i^0, (\mu^1)^{\otimes m}\right\}\right)$, dove $\mu_i^0 \doteq N(s_i,\sigma^2)$ mentre $\mu^1 \doteq N(0,\sigma^2)$ per ogni $i=1,\dots,m$, e poniamo $U_i((x_1,\dots,x_m)) \doteq x_i$ e $\mathcal{T} \doteq$ « $\mathcal{H}_0 : \vartheta=0$ contro $\mathcal{H}_1 : \vartheta=1$ ». (⚠ Così facendo, abbiamo posto $\mathcal{H}_0$: "il segnale è arrivato" contro $\mathcal{H}_1$: "il segnale NON è arrivato", o così stiamo preferendo lo stabilire a torto che il segnale sia emesso quando invece non è emesso, allo stabilire a torto che il segnale non sia emesso quando invece è emesso: sembrerebbe ragionevole quindo scambiare H_0 con H_1! In effetti, pur usando il lemma di Neyman–Pearson, vedremo che i nostri test D_H non saranno affatto soddisfacenti, perché neanche corretti!!

Per ciò, se μ è la misura di Lebesgue n-dim., allora abbiamo

$$\begin{cases} p^0(u_1,\dots,u_n) \doteq \frac{dP^0}{d\mu}(u_1,\dots,u_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (u_i-s_i)^2\right) > 0 \\ p^1(u_1,\dots,u_n) \doteq \frac{dP^1}{d\mu}(u_1,\dots,u_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n u_i^2\right) > 0 \end{cases},$$

$u_1,\dots,u_n \in \mathbb{R}$, e considerate le regioni $\boxed{D_\lambda \doteq \left[\frac{p^0}{p^1} < \lambda\right]}$ per $\lambda > 0$ $(\lambda \downarrow 0)$ vogliamo imporre nel sistema

$$\begin{cases} P^0[D_\lambda] \leq \alpha & , \ \alpha \in (0,1) \text{ con } \alpha \ll 1, \text{ scelto} \\ P^1[D_\lambda] \geq \beta & , \ \beta \in (0,1) \text{ „non piccolo", scelto} \end{cases}.$$

Dunque tutto è ricondotto a calcolo probabilistico, che svolgiamo in seguito: visto che $\frac{p^0}{p^1}(u_1,\dots,u_n) \overset{(imm.)}{=} \exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n s_i u_i - \frac{1}{2\sigma^2}\sum_{i=1}^n s_i^2\right)$, $=$

$= \exp\left(-\frac{1}{2\sigma^2}\|s\|^2\right) \exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n s_i U_i(\omega)\right)$ per ogni $\omega = (u_1,\dots,u_n) \in \Omega = \mathbb{R}^n$,

abbiamo $\frac{p^0}{p^1}(\omega) < \lambda \overset{(imm.)}{\iff} \exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n s_i U_i(\omega)\right) < \lambda \exp\left(+\frac{1}{2\sigma^2}\|s\|^2\right) \iff$

$\iff \sum_{i=1}^n s_i U_i(\omega) < \sigma^2 \log(\lambda) + \frac{1}{2}\|s\|^2$, $\omega \in \Omega$, o così

$\sum_{i=1}^n s_i U_i(\omega) < \tilde{\lambda}$ se $\tilde{\lambda} \doteq \sigma^2 \log(\lambda) + \frac{1}{2}\|s\|^2$ $_{(\downarrow -\infty \text{ per } \lambda \downarrow 0)}$; ma, adesso,

$$\sum_{i=1}^n s_i U_i \begin{cases} \sim N(\|s\|^2, \sigma^2\|s\|^2) \text{ sotto } P^0 \\ \sim N(0, \sigma^2\|s\|^2) \text{ sotto } P^1 \end{cases}, \text{ ossia}$$

$$\begin{cases} \left(\sum_{i=1}^{m} s_i U_i - \|s\|^2\right) \Big/ \sigma\sqrt{\|s\|^2} \sim N(0,1) & \text{sotto } P^0 \\ \left(\sum_{i=1}^{m} s_i U_i\right) \Big/ \sigma\sqrt{\|s\|^2} \sim N(0,1) & \text{sotto } P^1 \end{cases}$$

, da cui in definitiva

$$\begin{cases} P^0[D_\lambda] \leq \alpha \\ P^1[D_\lambda] \geq \beta \end{cases} \Leftrightarrow \begin{cases} P^0[\sum_{i=1}^{m} s_i U_i < \tilde{\lambda}] \leq \alpha \\ P^1[\sum_{i=1}^{m} s_i U_i < \tilde{\lambda}] \geq \beta \end{cases} \overset{(\|s\| \doteq \sqrt{\|s\|^2})}{\Leftrightarrow} \begin{cases} \Phi\left(\frac{\tilde{\lambda} - \|s\|^2}{\sigma\|s\|}\right) \leq \alpha \\ \Phi\left(\frac{\tilde{\lambda}}{\sigma\|s\|}\right) \geq \beta \end{cases},$$

$$\Leftrightarrow \begin{cases} \tilde{\lambda} \leq \sigma\|s\| q_\alpha + \|s\|^2 \\ \tilde{\lambda} \geq \sigma\|s\| q_\beta \end{cases} \Leftrightarrow \boxed{\sigma\|s\| q_\beta \leq \tilde{\lambda} \leq \sigma\|s\| q_\alpha + \|s\|^2}:$$

occorre ovvero che sia $\sigma\|s\| q_\beta \leq \sigma\|s\| q_\alpha + \|s\|^2$. Osserviamo subito che, per $\alpha \ll 1$, $\sigma\|s\| q_\alpha$ può essere anche molto negativo, a meno che $\|s\|$ non sia talmente grande da rendere $\sigma\|s\| q_\alpha + \|s\|^2 \gtrsim 0$, cosicché si possa prendere $\beta \to 1$; in effetti, il problema è quando $\|s\|$ è piccolo rispetto a σ: avremmo $\sigma\|s\| (q_\beta - q_\alpha) \leq \|s\|^2$ e così $q_\beta - q_\alpha \leq \frac{\|s\|}{\sigma}$ ($\leftarrow$ piccolo!), da cui per perendolo $\boxed{q_\beta - q_\alpha \leq 0}$, ossia $\beta \leq \alpha \ll 1$, il che è intollerabile...! $\checkmark$

Sempre per il modello stat. dominato $(\Omega, \mathcal{A}, \{P^0, P^1\}, \mu)$, per il quale poniamo $p^0 \doteq \frac{dP^0}{d\mu}$ e $p^1 \doteq \frac{dP^1}{d\mu}$, e per il relativo problema $\mathcal{T} \doteq$ $\doteq$ « $\mathcal{H}_0: \theta = 0$ contro $\mathcal{H}_1: \theta = 1$ », pensiamo ad un procedimento risolvente più sofisticato di quello corrispondente alle „sole" regioni critiche $D_M \doteq \{p^1 > Mp^0\}$, $M \geq 0$: pensiamo cioè a decodere per dei test aleatori ϕ opportuni. / Tutto funziona piuttosto bene optando per i TEST $\phi: \Omega \to [0,1]$ DI NEYMANN – PEARSON, ovvero tali che (per definizione) esiste $M \doteq M_\phi \geq 0$ (di costante M) tale per cui valga che

$\forall \omega \in \Omega$, ($\mu$-quasi ogni) $\begin{cases} \omega \in \{p^1 > Mp^0\} \Rightarrow \phi(\omega) = 1 \\ \omega \in \{p^1 < Mp^0\} \Rightarrow \phi(\omega) = 0 \end{cases}$, mentre su $\{p^1 = Mp^0\}$

(se $\neq \emptyset$) ϕ può essere ovviamente qualsiasi (per esempio costante).

In particolare, ci stiamo mettendo in un contesto più generale del precedente, in quanto chiaramente ogni test D_M (nel senso $\mathbb{1}_{D_M}$) è di Neymann-Pearson. Un test di questo tipo è apprezzabile già dalle sue stesse definizione, avendo ben presente il Lemma di Neymann-Pearson, perché notiamo subito che: se $\phi : \Omega \to [0,1]$ è in $\mathcal{L}^1(P^0) \cap \mathcal{L}^1(P^1)$, e se ϕ_M è un test di Neymann-Pearson (dove $M \geq 0$ è tale che ϕ sia $\equiv 1$ su $\{p^1 > Mp^0\}$, e ϕ sia $\equiv 0$ su $\{p^1 < Mp^0\}$), allora vale

$$(\phi - \phi_M)(p^1 - Mp^0) \leq 0 \quad \text{, da cui subito}$$
$$\mathbb{E}^{P^1}[\phi] - \mathbb{E}^{P^1}[\phi_M] \leq M(\mathbb{E}^{P^0}[\phi] - \mathbb{E}^{P^0}[\phi_M]) .$$

Comunque, possiamo affrontare subito il risultato principale e riguardo.

Teorema (di NEYMANN-PEARSON). Per il modello stat. dominato $(\Omega, \mathcal{F}, \{P^0, P^1\}, \mu)$, per il quale poniamo $p^0 \doteq \frac{dP^0}{d\mu}$ e $p^1 \doteq \frac{dP^1}{d\mu}$, e per il relativo problema $\mathcal{O} \doteq$ « $\mathcal{H}_0 : \vartheta = 0$ contro $\mathcal{H}_1 : \vartheta = 1$ », siano scelte come funzioni di test $\boxed{\phi_{M,\gamma} \doteq \mathbb{1}_{\{p^1 > Mp^0\}} + \gamma \mathbb{1}_{\{p^1 = Mp^0\}}}$, al variare di $M \geq 0$ e di $\gamma \in [0,1]$. (Dunque, ogni $\phi_{M,\gamma}$ è un test di Neymann-Pearson "di costante M", che vale $\equiv \gamma$ su $\{p^1 = Mp^0\}$.)

Allora valgono le seguenti proposizioni.

[1] Quale che sia $\alpha \in (0,1)$ ($\alpha \ll 1$), esistono $M \geq 0$ e $\gamma \in [0,1]$ $(=M_\alpha)$ $(=\gamma_\alpha)$ tali per cui valga esattamente $\alpha = \mathbb{E}^{P^0}[\phi_{M,\gamma}]$.

$\boxed{2}$ Per ogni $M \geq 0$ e $\gamma \in [0,1]$, $\phi_{M,\gamma}$ è ammissibile ed UPP rispetto ad ogni altro test ϕ di taglia $\mathbb{E}^{p^0}[\phi_{M,\gamma}]$.

$\boxed{3}$ Quale che sia ϕ^* test (per $\mathcal{T}$), ϕ^* è di Neyman-Pearson se, e solo se, ϕ^* è UPP di ogni altro test ϕ di taglia $\mathbb{E}^{p^0}[\phi^*]$.

Detto altrimenti, le due seguenti condizioni sono equivalenti.

$\boxed{a}$ ϕ^* è test di Neyman-Pearson.

$\Updownarrow$

$\boxed{b}$ Per ogni altro test ϕ di taglia $\mathbb{E}^{p^0}[\phi^*]$, è $\mathbb{E}^{p^1}[\phi] \leq \mathbb{E}^{p^1}[\phi^*]$.

$\boxed{4}$ Se ϕ^* è un test che risulta ammissibile fra tutti i test ϕ di taglia $\mathbb{E}^{p^0}[\phi^*]$, allora ϕ^* è di Neyman-Pearson.

5) Ogni test ϕ^* di Neyman-Pearson è corretto, nel senso che vale $\mathbb{E}^{p^1}[\phi^*] \geq \mathbb{E}^{p^0}[\phi^*]$.

Dim. $\boxed{1}$ Il fatto è che, $\forall M \geq 0$ e $\forall \gamma \in [0,1]$, è $\mathbb{E}^{p^0}[\phi_{M,\gamma}] =$
$= P^0[p^1 > Mp^0] + \gamma P^0[p^1 = Mp^0]$; per ciò, ricordando anche quanto già osservato su $f(x): [0,+\infty)_x \to [0,1]$, $f(x) \doteq P^0[p^1 > xp^0]$, cominciamo osservando che, $\forall \alpha \in (0,1)$ fissato, esiste $M_\alpha \geq 0$ tale che

$$P^0[p^1 > M_\alpha p^0] \leq \alpha \leq P^0[p^1 \geq M_\alpha p^0]$$, ed allora:

$\mapsto$ se $P^0[p^1 = M_\alpha p^0] = 0$, allora $\mathbb{E}^{p^0}[\phi_{M_\alpha,\gamma}] = P^0[p^1 > M_\alpha p^0] = \alpha$ per ogni $\gamma \in [0,1]$, e vale la tesi in modo molto semplice;

$\mapsto$ se invece $P^0[p^1 = M_\alpha p^0] \neq 0$, allora poniamo

$$\gamma_\alpha \doteq \frac{\alpha - P^0[p^1 > M_\alpha p^0]}{P^0[p^1 = M_\alpha p^0]}$$, in questo $\gamma_\alpha \in [0,1]$ e chiaramente

$\mathbb{E}^{p^0}[\phi_{M_\alpha,\gamma_\alpha}] = \alpha$.

[2] Per ogni $M \geq 0$ e $\gamma \in [0,1]$, se ϕ è un test con $\mathbb{E}^{p^0}[\phi] \overset{(*)}{\leq} \mathbb{E}^{p^0}[\phi_{M,\gamma}]$, allora da $(\phi - \phi_{M,\gamma})(p^1 - Mp^0) \leq 0$ deduciamo subito che $\mathbb{E}^{p^1}[\phi] \leq$ $\leq \mathbb{E}^{p^1}[\phi_{M,\gamma}]$, da cui risulta immediata la tesi.
(l'UPP è palese, e l'ammissibilità vale perché $\mathbb{E}^{p^0}[\phi] < \mathbb{E}^{p^0}[\phi_{M,\gamma}] \Rightarrow \mathbb{E}^{p^1}[\phi] < \mathbb{E}^{p^1}[\phi_{M,\gamma}]$)

[3] $\boxed{a} \Rightarrow \boxed{b}$: è esattamente come in (2), perché, se ϕ^* è un test di Neyman-Pearson "di costante M", allora per ogni altro test ϕ con $\mathbb{E}^{p^0}[\phi] \overset{(*)}{\leq} \mathbb{E}^{p^0}[\phi^*]$ vale $\mathbb{E}^{p^1}[\phi] - \mathbb{E}^{p^1}[\phi^*] \leq M(\mathbb{E}^{p^0}[\phi] - \mathbb{E}^{p^0}[\phi^*]) \overset{(*)}{\leq} 0$, da cui appunto $\mathbb{E}^{p^1}[\phi] \leq \mathbb{E}^{p^1}[\phi^*]$. ✓

$\boxed{b} \Rightarrow \boxed{a}$: sia dunque ϕ^* un test per il quale, se ϕ è un altro test avente $\mathbb{E}^{p^0}[\phi] \overset{(*)}{\leq} \mathbb{E}^{p^0}[\phi^*]$, allora pure $\mathbb{E}^{p^1}[\phi] \leq \mathbb{E}^{p^1}[\phi^*]$. Allora, per cominciare, se poniamo $\alpha \doteq \mathbb{E}^{p^0}[\phi^*]$ e se consideriamo quegli $M \doteq M_\alpha \geq 0$ (punto (1)) e $\gamma \doteq \gamma_\alpha \in [0,1]$ tali che $\mathbb{E}^{p^0}[\phi_{M,\gamma}] = \alpha$, allora appunto vale anche $\mathbb{E}^{p^1}[\phi_{M,\gamma}] \leq \mathbb{E}^{p^1}[\phi^*]$. Ma, attenzione: $\phi_{M,\gamma}$ è di Neyman-Pearson, e quindi, come conseguenza immediata di $\mathbb{E}^{p^0}[\phi_{M,\gamma}] = \mathbb{E}^{p^0}[\phi^*]$ (usando (2), oppure il $\boxed{a} \Rightarrow \boxed{b}$ appena verificato), è pure $\mathbb{E}^{p^1}[\phi^*] \leq \mathbb{E}^{p^1}[\phi_{M,\gamma}]$!

Deduciamo che
$$\begin{cases} \mathbb{E}^{p^0}[\phi_{M,\gamma}] = \mathbb{E}^{p^0}[\phi^*] \\ \mathbb{E}^{p^1}[\phi_{M,\gamma}] = \mathbb{E}^{p^1}[\phi^*] \end{cases}$$
, per cui la funzione integrabile $(\phi^* - \phi_{M,\gamma})(p^1 - Mp^0)$ è ≤ 0 e ha μ-integrale nullo su tutto Ω: dunque $(\phi^* - \phi_{M,\gamma})(p^1 - Mp^0) = 0$ μ-q.o., per cui è evidente che ϕ^* sia di Neyman-Pearson di costante M. //

[4] È una immediata conseguenza di (b)⇒(a) precedente, in quanto, se ϕ^* è ammissibile rispetto ai test ϕ con $\mathbb{E}^{p^0}[\phi] \overset{(*)}{\leq} \mathbb{E}^{p^0}[\phi^*]$, allora di certo non potrebbe essere $\mathbb{E}^{p^1}[\phi] > \mathbb{E}^{p^1}[\phi^*]$, altrimenti ϕ sarebbe

strettamente preferibile a ϕ^*, e per così dovremmo $E^{P_1}[\phi] \le E^{P_1}[\phi^*]$. ✓

(5) È sempre un corollario di (b) ⇒ (c) (e cioè della solita disug. fondamentale) in quanto, se ϕ^* è test di Neyman-Pearson e se $\phi \equiv E^{P_0}[\phi^*]$ (costante), allora essendo banalmente $E^{P_0}[\phi] = E^{P_0}[\phi^*]$ segue che anche $E^{P_0}[\phi^*] \overset{(c)}{=} E^{P_1}[\phi] \le E^{P_1}[\phi^*]$. ✓ □

NOTA(BENE). Dunque, sempre a riguardo di un modello stat. "binario" $(\Omega, \mathcal{F}, \{P^0, P^1\}, \mu)$ e del relativo problema $\mathcal{T} =$ «$H_0: \theta=0$ contro $H_1: \theta=1$», tutte le buone proprietà di un test di Neyman-Pearson ϕ^* di costante $M \ge 0$ (per $\mathcal{T}$), ricavate dal teorema precedente, hanno fondamento sostanzialmente dalla disuguaglianza $E^{P_1}[\phi] - E^{P_1}[\phi^*] \le M\,(E^{P_0}[\phi] - E^{P_0}[\phi^*])$, valida per ogni altro test aleatorio ϕ: prima che tutte, l'ammissibilità e l'"UPP" nella classe dei test ϕ aventi taglia $E^{P_0}[\phi] \le E^{P_0}[\phi^*]$, e l'ottimalità tra quei ϕ di questo avente taglia esattamente $E^{P_0}[\phi] = E^{P_0}[\phi^*]$, da cui in particolare le proprietà di correttezza di ϕ^* $E^{P_0}[\phi^*] \le E^{P_1}[\phi^*]$. ✓ Inoltre, comunque scelto $\alpha \in (0,1)$ (con $\alpha < 1$), conosciamo benissimo un test di Neyman-Pearson ϕ^* che abbia taglia $E^{P_0}[\phi^*] = \alpha$; (possiamo) e proposito, considerando di nuovo tutti quei test ϕ aventi taglia $E^{P_0}[\phi] \le E^{P_0}[\phi^*] = \alpha$, se anche ne trovassimo uno con $E^{P_0}[\phi] < E^{P_0}[\phi^*] = \alpha$, allora e questo $E^{P_1}[\phi] < E^{P_1}[\phi^*]$ e (quindi) ϕ^* potrebbe benissimo esser considerato come "preferibile" a ϕ, giacché le taglie dei test in gioco resta in ogni caso davvero bassa. Insomma, anche ragionando per un modello stat. astratto $(\Omega, \mathcal{F}, \{P^\theta\}_{\theta \in \Theta})$

e per il relativo problema $\mathcal{T} = $ « $\mathcal{A}_0 : \vartheta \in \Theta_0$ contro $\mathcal{A}_1 : \vartheta \notin \Theta_0$ », con $\phi \subsetneq \Theta_0 \subsetneq \Theta$ pre-scelto, stiamo pensando di poterci accontentare della relazione di UPP e fare a meno di quella di preferibilità (che è "migliore") all'interno di una classe di test ϕ (per $\mathcal{T}$) di "taglia molto bassa".
Spieghiamoci meglio: una volta fissato $\alpha \in (0,1)$ con $\alpha \ll 1$, ed una volta deciso di scegliere come test risolvente ϕ^* (per $\mathcal{T}$) uno che dei test ϕ aventi taglia $\sup_{\vartheta \in \Theta_0} \mathbb{E}^{P_\vartheta}[\phi] \leq \alpha$, sebbene potremmo considerare ϕ^* "migliore" di un certo ϕ di questi se ϕ^* è più potente di ϕ (, cioè se $\mathbb{E}^{P_\vartheta}[\phi] \leq \mathbb{E}^{P_\vartheta}[\phi^*]$ $\forall \vartheta \notin \Theta_0$), anche se ϕ^* avesse taglia superiore a quella di ϕ (, cioè $\mathbb{E}^{P_\vartheta}[\phi] \leq \mathbb{E}^{P_\vartheta}[\phi^*]$ $\forall \vartheta \in \Theta_0$ ed $\exists \vartheta_0 \in \Theta_0$ tale che $\mathbb{E}^{P_{\vartheta_0}}[\phi] < \mathbb{E}^{P_{\vartheta_0}}[\phi^*]$) ; per cui, in realtà, potrebbe anche succedere che tale ϕ sia, per definizione, strettamente preferibile a ϕ^* : quando esattamente

$$\begin{cases} \mathbb{E}^{P_\vartheta}[\phi] \leq \mathbb{E}^{P_\vartheta}[\phi^*] \ \forall \vartheta \in \Theta_0, \text{ ed } \exists \vartheta_0 \in \Theta_0 \text{ t.c. } \mathbb{E}^{P_{\vartheta_0}}[\phi] < \mathbb{E}^{P_{\vartheta_0}}[\phi^*] \\ \mathbb{E}^{P_\vartheta}[\phi] = \mathbb{E}^{P_\vartheta}[\phi^*] \ \forall \vartheta \notin \Theta_0 \end{cases}$$

Tuttavia, in un caso del genere, il guadagno di scegliere ϕ invece che ϕ^* consisterebbe (solo) nello scegliere un test di taglia un poco più bassa di un'altra taglia già molto bassa : ecco allora che potremmo sentirci molto più interessati a dei test ϕ^* UPP di ogni altro test ϕ di questi, possibilmente in senso stretto (cioè $\mathbb{E}^{P_\vartheta}[\phi] \leq \mathbb{E}^{P_\vartheta}[\phi^*]$ $\forall \vartheta \notin \Theta_0$ ed $\exists \bar{\vartheta} \notin \Theta_0$ tale che $\mathbb{E}^{P_{\bar{\vartheta}}}[\phi] < \mathbb{E}^{P_{\bar{\vartheta}}}[\phi^*]$ $(\forall \phi)$), $(\bar{\vartheta}_\phi)$ cosicché la taglia molto bassa forse neanche più un problema a maggior ragione, essendo in tal caso ϕ^* (anche) ammissibile.

Riassumendo : per due test di "taglia molto bassa", potremmo intendere sempre la relazione di preferibilità come la (più generale) relazione di maggior potenza. In questo senso, all'interno di una pre-fissata classe di test ϕ di taglia molto bassa, uno di tali test ϕ^* che risulti UPP di ogni altro test ϕ di questo può essere considerato come ottimale.

Il fatto è (pure) che adesso dovremmo provare a portare alla teoria estesa di test "buoni" per modelli stat. ben più complessi di quello binario, per cui ogni risultato teorico, o anche solo ogni idea o punto di vista, che siano resi più generali, più ampi, più elastici potrebbero tornarci comodo, come per esempio effettuare un'assunta elasticità nel giudicare un test migliore di un altro test.

Detto questo, l' $\boxed{\text{IDEA}}$ ora è quella di proseguire con la nostra teoria appoggiandoci proprio su questo appreso dei test di Neyman-Pearson su modelli stat. binari, lavorando per ciò con modelli stat. dominati $(\Omega, \mathcal{F}, \{P^\theta | \theta \in \Theta\}, \mu)$ aventi $\Theta \subseteq \mathbb{R}$ con $\#\Theta \geq 3$, idealmente infinito.

Infatti, visto che con Neyman-Pearson abbiamo affrontato il problema di decidere su quale delle due probabilità P^0 e P^1 su $\mathcal{F}$ fosse la "migliore" nel modellizzare quanto stiamo cercando di modellizzare, è naturale ammettere l'eventualità che entrambe una fra P^0 e P^1 sono migliori in questo senso, ma che nessuna delle due sia esattamente quella "giusta" P^{θ^*}, come se fosse stato possibile ~~non~~ NON considerarla in principio: è naturale quindi il tentare opportunamente di ragionare come "con Neyman-Pearson" per ogni coppia $P^{\theta_1}, P^{\theta_2}$ con $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$, idealmente "eliminando" via via le probabilità che così facendo sembrano meno soddisfacenti, e restando infine con le probabilità più "vicine" alle P^{θ^*} "giuste". In effetto, Potremmo pensare alla situazione ideale di una "buona" stima $\hat{\vartheta}(\omega^*) \approx \vartheta^*$: in questo caso, è molto naturale ritenere ogni P^ϑ con $\vartheta \approx \hat{\vartheta}(\omega^*)$ "migliore" rispetto a tutte le $P^{\vartheta'}$ con $\vartheta' \not\approx \hat{\vartheta}(\omega^*)$, da cui l'idea appunto del confronto per coppie. ✓

Vediamo pertanto con un esempio come potremmo ingegnarci per far funzionare bene le cose. Capiremo presto che converrà parecchio supporre una ben precisa relazione "algebrico-probabilistica" fra le densità $L^\vartheta(\omega) := \frac{dP^\vartheta}{d\mu}(\omega)$, $(\vartheta,\omega)\in\Theta\times\Omega$, del nostro modello stat.

ES. Proviamo a ragionare su di un modello stat. dominato "trivariato" $(\Omega,\mathcal{F},\{P^{\vartheta_0},P^{\vartheta_1},P^{\vartheta_2}\},\mu)$, essendo $\vartheta_0,\vartheta_1,\vartheta_2\in\mathbb{R}$ con $\vartheta_0<\vartheta_1<\vartheta_2$. (In particolare, potremmo immaginare μ come una combinazione convessa non-banale delle P^{ϑ_i}, $i=0,1,2$.) Dunque, quale che sia il problema di test $\mathcal{T}$ che vogliamo porci, abbiamo l'idea di poterlo risolvere ponendo dapprima per i seguenti modelli stat. bivariati, coi relativi "problemi di Neyman-Pearson":

$\mapsto (\Omega,\mathcal{F},\{P^{\vartheta_0},P^{\vartheta_1}\},\mu)$, $\mathcal{T}^{\vartheta_0,\vartheta_1} := \ll \mathcal{H}_0: \vartheta=\vartheta_0$ contro $\mathcal{H}_1: \vartheta=\vartheta_1 \gg$;

$\mapsto (\Omega,\mathcal{F},\{P^{\vartheta_0},P^{\vartheta_2}\},\mu)$, $\mathcal{T}^{\vartheta_0,\vartheta_2} := \ll \mathcal{H}_0: \vartheta=\vartheta_0$ contro $\mathcal{H}_1: \vartheta=\vartheta_2 \gg$;

$\mapsto (\Omega,\mathcal{F},\{P^{\vartheta_1},P^{\vartheta_2}\},\mu)$, $\mathcal{T}^{\vartheta_1,\vartheta_2} := \ll \mathcal{H}_0: \vartheta=\vartheta_1$ contro $\mathcal{H}_1: \vartheta=\vartheta_2 \gg$.

L'IDEA sembrerebbe essere quella di trovare un test aleatorio ϕ^* per $\mathcal{T}$ che sia anche un test di Neyman-Pearson per ciascuno dei tre sotto-problemi (di cui opportune costanti) $\mathcal{T}^{\vartheta_0,\vartheta_1}$, $\mathcal{T}^{\vartheta_0,\vartheta_2}$, $\mathcal{T}^{\vartheta_1,\vartheta_2}$, cosicché per ogni altro test ϕ per $\mathcal{T}$ valgano

$$\begin{cases} \mathbb{E}^{P^{\vartheta_1}}[\phi]-\mathbb{E}^{P^{\vartheta_1}}[\phi^*] \le M^{\vartheta_0,\vartheta_1}\left(\mathbb{E}^{P^{\vartheta_0}}[\phi]-\mathbb{E}^{P^{\vartheta_0}}[\phi^*]\right) \\ \mathbb{E}^{P^{\vartheta_2}}[\phi]-\mathbb{E}^{P^{\vartheta_2}}[\phi^*] \le M^{\vartheta_0,\vartheta_2}\left(\mathbb{E}^{P^{\vartheta_0}}[\phi]-\mathbb{E}^{P^{\vartheta_0}}[\phi^*]\right) \\ \mathbb{E}^{P^{\vartheta_2}}[\phi]-\mathbb{E}^{P^{\vartheta_2}}[\phi^*] \le M^{\vartheta_1,\vartheta_2}\left(\mathbb{E}^{P^{\vartheta_1}}[\phi]-\mathbb{E}^{P^{\vartheta_1}}[\phi^*]\right) \end{cases}$$

, per opportune costanti $M^{\vartheta_0,\vartheta_1}, M^{\vartheta_0,\vartheta_2}, M^{\vartheta_1,\vartheta_2} \ge 0$: così, avremo per qualche $\mathcal{T}$, ϕ^*

dovrebbe risultare "preferibile" rispetto a "tanti altri" test ϕ. (Osserviamo subito che, sempre in tale caso, otteniamo subito che $E^{p_{\theta_0}}[\phi^*] \le \le E^{p_{\theta_1}}[\phi^*] \le E^{p_{\theta_2}}[\phi^*]$.) Dunque, cominciamo per esempio col problema $\mathcal{T} =$ « $H_0: \theta=\theta_0$ contro $\underset{(o\ \theta=\theta_1\ o\ \theta=\theta_2)}{H_1: \theta\neq\theta_0}$ », = « $H_0: \theta\le\theta_0$ contro $H_1: \theta>\theta_0$ ».
Comunque possa essere scelto un test aleatorio ϕ per $\mathcal{T}$, ϕ avrebbe taglia $E^{p_{\theta_0}}[\phi] = \int_\Omega \phi(\omega) L^{\theta_0}(\omega)\, d\mu(\omega)$ e potenza "data da"

$$\begin{cases} E^{p_{\theta_1}}[\phi] = \int_\Omega \phi L^{\theta_1}\, d\mu \\ E^{p_{\theta_2}}[\phi] = \int_\Omega \phi L^{\theta_2}\, d\mu \end{cases}$$

, per cui un'idea potrebbe "essere" quella di "basare" ϕ su eventi della forma $\{L^{\theta_1} > M^{\theta_0,\theta_1} L^{\theta_0}\} \cap \{L^{\theta_2} > M^{\theta_0,\theta_2} L^{\theta_0}\}$, $M^{\theta_0,\theta_2}, M^{\theta_0,\theta_1} \ge 0$, o anche, volendo (eventualmente) essere "più precisi", su eventi della forma $\{L^{\theta_1} > M^{\theta_0,\theta_1} L^{\theta_0}\} \cup \{L^{\theta_2} > M^{\theta_0,\theta_2} L^{\theta_0}\}$.
Tuttavia, anche considerando $D_M = \{L^{\theta_1} > M L^{\theta_0}\} \cup \{L^{\theta_2} > M L^{\theta_0}\}$ per uno stesso $M \ge 0$, NON è affatto scontata la scelta di un test ϕ_M (basato su D_M) che soddisfi una "disuguaglianza fondamentale" come quella che invece avevamo, in analogia ai modelli stat. binari e ai test di Neyman–Pearson ...!
(Esempio banale: se fosse $L^{\theta_1}(\omega) \le L^{\theta_2}(\omega)$ $\forall \omega \in \Omega$ (ma $L^{\theta_1} \not\equiv L^{\theta_2}$), allora $D_M = \underset{(L^{\theta_1}\ \text{con l'}\Omega)}{\{L^{\theta_2} > M L^{\theta_0}\}}$, $M \ge 0$, sarebbe più ben poco "familiare" nel senso dei test di Neyman–Pearson!) Inoltre, (anche) alla luce dell'idea (iniziale) di trovare un test ϕ^* per $\mathcal{T}$ che sia di Neyman–Pearson per ciascuno dei problemi "binari" relativi, basare ϕ^* su $\{L^{\theta_1} > M^{\theta_0,\theta_1} L^{\theta_0}\} \cup \{L^{\theta_2} > M^{\theta_0,\theta_2} L^{\theta_0}\}$ renderebbe ϕ^* stesso "indipendente" dalle coppie $L^{\theta_1}, L^{\theta_2}$, il che non pare di certo una buona cosa ...! Viene così l'idea di supporre un qualche, opportuno legame tra le (coppie di) v.a.o.r. $L^{\theta_0}, L^{\theta_1}$, $L^{\theta_0}, L^{\theta_2}$, e $L^{\theta_1}, L^{\theta_2}$.

Con lo scopo appunto di trovare una "ragionevole" ipotesi sulle verosimiglianze, passiamo per il momento al caso di un modello stat. (dominato) dotato di una statistica esaustiva reale $T(\omega): \Omega \to \mathbb{R}$: così, per Neyman-Fischer, esistono $h(\omega): \Omega \to [0,+\infty)$ e, $\forall \theta \in \Theta$, $g^{\theta}(x): \mathbb{R} \to [0,+\infty)$ misurabili, tali che

$\mathcal{L}^{\theta}(\omega) = g^{\theta}(T(\omega))\, h(\omega) \quad \forall (\theta,\omega) \in \Theta \times \Omega$, quindi, nel nostro semplice caso,

$$\begin{cases} \mathcal{L}^{\theta_0}(\omega) = g^{\theta_0}(T(\omega))\, h(\omega) \\ \mathcal{L}^{\theta_1}(\omega) = g^{\theta_1}(T(\omega))\, h(\omega) \\ \mathcal{L}^{\theta_2}(\omega) = g^{\theta_2}(T(\omega))\, h(\omega) \end{cases}$$

, $\forall \omega \in \Omega$. Pertanto, deduciamo subito che, immaginando che oltre che $g^{\theta_0}(x), g^{\theta_1}(x)$ siano $\neq 0$ (>0) per ogni $x \in \mathbb{R}$,

$$\begin{cases} \mathcal{L}^{\theta_1}(\omega) = \mathcal{L}^{\theta_0}(\omega) \cdot \dfrac{g^{\theta_1}(T(\omega))}{g^{\theta_0}(T(\omega))} \\ \mathcal{L}^{\theta_2}(\omega) = \mathcal{L}^{\theta_0}(\omega) \cdot \dfrac{g^{\theta_2}(T(\omega))}{g^{\theta_0}(T(\omega))} \\ \left(\mathcal{L}^{\theta_2}(\omega) = \mathcal{L}^{\theta_1}(\omega) \cdot \dfrac{g^{\theta_2}(T(\omega))}{g^{\theta_1}(T(\omega))} \right) \end{cases}$$

, $\forall \omega \in \Omega$: ovvero, ponendo $f^{\theta_0,\theta_1} \doteq \dfrac{g^{\theta_1}}{g^{\theta_0}}$, $f^{\theta_0,\theta_2} \doteq \dfrac{g^{\theta_2}}{g^{\theta_0}}$ e $f^{\theta_1,\theta_2} \doteq \dfrac{g^{\theta_2}}{g^{\theta_1}}$, allora avremo appunto, $\forall \omega \in \Omega$,

$$\begin{cases} \mathcal{L}^{\theta_1}(\omega) = f^{\theta_0,\theta_1}(T(\omega))\, \mathcal{L}^{\theta_0}(\omega) \\ \mathcal{L}^{\theta_2}(\omega) = f^{\theta_0,\theta_2}(T(\omega))\, \mathcal{L}^{\theta_0}(\omega) \\ \mathcal{L}^{\theta_2}(\omega) = f^{\theta_1,\theta_2}(T(\omega))\, \mathcal{L}^{\theta_1}(\omega) \end{cases}$$

, dove ovviamente $f^{\theta_0,\theta_1} \cdot f^{\theta_1,\theta_2} = f^{\theta_0,\theta_2}$.

Ecco finalmente un' $\boxed{\text{IDEA}}$: visto che la statistica T accomuna "parecchio" tutte le coppie $\mathcal{L}^{\theta_1}, \mathcal{L}^{\theta_0}$, $\mathcal{L}^{\theta_2}, \mathcal{L}^{\theta_0}$ e $\mathcal{L}^{\theta_2}, \mathcal{L}^{\theta_1}$, e ricordando che il problema $\mathcal{T}$ una volta pone i tre sotto-problemi $\mathcal{T}^{\theta_0,\theta_1}, \mathcal{T}^{\theta_0,\theta_2}$ e $\mathcal{T}^{\theta_1,\theta_2}$, potremmo supporre $f^{\theta_0,\theta_1}, f^{\theta_0,\theta_2}$ e f^{θ_1,θ_2} <u>MONOTÒNE</u> su $\mathbb{R}$

(non necessariamente in senso stretto), o meglio suppole o tutte crescenti o tutte decrescenti. Infatti, sotto queste ipotesi, potremmo basare il test randomizzato (test aleatorio) ϕ^* su una comune tipologia di eventi: o quello della forma $\{T > c\}$, o quello della forma $\{T < c\}$, con $c \in \mathbb{R}$, a seconda delle crescenze o decrescenze rispettivamente delle f^{θ_i,θ_j}!

$\longrightarrow$ **NOTA** Per i nostri scopi, dunque, non ha alcuna importanza che T sia esaustiva, quanto piuttosto il poter basare un test ϕ^* su degli eventi che non dipendono dai parametri: per ciò, ci basta che, $\forall \theta', \theta'' \in \Theta$ con $\theta' < \theta''$, esistano $f^{\theta',\theta''}(x) : \mathbb{R} \to [0, +\infty)$ o sempre crescenti o sempre decrescenti, e che esiste una statistica $T : \Omega \to \mathbb{R}$ reale, tale che

$$\boxed{L^{\theta''}(\omega) = f^{\theta',\theta''}(T(\omega))\, L^{\theta'}(\omega)} \quad \forall \theta' < \theta'' \text{ in } \Theta \text{ e } \forall \omega \in \Omega.$$

Ma allora, in una situazione come questa, a meno di sostituire T con $-T$, potremmo sempre supporre che, $\forall \theta', \theta'' \in \Theta$ con $\theta' < \theta''$, $f^{\theta',\theta''}$ sia crescente su $\mathbb{R}$.

[Infatti, semplicemente, se anche le $f^{\theta',\theta''}$ fossero decrescenti su $\mathbb{R}$, $\theta' < \theta''$ in Θ, allora ovviamente le funzioni $\tilde{f}^{\theta',\theta''}(x) : \mathbb{R} \to [0, +\infty)$, $\tilde{f}^{\theta',\theta''}(x) := f^{\theta',\theta''}(-x)$ per ogni $x \in \mathbb{R}$, $\theta' < \theta''$ in Θ, sarebbero crescenti e tali che $L^{\theta''}(\omega) = \tilde{f}^{\theta',\theta''}(-T(\omega)) L^{\theta'}(\omega)$ $\forall \theta' < \theta''$ in Θ e $\forall \omega \in \Omega$, ovviamente.]

(**ALTRA NOTICINA** L'essere f^{θ_0,θ_1}, f^{θ_0,θ_2} e f^{θ_1,θ_2} tutte crescenti NON crea alcuna contraddizione con l'uguaglianza $f^{\theta_0,\theta_1} \cdot f^{\theta_1,\theta_2} = f^{\theta_0,\theta_2}$, in quanto invece il prodotto di funzioni crescenti non-negative resta una funzione crescente (non-negativa).)

Pertanto, per il nostro modello stat. dominato finito $(\Omega, \mathcal{A}, \{P^{\theta_0}, P^{\theta_1}, P^{\theta_2}\}, \mu)$, $\theta_0 < \theta_1 < \theta_2$ in $\mathbb{R}$, supponiamo l'esistenza di una statistica reale $T : \Omega \to \mathbb{R}$ e di tre funzioni crescenti non-negative f^{θ_0,θ_1}, f^{θ_0,θ_2} e $f^{\theta_1,\theta_2} : \mathbb{R} \to [0, +\infty)$ tali che valga, $\forall \omega \in \Omega$,

$$\begin{cases} L^{\theta_1}(\omega) = f^{\theta_0,\theta_1}(T(\omega))\, L^{\theta_0}(\omega) \\ L^{\theta_2}(\omega) = f^{\theta_0,\theta_2}(T(\omega))\, L^{\theta_0}(\omega) \\ L^{\theta_2}(\omega) = f^{\theta_1,\theta_2}(T(\omega))\, L^{\theta_1}(\omega) \end{cases}$$ (da cui, in realtà, $f^{\theta_0,\theta_2} \equiv f^{\theta_0,\theta_1} \cdot f^{\theta_1,\theta_2}$).

Allora, per ogni $\theta', \theta'' \in \Theta = \{\theta_0, \theta_1, \theta_2\}$ con $\theta' < \theta''$ (o, cioè, θ' è θ_0 o θ_1 mentre θ'' è θ_1 o θ_2, e secondo), vale che, $\forall C \in \mathbb{R}$,

$$\begin{cases} \{L^{\theta''} > f^{\theta',\theta''}(C)\, L^{\theta'}\} \subseteq \{T > C\} \\ \{L^{\theta''} < f^{\theta',\theta''}(C)\, L^{\theta'}\} \subseteq \{T < C\} \end{cases}$$

, e quindi chiaramente il test aleatorio $\boxed{\Phi_{C,\gamma} \doteq 1\!\!1_{\{T>C\}} + \gamma 1\!\!1_{\{T=C\}}}$, $C \in \mathbb{R}$ e $\gamma \in [0,1]$, è un test di Neyman-Pearson per il modello stat. binario $(\Omega, \mathcal{A}, \{p^{\theta'}, p^{\theta''}\}, \mu)$ ed il relativo problema $\mathcal{T}^{\theta',\theta''} = $ «$H_0: \theta = \theta'$ contro $H_1: \theta = \theta''$», precisamente di costante $M^{\theta',\theta''} := f^{\theta',\theta''}(C)$.

[Infatti, $\forall \omega \in \Omega$, $T(\omega) \leq C \Rightarrow f^{\theta',\theta''}(T(\omega)) \leq f^{\theta',\theta''}(C)$, $\Rightarrow L^{\theta''}(\omega) = f^{\theta',\theta''}(T(\omega)) L^{\theta'}(\omega) \leq$ $\leq f^{\theta',\theta''}(C) L^{\theta'}(\omega)$. / È evidente l'analogia per l'altro $\subseteq$. /]

Per ciò, segue subito che $\mathbb{E}^{p^{\theta''}}[\Phi] - \mathbb{E}^{p^{\theta''}}[\Phi_{C,\gamma}] \leq M^{\theta',\theta''}\left(\mathbb{E}^{p^{\theta'}}[\Phi] - \mathbb{E}^{p^{\theta'}}[\Phi_{C,\gamma}]\right)$, e che $\mathbb{E}^{p^{\theta_0}}[\Phi_{C,\gamma}] \leq \mathbb{E}^{p^{\theta_1}}[\Phi_{C,\gamma}] \leq \mathbb{E}^{p^{\theta_2}}[\Phi_{C,\gamma}]$, quali che siano $C \in \mathbb{R}$, $\gamma \in [0,1]$ e Φ altro test aleatorio. / Tornando quindi al problema iniziale $\mathcal{T} = $ «$H_0: \theta = \theta_0$ contro $H_1: \theta \neq \theta_0$», ogni test aleatorio Φ per $\mathcal{T}$ ha taglia $\mathbb{E}^{p^{\theta_0}}[\Phi]$ e potenza data da $\mathbb{E}^{p^{\theta_1}}[\Phi]$ e $\mathbb{E}^{p^{\theta_2}}[\Phi]$, per cui è ormai chiaro che ogni $\Phi_{C,\gamma}$ sia <u>ammissibile ed UPP</u> fra tutti i Φ che abbiano taglia $\mathbb{E}^{p^{\theta_0}}[\Phi] \leq \mathbb{E}^{p^{\theta_0}}[\Phi_{C,\gamma}]$, ed <u>ottimale</u> nel caso di $=$. / Inoltre, tali $\Phi_{C,\gamma}$ sono corretti e hanno il pregio che, comunque scelto un $\alpha \in (0,1)$ (con $\alpha \ll 1$), esistono $C \doteq C_\alpha \in \mathbb{R}$ e $\gamma \doteq \gamma_\alpha \in [0,1]$ tali per cui $\mathbb{E}^{p^{\theta_0}}[\Phi_{C,\gamma}] = \alpha$, ragionando in modo del tutto

analogo alla dimostrazione del punto $\boxed{I}$ del teorema di Neyman–Pearson.
Insomma, tali $\phi_{c,\gamma}$ ci forniscono veramente.

(8?) Se invece ci ponessimo $\mathcal{T} =$ « $H_0: \underset{(\theta=\theta_0,\,\theta=\theta_1)}{\theta \le \theta_1}$ contro $H_1: \underset{(\theta=\theta_2)}{\theta > \theta_1}$ » ? In questo caso, le regioni "critiche" dovrebbero essere della forma $\{L^{\theta_2} > \text{cost.}\, L^{\theta_0}\} \cup \{L^{\theta_2} > \text{cost.}\, L^{\theta_1}\}$, dunque ci sarebbe (potrebbe tornare) molto comodo il "rivedere" gli esiti del tipo $\{L^{\theta''} > \text{cost.}\, L^{\theta'}\}$ con $\theta' < \theta''$ "indipendenti" da θ', θ'' stessi ...! Supponendo per ciò di essere (dunque) nella situazione precedente per il modello, ogni test aleatorio ϕ per $\mathcal{T}$ che sia di taglia $\max\{\mathbb{E}^{P_{\theta_0}}[\phi], \mathbb{E}^{P_{\theta_1}}[\phi]\}$ e potenza $\mathbb{E}^{P_{\theta_2}}[\phi]$, ed è immediato riconoscere quanto segue: ogni test $\phi_{c,\gamma}$ che taglia $\mathbb{E}^{P_{\theta_1}}[\phi_{c,\gamma}]$ (volendo $= \alpha \,(\ll 1)$ per opportuni c_α e γ_α), è corretto ed è ammissibile ed UPP che tutti i test ϕ aventi taglia $\le \mathbb{E}^{P_{\theta_1}}[\phi_{c,\gamma}]$, ma NON necessariamente ottimale (nemmeno) nel caso di $=$ (invece sì per quei ϕ aventi $\mathbb{E}^{P_{\theta_0}}[\phi] \le \mathbb{E}^{P_{\theta_1}}[\phi]$). Pertanto, anche per questo $\mathcal{T}$ possiamo ritenerci soddisfatti. ✓

⟶ Mentre per $\mathcal{T} =$ « $H_0: \underset{(\theta=\theta_1,\,\theta=\theta_2)}{\theta \ge \theta_1}$ contro $H_1: \underset{(\theta=\theta_0)}{\theta < \theta_2}$ » ? Dovrebbe essere ben fattibile ragionare in modo analogo ai ~~questo~~ due precedenti esempi di problemi, in quanto ~~per~~ questo $\mathcal{T}$ è esattamente il problema "speculare" del primo analizzato, o cioè di « $H_0: \theta = \theta_0$ contro $H_1: \theta \neq \theta_0$ », ed infatti le relative regioni critiche stavolta dovrebbero essere della forma
$\{L^{\theta_0} > \text{cost.} \cdot L^{\theta_1}\} \cup \{L^{\theta_0} > \text{cost.} \cdot L^{\theta_2}\}$, o cioè della forma
$\{L^{\theta_1} < \text{cost.} \cdot L^{\theta_0}\} \cup \{L^{\theta_2} < \text{cost.} \cdot L^{\theta_0}\}$, tale per cui insieme contano quelle regioni di Ω sulle quali vale la stessa disuguaglianza
$L^{\theta''} < \text{cost.} \cdot L^{\theta'}$ per ~~tutti~~ i $\theta', \theta'' \in \Theta$ che rientrano con $\theta' < \theta''$:

pertanto, in questo caso potremo «basare» i candidati test aleatori «buoni» sugli eventi del tipo $\{T<C\}$, $C\in\mathbb{R}$, com'è chiaro per diverse ragioni. (⚠ Per $\mathcal{T} \doteq \{H_0: \theta=\theta_1$ contro $H_1: \theta\neq\theta_1\}$ (invece $(\theta\notin\{\theta_1\})$), NON potremo sperare che la monotonia delle $f^{\theta',\theta''}$ ci fornisca qualcosa, $\theta'<\theta''$ in Θ, ossia: adesso le «buone» regioni critiche per $\mathcal{T}$ paiono della forma $\{L^{\theta_0} > \text{cost.}\cdot L^{\theta_1}\}\cup$ $\cup\{L^{\theta_2}>\text{cost.}\cdot L^{\theta_1}\}$, «$\subseteq \{T<C_{cost}\}\cup\{T>C_{cost}\}$» ($C\in\mathbb{R}$ opportuno) !!

Riassumendo, la monotonia delle $f^{\theta',\theta''}$, $\theta'<\theta''$ in Θ, è una ragionevole ipotesi (e comoda) per risolvere problemi in cui ci si chiede se il parametro «giusto» stia o sotto o sopra di una certa soglia, e NON sia sotto che sopra di due soglie, anziché no. //

Ci sentiamo ispirati per generalizzare: mettiamoci dunque su di un modello stat. dominato $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\},\mu)$, possibilmente regolare, avente $\Theta\subseteq\mathbb{R}$ con $\#\Theta\geq 3$, possibilmente Θ infinito, e per ogni fissato $\vartheta_0\in\Theta$ consideriamo il problema di test basato su $\Theta_0 \doteq (-\infty,\vartheta_0]\cap\Theta$ (in modo che può risultare $\Theta_0\neq\Theta$), ossia

$$\boxed{\mathcal{T}\doteq \text{« } H_0:\vartheta\le\vartheta_0 \text{ contro } H_1:\vartheta>\vartheta_0 \text{ »}}$$

: ogni test aleatorio $\phi(\omega)$ per $\mathcal{T}$ è UN TEST UNILATERALE (per il modello). Dunque, è da intendersi come test unilaterale anche ogni test per un problema $\mathcal{T}'$ basato su $\Theta_0' \doteq [\vartheta_0,+\infty)\cap\Theta$ (se $\neq\Theta$, di nuovo), $\vartheta_0\in\Theta$, ossia $\mathcal{T}'\doteq$ « $H_0:\vartheta\ge\vartheta_0$ contro $H_1:\vartheta<\vartheta_0$ ». Infatti, visto che ovviamente $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\},\mu) = (\Omega,\mathcal{F},\{P^{-\theta'}|\theta'\in-\Theta\},\mu)$, essendo $-\Theta \doteq \{-\theta|\theta\in\Theta\}$, il porci un problema del tipo $\mathcal{T}' =$ « $H_0:\theta'\le\theta_0'$ contro $H_1:\theta'>\theta_0'$ »
$(-\theta\le-\vartheta_0)$ $(-\theta>-\vartheta_0)$

equivale al porsi un problema del tipo $\mathcal{T}' =$ «Ado: $\vartheta \geq \vartheta_0$ contro Ad$_1$: $\vartheta < \vartheta_0$» , effettuato, da cui anche l'equivalenza nelle rispettive risoluzioni. $\checkmark$

Concentriamoci comunque su $\mathcal{T}$, e supponiamo che il modello stat. sia A RAPPORTO DI VEROSIMIGLIANZA MONOTÒNO, o CRESCENTE o DECRESCENTE, in senso "debole" (non stretto), nel senso che esiste una statistica reale $T(\omega): \Omega \to \mathbb{R}$ ed esiste per ogni coppia $\vartheta_1, \vartheta_2 \in \Theta$ con $\vartheta_1 < \vartheta_2$ una funzione mis. non-negativa $f^{\vartheta_1,\vartheta_2}(x): \mathbb{R} \to [0,+\infty)$ o crescente o decrescente (stesso tipo di monotonia per ogni $\vartheta_1 < \vartheta_2$ in Θ) tale che $\boxed{L^{\vartheta_2}(\omega) = f^{\vartheta_1,\vartheta_2}(T(\omega)) \cdot L^{\vartheta_1}(\omega)}$ per ogni $\omega \in \Omega$.

In questo caso, per lo sostituire T con $-T$, possiamo supporre che le $f^{\vartheta_1,\vartheta_2}$ siano tutte crescenti, $\vartheta_1 < \vartheta_2$ in Θ. [Infatti, se anche fosse che $f^{\vartheta_1,\vartheta_2}$ siano tutte decrescenti, allora le $\tilde{f}^{\vartheta_1,\vartheta_2}(x) \doteq f^{\vartheta_1,\vartheta_2}(-x)$ (≥ 0) sarebbero tutte crescenti e tale che $L^{\vartheta_2} = \tilde{f}^{\vartheta_1,\vartheta_2}(-T) \cdot L^{\vartheta_1}$. $\checkmark$]

Allora lo supponiamo, e così $\forall C \in \mathbb{R}$ e $\forall \vartheta_1 < \vartheta_2$ in Θ, chiaramente vale

$$\begin{cases} \{L^{\vartheta_2} > f^{\vartheta_1,\vartheta_2}(C) L^{\vartheta_1}\} \subseteq \{T > C\} \\ \{L^{\vartheta_2} < f^{\vartheta_1,\vartheta_2}(C) L^{\vartheta_1}\} \subseteq \{T < C\} \end{cases}$$

per cui: per ogni $\vartheta_1 < \vartheta_2$ in Θ, per il modello stat. chiamato $(\Omega, \mathcal{F}, \{P^{\vartheta_1}, P^{\vartheta_2}\}, \mu)$ e per il relativo problema $\mathcal{T}^{\vartheta_1,\vartheta_2} =$ «Ado: $\vartheta = \vartheta_1$ contro Ad$_1$: $\vartheta = \vartheta_2$», ogni test aleatorio della forma $\boxed{\phi_{C,\gamma} \doteq \mathbb{1}_{\{T > C\}} + \gamma \mathbb{1}_{\{T = C\}}}$, $C \in \mathbb{R}$ e $\gamma \in [0,1]$, è un test di Neyman–Pearson di costante $f^{\vartheta_1,\vartheta_2}(C) =: M^{\vartheta_1,\vartheta_2}$, ed in particolare $E^{P^{\vartheta_2}}[\phi] - E^{P^{\vartheta_2}}[\phi_{C,\gamma}] \leq M^{\vartheta_1,\vartheta_2} \left(E^{P^{\vartheta_1}}[\phi] - E^{P^{\vartheta_1}}[\phi_{C,\gamma}] \right)$ per ogni altro test aleatorio ϕ, e $E^{P^{\vartheta_1}}[\phi_{C,\gamma}] \leq E^{P^{\vartheta_2}}[\phi_{C,\gamma}]$. Inoltre, per ogni fissato $\bar{\vartheta} \in \Theta$, e per ogni fissato $\alpha \in (0,1)$ (con $\alpha < 1$), mostriamo

$C \doteq C_\alpha \in \mathbb{R}$ e $\gamma \doteq \gamma_\alpha \in [0,1]$ tali che $E^{P^{\theta_0}}[\phi_{c,\gamma}] = \alpha$. Dunque, tornando al nostro modello stat. e al problema $\mathcal{T}$, ogni test aleatorio $\phi(\omega)$ per $\mathcal{T}$ che taglia $\sup_{\theta \le \theta_0} E^{P^\theta}[\phi]$ e potenza data da $E^{P^\theta}[\phi]$ per $\theta > \theta_0$, quando, per ogni $C \in \mathbb{R}$ e $\gamma \in [0,1]$, il test $\phi_{c,\gamma}$ che taglia

$$\sup_{\theta \le \theta_0} E^{P^\theta}[\phi_{c,\gamma}] = E^{P^{\theta_0}}[\phi_{c,\gamma}] \le E^{P^\theta}[\phi_{c,\gamma}] \quad \forall \theta > \theta_0,$$

per cui $\phi_{c,\gamma}$ è corretto e, volendo, possiamo comunque che $E^{P^{\theta_0}}[\phi_{c,\gamma}] = \alpha$ per un pre-scelto $\alpha \in (0,1)$ con $\alpha \ll 1$ (e $C \doteq C_\alpha \in \mathbb{R}$, $\gamma \doteq \gamma_\alpha \in [0,1]$, opportuni).
Inoltre, per tutti i test ϕ aventi taglia $\sup_{\theta \le \theta_0} E^{P^\theta}[\phi] \le E^{P^{\theta_0}}[\phi_{c,\gamma}]$ (risp. $<$), vale che, $\forall \theta > \theta_0$, $E^{P^\theta}[\phi] \le E^{P^\theta}[\phi_{c,\gamma}]$ (risp. $<$), da cui risulta il fatto che $\phi_{c,\gamma}$ è ammissibile ed UPP (ma non ottimale, in generale) per tutti i test ϕ di taglia $\le E^{P^{\theta_0}}[\phi_{c,\gamma}]$, e pertanto siamo soddisfatti. //

(ES.) Mettiamoci in un modello stat. esponenziale (ma con $n=1$): supponiamo cioè che la misura μ su $(\Omega, \mathcal{A})$ e l'insieme $\Theta \subseteq \mathbb{R}$ siano tali che Θ sia aperto e convesso in $\mathbb{R}$, e che esiste una statistica $T : \Omega \to \mathbb{R}$ reale tale che risulti $0 < \int_\Omega e^{\theta T(\omega)} d\mu(\omega) < +\infty$ per ogni $\theta \in \Theta$, cosicché ponendo $f(\theta) \doteq \log\left(\int_\Omega e^{\theta \cdot T(\omega)} d\mu(\omega)\right)$, $\theta \in \Theta$, supponiamo che sia

$$L^\theta(\omega) = \exp\left(\theta \cdot T(\omega) - f(\theta)\right)$$ per ogni $(\theta, \omega) \in \Theta \times \Omega$. Allora $L^\theta(\omega)$ è sempre >0, e per ogni coppia $\theta_1 < \theta_2$ in Θ vale

$$\frac{L^{\theta_2}(\omega)}{L^{\theta_1}(\omega)} = \exp\left((\theta_2 - \theta_1) T(\omega)\right) \exp\left(f(\theta_1) - f(\theta_2)\right)$$ per ogni $\omega \in \Omega$, ossia

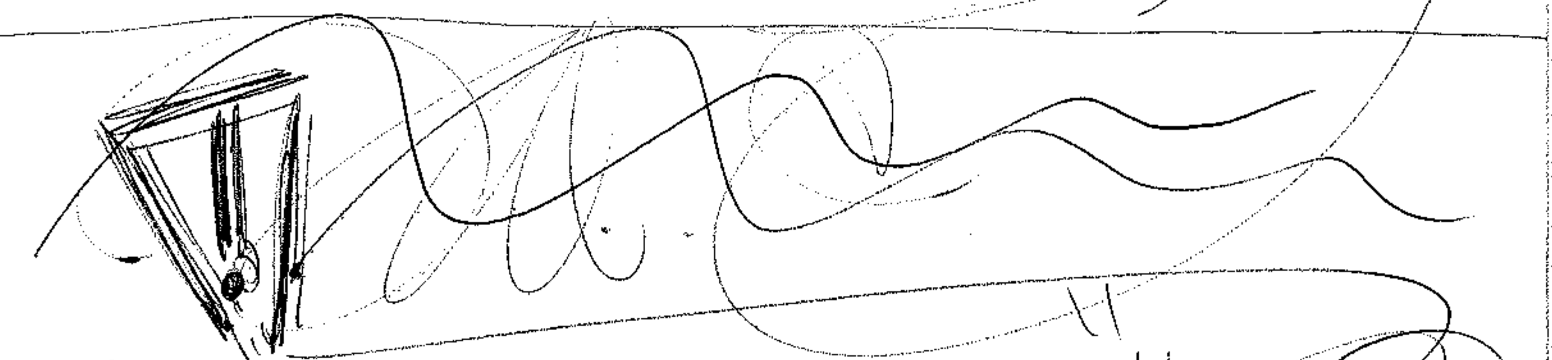
$L^{\theta_2}(\omega) = h^{\theta_1,\theta_2}(T(\omega)) L^{\theta_1}(\omega)$ se $h^{\theta_1,\theta_2}(x) \doteq e^{(\theta_2-\theta_1)x} \cdot \underbrace{e^{f(\theta_1) - f(\theta_2)}}_{>0}$, $x \in \mathbb{R}$, la quale è chiaramente monotona crescente (essendo $\theta_2 - \theta_1 > 0$). Per cioè, in effetti, un "buon" test per un problema del tipo $\mathcal{T} =$ «$\theta\le\theta_0$ contro $\theta>\theta_0$», e $\theta_0\in\Theta$ fissato, è $\phi_{c,\gamma} = 1\!\!1_{\{T>c\}} + \gamma 1\!\!1_{\{T=c\}}$ con $c\in\mathbb{R}$ e $\gamma\in[0,1]$ tale che $E^{P^{\theta_0}}[\phi_{c,\gamma}]$ sia "piccolo a piacere": infatti, un tale $\phi_{c,\gamma}$ è corretto e comunque ammissibile nel UPP di ogni altro test ϕ di taglia $\sup_{\theta\le\theta_0} E^{P^{\theta}}[\phi] \le E^{P^{\theta_0}}[\phi_{c,\gamma}]$. ✓

Ora non resterebbe che immergerci in casi concreti, e tuttavia prima vogliamo accennare al caso di problemi di test basati su un $\Theta_0 \subsetneq \Theta$ della forma $\Theta_0 = [\theta_1,\theta_2]\cap\Theta$, con $\theta_1\le\theta_2$ in Θ, o cioè $\mathcal{T} =$ «$H_0 : \theta\in[\theta_1,\theta_2]$ contro $H_1 : \theta\notin[\theta_1,\theta_2]$»: ogni test aleatorio ϕ per $\mathcal{T}$ è UN TEST BILATERALE (per il modello).

Teorema (TEST BILATERALI PER MODELLI ESPONENZIALI). Sia $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\},\mu)$ un modello stat., avente $\Theta\subseteq\mathbb{R}$, che sia esponenziale rispetto a μ, a Θ e ad una statistica reale $T:\Omega\to\mathbb{R}$, tale che inoltre $T(\mu)$ sia σ-finita e diffusa su $\mathcal{B}(\mathbb{R})$. Allora, considerati $\theta_1\le\theta_2$ in Θ ed il problema di test $\mathcal{T} =$ «$H_0:\theta\in[\theta_1,\theta_2]$ contro $H_1:\theta\notin[\theta_1,\theta_2]$», e posto $D_{1,2} = \{T\notin[c_1,c_2]\}$ per ogni $c_1\le c_2$ in $\mathbb{R}$, vale che: per ogni scelta $\alpha\in(0,1)$ (con $\alpha\ll 1$), esistono $c_1 = c_{1,\alpha}$ e $c_2 = c_{2,\alpha}$ in $\mathbb{R}$ con $c_1\le c_2$ tale che $D_\alpha = D_{1,2} = \boxed{\{T\notin[c_1,c_2]\}}$ soddisfi o $\begin{cases} P^{\theta_1}[D_\alpha]=\alpha \\ P^{\theta_2}[D_\alpha]=\alpha \end{cases}$ se $\boxed{\theta_1<\theta_2}$ o $\begin{cases} P^{\theta_1}[D_\alpha]=\alpha \\ \frac{dP^{\theta}[D_\alpha]}{d\theta}\Big|_{\theta=\theta_1}=0 \end{cases}$ se $\boxed{\theta_1=\theta_2}$, e tale D_α è una regione critica di taglia esattamente α, è corretta come test, ed è UPP fra tutti i test corretti di taglia $\leq \alpha$.

NOTA A riguardo del precedente teorema, osserviamo che $\theta \mapsto P^\theta[D_\alpha]$ è in $C^\infty(\Theta)$: infatti, $\forall \theta \in \Theta$, $P^\theta[D_\alpha] = \int_{D_\alpha} L^\theta \, d\mu =$

$$= e^{-f(\theta)} \int_{D_\alpha} e^{\theta T} d\mu \overset{(\text{def. } D_\alpha)}{=} e^{-f(\theta)} \int_{[c_1,c_2]^c} e^{\theta t} \, dT(\mu)(t) \; . \checkmark$$



DA ORA BASTA LA BELLA !! DAI

ES (Test unilaterale al $\leq$) $(X_1,\dots,X_n)$ campione di legge $\mathcal{E}(\theta)$, $\theta>0$ $\to$ $(\Omega, \mathcal{A}, \{P^\theta \mid \theta\in\Theta\}, \mu) =$

$$= \left((0,+\infty)^n, \mathcal{B}((0,+\infty)^n), \{\mu^{\theta,\otimes n} \mid \theta\in(0,+\infty)\}, \lambda^{\otimes n}\right),$$

λ di Leb. su $\mathcal{B}(0,+\infty)$, $\frac{d\mu^\theta}{d\lambda}(x) = \theta e^{-\theta x}$ $(\theta>0, x>0)$,

$X_i(x_1,\dots,x_n) = x_i$ ("$X = id$") $\to$ modello regolare con

$L^\theta(x_1,\dots,x_n) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$ $\to$ se $\boxed{T = -\sum_{i=1}^n X_i}$, (≤ 0)

allora $L^\theta(\omega) = \theta^n e^{\theta T(\omega)}$ $\to$ $\forall \theta_1 < \theta_2$ in Θ,

$$\frac{L^{\theta_2}(\omega)}{L^{\theta_1}(\omega)} = \left(\frac{\theta_2}{\theta_1}\right)^n e^{(\theta_2-\theta_1)T(\omega)} =: f^{\theta_1,\theta_2}(x)\Big|_{x=T(\omega)}$$

→ rispetto al verosimiglianza crescente: le buone regioni critiche si "basano" su T per ogni test unilatero! Ad es., per $\theta \le \boxed{1}_{\theta_0}$ contro $\theta > 1$, c l'intervallo $\{T \geq c\}$, $c \in \mathbb{R}$ ed anzi $c \le 0$, cioè $\{\sum_{i=1}^n X_i \le a\}$ con $a \ge 0$:

ma $X_1, \dots, X_n$ indip. $\sim \underset{(=\Gamma(1,\theta))}{\mathcal{E}(\theta)} \Rightarrow \sum_{i=1}^n X_i \sim \Gamma(n, \theta)$, $\Rightarrow$

$$P^1\left[\sum_{i=1}^n X_i \le a\right] = \int_0^a \frac{1}{(n-1)!} \cancel{\theta^n}^{\,\theta=1} x^{n-1} e^{-\cancel{\theta}^{\,\theta=1} x} \, dx =$$

$$= \frac{1}{(n-1)!} \int_0^a \underset{\substack{\uparrow \\ \text{der.}}}{x^{n-1}} \, \underset{\substack{\uparrow \\ \text{int.}}}{e^{-x}} \, dx = \frac{1}{(n-1)!} \Bigg[\underbrace{\left[x^{n-1} e^{-x}\right]_a^0}_{-a^{n-1} e^{-a}} + \int_0^a (n-1) x^{n-2} e^{-x} \, dx \Bigg] =$$

$$= e^{-a} \cdot \left(- \frac{a^{n-1}}{(n-1)!} \right) + \underbrace{\frac{1}{(n-2)!} \int_0^a x^{n-2} e^{-x}}_{(\text{"analogo"!})} =$$

$$= e^{-a} \left[- \frac{a^{n-1}}{(n-1)!} - \frac{a^{n-2}}{(n-2)!} - \dots - a \right] + \underbrace{\int_0^a e^{-x} \, dx}_{1 - e^{-a}} =$$

$$= e^{-a} \left[- \frac{a^{n-1}}{(n-1)!} - \frac{a^{n-2}}{(n-2)!} - \dots - a - 1 \right] + 1 . \checkmark$$

Es (TEST UNILATERALE con $\ge$) $(X_1, \dots, X_n)$ campione di legge μ^θ, $\theta > -1$, di densità rispetto alla mis. di Leb. 1-dim. $f^\theta(x) \doteq (\theta+1) x^\theta 1_{(0,1)}(x)$

$\Rightarrow \left(\Omega, \mathcal{A}, \{P^\theta | \theta \in \Theta\}, \mu\right) = \left((0,1)^n, \mathcal{B}((0,1)^n), \{\mu^{\theta, \otimes n} | \theta \in (-1, +\infty)\}, \nu^{\otimes n}\right)$,

$\frac{d\mu^\theta}{d\nu}(x) = (\theta+1)x^\theta \quad \forall \theta > -1, \forall x \in (0,1)$, "$X = id$" (X_i le funzioni ξ_i)

$\rightarrow$ Modello regolare con $\mathcal{L}^\theta(x_1, \dots, x_n) = (\theta+1)^n \left(\prod_{i=1}^n x_i\right)^\theta \rightarrow$ se

$\boxed{T \doteq \prod_{i=1}^n X_i}$, allora $\mathcal{L}^\theta(\omega) = (\theta+1)^n T(\omega)^\theta$ e

$\forall \theta_1 < \theta_2$ in Θ $\quad \frac{\mathcal{L}^{\theta_2}(\omega)}{\mathcal{L}^{\theta_1}(\omega)} = \left(\frac{\theta_2+1}{\theta_1+1}\right)^n T(\omega)^{\theta_2 - \theta_1} \quad (\forall \omega \in \Omega) =$

$= f^{\theta_1, \theta_2}(x) \Big|_{x = T(\omega)}$ $\rightarrow$ ~~...~~ rispetto a cui è monotona crescente,

$\Rightarrow$ per, ad es., « $\theta \geq \theta_0$ contro $\theta < \theta_0$ » ci interessano gli

eventi $\{T < C\}$, $C \geq 0$: Come nell'Es. prec.,

dobbiamo calcolare per esercizio $P^{\theta}[T < C]$.

$\boxed{\text{OSS.}}$ $\forall i = 1, \dots, n$, X_i è a valori in $(0,1)$ e, $\forall A \in \mathcal{B}(0,1)$,

$P^\theta[X_i \in A] = \int_A (\theta+1)x^\theta dx \quad \forall \theta \in \Theta$, $\Rightarrow P^0[X_i \in A] = \nu(A)$

$A \cap (0,1)$

Allora $\boxed{-\log(X_i) \overset{P^0}{\sim} \mathcal{E}(1)}$, o cioè $\forall a \geq 0$

$P^0[-\log(X_i) \leq a] = 1 - e^{-a}$, perché appunto

$P^0[-\log(X_i) \leq a] \overset{(\text{ovvio})}{=} P^0[X_i \geq e^{-a}] = 1 - P^0[X_i < e^{-a}] =$

$= 1 - e^{-a}$ $\quad$ $\checkmark$ $\quad$ Allora possiamo

$P^{\theta}[T < C] = P^{\theta}\left[\sum_{i=1}^{n}(-\log X_i) > -\log(C)\right]$ e $\sim$, ✓, per analogia

$\left(= \prod_{i=1}^{n} X_i\right)$

$\sum_{i=1}^{n}(-\log X_i) \sim \Gamma(n,1)$ con densità $\frac{1}{(n-1)!}x^{n-1}e^{-x}$, $x>0$, ecc. ✓

ES. (TEST "VARI" NEL CASO GAUSSIANO) $X=(X_1,\dots,X_n)$ campione di legge $N(m,\sigma^2)$,

$m\in\mathbb{R}$ e $\sigma^2>0$, con σ^2 NOTA (per il momento) $\rightarrow$

$(\Omega,\mathcal{A},\{P^\theta \mid \theta\in\Theta\},\mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{(\mu^m)^{\otimes n} \mid \theta = m\in\mathbb{R}\}, \nu^{\otimes n})$, con

ν di Leb. 1-dim., $\frac{d\mu^m}{d\nu}(x) = (2\pi\sigma^2)^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$ $(x\in\mathbb{R})$,

$X = id$ (X_i proiezione i-esima) $\rightarrow$ modello regolare con

$$L^m(x_1,\dots,x_n) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\Big(-\frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}\overbrace{(x_i-m)^2}^{(x_i^2+m^2-2x_im)}}_{\left(\sum_{i=1}^{n}x_i^2 + nm^2 - 2m\sum_{i=1}^{n}x_i\right)}\Big) \rightarrow$$

$\forall\, m_1 < m_2$ in $\mathbb{R}$, $\frac{L^{m_2}(x_1,\dots,x_n)}{L^{m_1}(x_1,\dots,x_n)} = \exp\left(\frac{m_2-m_1}{\sigma^2}\sum_{i=1}^{n}x_i\right)\text{cost}^{\text{nuova}}$

$\rightarrow$ rapporto di verosimiglianza crescente rispetto a $\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n}X_i}$ (ovviamente).

$\boxed{\text{Problema di test } m \le m_0 \text{ contro } m > m_0}$: regione critica delle forme $\{\overline{X} > C\}$, $C\in\mathbb{R}$.

per cui $\frac{d}{d\mu} P^{\mu}[\overline{X} \notin [c_1, c_2]]\Big|_{\mu=\mu_0} = 0$ se, e solo se,

$$\varphi\left(\sqrt{n}\,\frac{c_1-\mu_0}{\sigma}\right) = \varphi\left(\sqrt{n}\,\frac{c_2-\mu_0}{\sigma}\right), \quad \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

cioè $\cancel{\sqrt{n}}\,\frac{c_1-\mu_0}{\cancel{\sigma}} = \pm \cancel{\sqrt{n}}\,\frac{c_2-\mu_0}{\cancel{\sigma}}$ $\Rightarrow$ $\boxed{c_1 = c_2}$ oppure $\boxed{c_1 + c_2 = 2\mu_0}$

$\longrightarrow$ $c_1 = c_2$ NO, perché $\alpha < 1$ mentre $P^{\mu_0}[\overline{X} \neq c_1] = 1$.

Allora abbiamo $\boxed{\begin{cases} c_1 = \mu_0 - c \\ c_2 = \mu_0 + c \end{cases}, \quad c > 0} \longrightarrow$

$\longrightarrow$ $P^{\mu_0}[\overline{X} \notin [\mu_0 - c, \mu_0 + c]] = \alpha \iff$ (come prima)

$$P^{\mu_0}[\underbrace{\overline{X} < \mu_0 - c}_{\left(\sqrt{n}\frac{\overline{X}-\mu_0}{\sigma} < -c\frac{\sqrt{n}}{\sigma}\right)}] + 1 - P^{\mu_0}[\underbrace{\overline{X} < \mu_0 + c}_{\left(\sqrt{n}\frac{\overline{X}-\mu_0}{\sigma} < \frac{\sqrt{n}}{\sigma}c\right)}] = \alpha \iff$$

$$\iff \underbrace{1 - \Phi\left(\frac{\sqrt{n}}{\sigma}c\right) + \Phi\left(-c\frac{\sqrt{n}}{\sigma}\right)}_{2\left(1 - \Phi\left(\frac{\sqrt{n}}{\sigma}c\right)\right)} = \alpha \iff$$

$$\iff 1 - \frac{\alpha}{2} = \Phi\left(\frac{\sqrt{n}}{\sigma}c\right) \iff \frac{\sqrt{n}}{\sigma}c = q_{1-\frac{\alpha}{2}} \iff$$

$$\iff \boxed{c = \frac{\sigma}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}}$$

☺ ✓

→ Proviamo a calcolare $P^{\mu_0}[\overline{X} > c]$: è =

$$= 1 - P^{\mu_0}[\overline{X} \le c] = 1 - P^{\mu_0}\left[\underbrace{\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}}_{\left(\overset{P^{\mu_0}}{\sim} N(0,1)\right)} < \sqrt{n}\,\frac{c-\mu_0}{\sigma}\right] =$$

$\left(\overline{X} \overset{P^{\mu_0}}{\sim} N(\mu_0, \frac{\sigma^2}{n})\right)$

$$= 1 - \Phi\left(\sqrt{n}\cdot\frac{c-\mu_0}{\sigma}\right).$$

✓ Allora, $\forall \alpha \in (0,1)$ (con $\alpha \ll 1$), $P^{\mu_0}[\overline{X} > c] = \alpha \iff$

$$\iff 1-\alpha = \Phi\left(\sqrt{n}\cdot\frac{c-\mu_0}{\sigma}\right) \iff \sqrt{n}\,\frac{c-\mu_0}{\sigma} = q_{1-\alpha} \iff$$

$$\iff \boxed{c = \frac{\sigma}{\sqrt{n}}\, q_{1-\alpha} + \mu_0}$$

→ basta la regione critica! ($\phi_{c,\gamma}$ "con $\gamma = 0$"). ✓

$\boxed{\text{Problema del test } \mu = \mu_0 \text{ contro } \mu \neq \mu_0 :}$ Vogliamo trovare $c_1 \le c_2 \in \mathbb{R}$ tali che

$$\begin{cases} P^{\mu_0}[\overline{X} \notin [c_1, c_2]] = \alpha \\ \left.\dfrac{d}{d\mu} P^{\mu}[\overline{X} \notin [c_1,c_2]]\right|_{\mu=\mu_0} = 0 \end{cases}$$

Ma, $\forall \mu \in \mathbb{R}$,

$$P^{\mu}[\overline{X} \notin [c_1,c_2]] = P^{\mu}[\overline{X} \le c_1] + 1 - P^{\mu}[\overline{X} < c_2] =$$

$$= \Phi\left(\sqrt{n}\,\frac{c_1-\mu}{\sigma}\right) + 1 - \Phi\left(\sqrt{n}\,\frac{c_2-\mu}{\sigma}\right),$$ ed inoltre

$$\frac{d}{d\mu} P^{\mu}[\overline{X} \notin [c_1,c_2]] = -\frac{\sqrt{n}}{\sigma}\,\underset{\text{densità } N(0,1)}{\varphi}\left(\sqrt{n}\,\frac{c_1-\mu}{\sigma}\right) + \frac{\sqrt{n}}{\sigma}\,\varphi\left(\sqrt{n}\,\frac{c_2-\mu}{\sigma}\right),$$

(?) Per $m \in \mathbb{R}$ NOTA e $\vartheta = \sigma^2 > 0$ parametro? Vediamo:

usando $\mathcal{L}^\vartheta = \mathcal{L}^{\sigma^2} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - m)^2\right)$,

$\left(= \sum_{i=1}^{n} X_i^2 + n m^2 - 2m \sum_{i=1}^{n} X_i \right)$

(quasi) esponenziale rispetto a tale $\sigma^2 \mapsto -\frac{1}{2\sigma^2}$ e tale statistica

$\rightarrow \forall \sigma_1^2 < \sigma_2^2$, $(\vartheta_1 < \vartheta_2)$, $\frac{\mathcal{L}^{\vartheta_2}}{\mathcal{L}^{\vartheta_1}} = \frac{\mathcal{L}^{\sigma_2^2}}{\mathcal{L}^{\sigma_1^2}} = \text{cost.}^{\sigma_1^2, \sigma_2^2} \exp\left(\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right) \sum_{i=1}^{n} (X_i - m)^2\right)$

> 0 !

$\rightarrow$ rapporto a verosimiglianza crescente rispetto alla statistica

$$\boxed{T \doteq \sum_{i=1}^{n} (X_i - m)^2}$$

(. $\boxed{\text{NOTA}}$ $\forall i = 1, \dots, n$, $X_i \overset{P^{\sigma^2}}{\sim} N(m, \sigma^2)$ $\forall \sigma^2 > 0$ ($m \in \mathbb{R}$ era già pre-scelta), dunque ...

$\frac{X_i - m}{\sigma} \sim N(0,1)$, e allora $T = \sum_{i=1}^{n} (X_i - m)^2 = \sigma^2 \sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right)^2 \sim$

$\sim \sigma^2 \chi^2(n)$ ($X_1, \dots, X_n$ sono indip.), $\forall \sigma^2 > 0$.)

$\boxed{\text{Problema di test unilaterale } \text{"}\sigma^2 \leq \sigma_0^2 \text{ contro } \sigma^2 > \sigma_0^2\text{"}}$:

("ϑ_0")

e cerchiamo le regioni del tipo $\{T > C\}$, $C \in \mathbb{R}_+$.

Calcoliamo $P^{\sigma_0^2}[T > C]$: è $= P^{\sigma_0^2}\left[\sum_{i=1}^{n} (X_i - m)^2 > C\right] =$

$= P^{\sigma_0^2}\left[\sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma_0}\right)^2 > \frac{C}{\sigma_0^2}\right] = 1 - P^{\sigma_0^2}\left[\sum_{i=1}^{n} \frac{(X_i - m)^2}{\sigma_0^2} \leq \frac{C}{\sigma_0^2}\right]$,

$(\sim \chi^2(n))$

$= \alpha \iff \frac{C}{\sigma_0^2} = \chi^2_{1-\alpha, n}$, cioè $\boxed{C = \sigma_0^2 \chi^2_{1-\alpha, n}}$

$(\alpha \in (0,1), \alpha \ll 1)$

$\to$ Viene più tutto bene con la regione anche $\{T>C\}$ (comunque, così test $\varphi_{C,\gamma} \doteq 1_{\{T>C\}} + \gamma 1_{\{T=C\}}$ con $\gamma=0$). ✓

$\boxed{\text{Problema di test } \sigma^2=\sigma_0^2 \text{ contro } \sigma^2\neq\sigma_0^2}$: fissato $\alpha\in(0,1)$, $\alpha\ll1$, dovremmo ~~...~~ trovare $C_1\leq C_2$ in $\mathbb{R}$ tali che

$$\begin{cases} P^{\sigma_0^2}\left[\sum_{i=1}^{n}(X_i-m)^2 \notin [C_1,C_2]\right]=\alpha \\ \left.\dfrac{d P^{\sigma^2}\left[\sum_{i=1}^{n}(X_i-m)^2\notin[C_1,C_2]\right]}{d\sigma^2}\right|_{\sigma^2=\sigma_0^2}=0 \end{cases}$$

$\to$ (NOTA Ricorda che $\chi^2(n)=\Gamma(\frac{n}{2},\frac{1}{2})$!)

$\to$ Molto complicato, spesso insolubile .....

("lo vogliamo vedere")

$\to$ OSS Se il modello stat. $(\Omega,\mathcal{F},\{P^\theta|\theta\in\Theta\},\mu)$ (dominato) "che" come insieme dei parametri Θ un prodotto cartesiano

$$\boxed{\underset{\theta}{\Theta}=\Lambda_\lambda\times M_m}$$

dove solo Λ_λ ci interessano come insieme dei "parametri realmente ~~...~~ interessanti" allora i parametri $m\in M$ sono PARAMETRI FANTASMA (comunemente chiamati "parametri di disturbo"!). In queste situazioni, un problema di test dovrà basarsi necessariamente su un $\Theta_0\subsetneq\Theta$ della forma $\boxed{\Theta_0=\Lambda_0\times M}$

$(\Rightarrow \Theta_1\doteq\Theta\setminus\Theta_0=(\Lambda\setminus\Lambda_0)\times M)$.

(Nel complesso, nell'ES. per oltre un campione gaussiano ↗ media sconosciuta, ovvero nota ↘ media nota, ovvero sconosciuta

nessun "vero" Vir presenza di un parametro di disturbo. ✓

→ Anche Mettendo a confronto i due casi per un campione gaussiano visti nel precedente ES., l'idea per risolvere problemi di test unilaterale/bilaterale in presenza di un parametro di disturbo m è che conviene (per i calcoli) trovare una statistica T che NON dipende dal parametro di disturbo m stesso. ✓

(MA vediamo solo 2 esempi "importanti")

ES (1) TEST VARIANZA

$X = (X_1, \dots, X_n)$ campione di legge $N(m, \sigma^2)$, m e σ^2 entrambi NON NOTI, con lo scopo di fare un test sulla sola $\sigma^2 > 0$. Dunque è $L^{\theta} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - m)^2\right)$...

...... Consideriamo $\boxed{T \doteq \sum_{i=1}^{n}(X_i - \overline{X})^2}$, statistica non-negativa che, per COCHRAN, ha $T \overset{P^{m,\sigma^2}}{\sim} \sigma^2 \underbrace{\chi^2(n-1)}_{}$, e allora mettiamoci nel modello stat. immagine di $\Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$

$(\Omega, \mathcal{F}, \{P^{\theta} | \theta \in \Theta\}, \mu)$ di X mediante T, cioè in

$((0,+\infty), \mathcal{B}((0,+\infty)), \{T(P^{\theta}) | \theta \in \Theta\}, \cancel{T(\mu)})$ dove la verosimiglianza è

$$f^{\sigma^2}(x) \doteq \frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{1}{2}\right)^{\frac{n-1}{2}} \sigma^{-(n-1)} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2\sigma^2}}, \quad x > 0 \quad ✓$$

[Sappiamo infatti che $\frac{T}{\sigma^2} \overset{(P^{\sigma^2})}{\sim} \chi^2(n-1) = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$, e allora $\forall a>0$:

$$P[T \le a] = P\left[\frac{T}{\sigma^2} \le \frac{a}{\sigma^2}\right] =$$

$$= \int_0^{a/\sigma^2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n-1}{2}} y^{\frac{n-1}{2}-1} e^{-\frac{y}{2}}\, dy = $$

$\left(y = \frac{x}{\sigma^2} \right.$

$dy = \frac{1}{\sigma^2} dx$

$\left. [0, \frac{a}{\sigma^2}] \mapsto [0,a] \right)$

$$= \int_0^{a} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n-1}{2}} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2\sigma^2}} \underbrace{(\sigma^2)^{-\left(\frac{n-1}{2}-1\right)} \frac{1}{\sigma^2}}_{\left((\sigma^2)^{-\frac{n-1}{2}}\right)}\, dx \;. \checkmark]$$

Allora, $\forall \sigma_1^2 < \sigma_2^2$ $(\checkmark)$ $(e\ \forall x>0)$, $\frac{f^{\sigma_2^2}(x)}{f^{\sigma_1^2}(x)} = \text{cost.}^{\sigma_1^2,\sigma_2^2} \exp\Big(\underbrace{\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right)}_{>0} \overset{\text{"}T\text{"}}{x}\Big)$

$\rightsquigarrow$ *monotonia crescente!* Allora, per

$\boxed{\sigma^2 \le \sigma_0^2 \text{ cerchiamo } >}$ dobbiamo risolvere $\boxed{P^{\sigma_0^2}[T > C] = \alpha}$

(per $\alpha \in (0,1)$, $\alpha \ll 1$, fissato, e $C \ge 0$ incognita):

$$P^{\sigma_0^2}\Big[\frac{1}{\sigma_0^2}\underbrace{\sum_{i=1}^n (X_i - \overline{X})^2}_{(=T)} > \frac{C}{\sigma_0^2}\Big] = 1 - P^{\sigma_0^2}\Big[\underbrace{\frac{1}{\sigma_0^2} T}_{(\sim \chi^2(n-1))} \le \frac{C}{\sigma_0^2}\Big] =$$

$$= \alpha \iff \boxed{C = \sigma_0^2 \chi^2_{1-\alpha, n-1}} \;. \checkmark$$

ES. 2 (TEST DI STUDENT) Cominciamo ricordando che una v.a.

TEST MEDIA

$Y \sim t(n)$ ($\leftarrow$ legge di Student a n gradi di libertà) se esistono $Z \sim N(0,1)$ e $X \sim \chi^2(n)$
($n \geq 1$ opportuno)

tali che $Y = \sqrt{n} \dfrac{Z}{\sqrt{X}}$ (q.c.), X, Z INDIPENDENTI.

→ $\boxed{\text{NOTA}}$ Date $X_1, \dots, X_n$ indipendenti $N(\mu, \sigma^2)$ ($n \geq 2$), e posto $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ e $S := \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$, allora per COCHRAN $\sqrt{n} \dfrac{\overline{X} - \mu}{S} \sim t(n-1)$.

$\boxed{\text{Def.}}$ Y ha la legge di Student, a n gradi di libertà, decentrata di $a \in \mathbb{R}$ se esistono $Z \sim N(a, 1)$ e $X \sim \chi^2(n)$ INDIPENDENTI tali che $Y \overset{q.c.}{=} \sqrt{n} \dfrac{Z}{\sqrt{X}}$.

→ Se $(X_1, \dots, X_n)$ è un campione di legge $N(\mu, \sigma^2)$, $n \geq 2$, allora $\boxed{T := \sqrt{n} \dfrac{\overline{X}}{S}}$ ha la legge di Student a $n-1$ gradi di libertà decentrata di $\dfrac{\sqrt{n}}{\sigma} \mu$.

Il fatto è che, (sempre per COCHRAN) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ e $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$ e sono indipendenti: basta allora osservare che

$\frac{\sqrt{n}}{\sigma}\overline{X} \sim N\left(\frac{\sqrt{n}}{\sigma}\mu , 1\right)$ e che

$$T \overset{(\text{def.})}{=} \sqrt{n}\,\frac{\overline{X}}{S} \overset{(\text{imm.})}{=} \underbrace{\sqrt{n-1}}_{\text{“}n-1\text{”}}\;\frac{\overbrace{\left(\frac{\sqrt{n}}{\sigma}\overline{X}\right)}^{N\left(\frac{\sqrt{n}}{\sigma}\mu,\,1\right)}}{\boxed{\frac{\sqrt{n-1}}{\sigma}S}} \quad \text{(indip.)} \quad \text{“}\sqrt{\chi^2(n-1)}\text{”} \quad . \checkmark \Big]$$

Come prima, se $X=(X_1,\dots,X_n)$ campione di legge $N(\mu,\sigma^2)$

e se $\boxed{T \doteq \sqrt{n}\,\frac{\overline{X}}{S}}$, e DIAMO PER BUONO il fatto

che la statistica reale T trasforma il modello stat. di

campione in un (altro) modello stat. che abbia rapporto di

verosimiglianza crescente in $\sqrt{n}\,\frac{\mu}{\sigma}$ (come parametro).

Allora l'idea è quella di fare un test su μ su quest'ultimo modello anche su quello di partenza, perché quest'ultimo è a v.r. crescente rispetto ad un parametro nel quale comunque compare μ!

(Come nell'es. prec., ma con μ invece che σ.)

$\boxed{\text{Test. di } \frac{\mu}{\sigma} \le 0 \;(\mu_0=0) \text{ contro } > 0 \text{ »}}$ $\longrightarrow$ $\boxed{P^0[T > \underset{\text{(GR incognita)}}{c}] = \underset{((0,1),\ \text{dato})}{\alpha}}$:

sotto P^0, $T = \sqrt{n}\,\frac{\overline{X}}{S} \sim t(n-1)$, per cui come sempre

$\alpha = 1 - P^0[T \le c] \Leftrightarrow c = t_{1-\alpha,\,n-1}$. ✓

$\boxed{\text{Test di } \frac{\mu}{\cancel{\sigma}} \le \frac{\mu_0}{\cancel{\sigma}} \text{ contro } >\text{"}}$ $\to P^{\mu,\sigma}[T > c] = \alpha \ldots$ (Student decentrata di $\sqrt{n}\frac{\mu_0}{\sigma}$)

Proviamo a ricondurci al caso precedente: se, $\forall i = 1, \ldots, n$,

$Y_i \doteq X_i - \mu_0$, allora $Y = (Y_1, \ldots, Y_n)$ è un campione di legge $N(\boxed{\mu - \mu_0}, \sigma^2)$, sotto P^{μ,σ^2}, ed

il test è di $\boxed{\mu - \mu_0 \le 0}$ contro $>$". Inoltre

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{X} - \mu_0 \text{, e (quindi)}$$

$$S_Y^2 \doteq \frac{1}{n-1}\sum_{i=1}^{n}(\underbrace{Y_i}_{X_i-\mu_0} - \underbrace{\overline{Y}}_{\overline{X}-\mu_0}) = S^2$$

per il test,

Consideriamo $T \doteq \sqrt{n}\,\frac{\overline{Y}}{S_Y} = \boxed{\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}}$, e allora

anziché $\{\sqrt{n}\frac{\overline{X}}{S} > c\}$ abbiamo $\{\underbrace{\sqrt{n}\frac{\overline{X}-\mu_0}{S}}_{(\sim t(n-1))} > c\}$, CGR ✓

TEST DEL RAPPORTO DI VEROSIMIGLIANZA

Per $(\Omega, \mathcal{A}, \{P^\theta \mid \theta \in \Theta\}, \mu)$ modello stat. dominato, e per il problema di test "generale" di "$\theta \in \Theta_0$ contro $\theta \notin \Theta_0$", facciamo un test aleatorio $\phi(\omega)$ del rapporto di verosimiglianza se "bruciamo" $\phi(\omega)$

nelle regioni della forma $\left\{ \sup_{\vartheta_0 \in \Theta_0} L^{\vartheta_0} < C \sup_{\vartheta_1 \in \Theta_1} L^{\vartheta_1} \right\}$, $C \in \mathbb{R}$ $(C>0)$

$(\Theta_1 = \Theta \setminus \Theta_0)$

(supponendo esistano ϑ_0, ϑ_1 q.e.n.)

(L'idea è tipo la seguente: ogni $\forall \omega \in \Omega$ tale che

$\sup_{\vartheta_1 \in \Theta_1} L^{\vartheta_1}(\omega) >> \sup_{\vartheta_0 \in \Theta_0} L^{\vartheta_0}(\omega)$, è tale che almeno

una $L^{\vartheta}(\omega)$ con $\vartheta \notin \Theta_0$ conta molto di più di tutte le (sup P^{ϑ})

$L^{\vartheta_0}(\omega)$, $\vartheta_0 \in \Theta_0$).

e se $\vartheta \mapsto L^{\vartheta}(\omega)$ è continua ($\forall \omega \in \Omega$)

Se $\Theta_0 = \{\vartheta_0\}$, cioè il problema è $\vartheta = \vartheta_0$ contro $\vartheta \neq \vartheta_0$,

allora la regione diventa $\boxed{\left\{ L^{\vartheta_0} < C \sup_{\vartheta \in \Theta} L^{\vartheta} \right\}}$: se

esiste uno stimatore $\hat{\vartheta}(\omega)$ (di ϑ) di massima verosimiglianza,

allora $\left\{ L^{\vartheta_0}_{(\cdot)} < C\, L^{\hat{\vartheta}(\cdot)}_{(\cdot)} \right\}$.

Vogliamo vedere due esempi.

ES. (a) Torniamo ad un campione $X=(X_1,\dots,X_n)$ di legge $N(m,\sigma^2)$ con entrambi m e σ^2 sconosciuti, e vogliamo fare un problema di test sulla media m ($\to$ TEST DI STUDENT), precisamente

« $m=0$ contro $m\neq 0$ » (NON calcolabile, perché qui/anche σ^2 è sconosciuta). Allora cerchiamo le regioni tipo

$$\boxed{\left\{ \sup_{\sigma^2>0} L^{(0,\sigma^2)} < C \sup_{\substack{m\in\mathbb{R},\\ \sigma^2>0}} L^{(m,\sigma^2)} \right\}}, \quad C>0, \text{ dove}$$

$$L^{(m,\sigma^2)}(x_1,\dots,x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-m)^2\right) =$$

$$= \exp\left(\underbrace{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{m}{\sigma^2}\sum_{i=1}^n x_i - \frac{n m^2}{2\sigma^2}} - \frac{n}{2}\cancel{\log(2\pi\sigma^2)}\right), \text{ e}$$

dove (ricordando) sono stimatori di massima verosimiglianza

$$\begin{cases} \xrightarrow[\sigma^2>0]{m=0} \dfrac{1}{n}\sum_{i=1}^n X_i^2 \\[2ex] \xrightarrow[\sigma^2>0]{m\in\mathbb{R},} \dfrac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2 \end{cases}, \text{ da cui}$$

$$\begin{cases} \sup_{\sigma^2>0} L^{(0,\sigma^2)} = L^{\left(0,\frac{1}{n}\sum_{i=1}^n X_i^2\right)} = (2\pi)^{-\frac{n}{2}}\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{-\frac{n}{2}} e^{-\frac{n}{2}} \\[2ex] \sup_{\substack{m\in\mathbb{R},\\ \sigma^2>0}} L^{(m,\sigma^2)} = L^{\left(\bar{X},\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2\right)} = (2\pi)^{-\frac{n}{2}}\left(\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2\right)^{-\frac{n}{2}} e^{-\frac{n}{2}} \end{cases}$$

$\Longrightarrow$ le regioni sono del tipo

$$\left\{\left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^{-\frac{n}{2}} < c\right\} = \left\{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n} x_i^2} < c^{\frac{2}{n}}\right\} =$$

$$\sum_{i=1}^{n}(x_i^2+\bar{x}^2-2x_i\bar{x})$$

$$\to n(\bar{x}^2-2\bar{x}^2) = -n\bar{x}^2$$

$$= \left\{1 + \frac{\boxed{n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n} x_i}}{\sum_{i=1}^{n} x_i^2} < c^{\frac{2}{n}}\right\} =$$

$$= \left\{1 - n\frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2} < c^{2/n}\right\} =$$

e questo, $\sum_{i=1}^{n} x_i =$
$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n\bar{x}^2$

$$\Rightarrow \left\{\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 + n\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^{-\frac{n}{2}} < c\right\} =$$

$$= \boxed{\left\{\left(1 + n\left(\frac{\bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right)^2\right)^{-\frac{n}{2}} < c\right\}} \, , =$$

$(a = a(c))$
(>0)

$$= \left\{\frac{|\bar{x}|}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} > a\right\}$$

: tutto questo per

avere ora che, finalmente,

$$\frac{\overline{X}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} \overset{(\text{inn.})}{=} \frac{1}{\sqrt{n(n-1)}}\sqrt{n}\underbrace{\frac{\overline{X}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}}}_{\left(=\sqrt{n}\frac{\overline{X}}{S}\right)} = \frac{1}{\sqrt{n(n-1)}}\sqrt{n}\frac{\overline{X}}{S}$$

(Cochran) $\Rightarrow$ è $\sim t(n-1)$ sotto $m=0$!)

quindi possiamo calcolare. ✓✓
(La legge di Student è nominata!)

ES. (b) Campione $(X_1,\dots,X_n)$ di legge μ^θ, $\theta>0$, avente densità rispetto alla Lebesgue λ $\boxed{\frac{d\mu^\theta}{d\lambda}(x) = e^{-(x-\theta)}\mathbb{1}_{[\theta,+\infty)}(x)}$,

$x\in\mathbb{R}$ (o $x>0$).

$\boxed{\text{Oss. che, in effetti, } \int_\theta^{+\infty} e^{-(x-\theta)}\,dx = \left[e^{-(x-\theta)}\right]_{x=+\infty}^{x=\theta} = 1.}$

$\Rightarrow (\Omega,\mathcal{B},\{P^\theta \mid \theta\in\Theta\},\mu) = ((0,+\infty)^n, \mathcal{B}((0,+\infty)^n), \{(\mu^\theta)^{\otimes n} \mid \theta\in(0,+\infty)\}, \lambda^n)$,
$X_i(x_1,\dots,x_n)=x_i$. Allora la verosimiglianza è, $\forall\theta>0$,

$$L^\theta(x_1,\dots,x_n) \underset{(\forall x_i>0)}{=} \exp\left(-\left(\sum_{i=1}^n x_i - n\theta\right)\right)\mathbb{1}_{[\theta,+\infty)}\left(\min_{i=1,\dots,n} x_i\right),$$

ossia $\boxed{L^\theta = e^{-\left(\sum_{i=1}^n X_i - n\theta\right)}\mathbb{1}_{[\theta,+\infty)}\left(\min_{i=1,\dots,n} X_i\right)}$, $\theta>0$.

("CASINOSA" ....)

$\boxed{\text{Problema di test « } \theta=1 \text{ contro } \theta\neq 1 \text{ »}}$: regioni

$$\left\{ L^1 < C \sup_{\theta\neq 1} L^\theta \right\}, \quad C>0 .$$

Quindi, $L^1 = e^{-\left(\sum_{i=1}^n X_i - n\right)} 1\!\!1_{[1,+\infty)}(\min_{i=1,..,n} X_i)$:

chi è invece $\sup_{\theta\neq 1} L^\theta$?? Viene (ovvio ...)

$$\sup_{\theta\neq 1} L^\theta = \sup_{\theta>0} L^\theta = \exp\left(n(\min_{i=1,..,n} X_i) - \sum_{i=1}^n X_i \right) \quad (>0),$$

ed il rapporto è $\dfrac{L^1}{\sup_{\theta\neq 1} L^\theta} = \begin{cases} 0 & \text{se } \min_{i=1,..,n} X_i < 1 \\ e^{n(1-\min_{i=1,..,n} X_i)} & \text{se } \min_i X_i \geq 1, \end{cases}$

per cui hanno senso i $C \in (0,1)$, e allora

$\forall C \in (0,1)$, le regioni sono

$$\{\min_i X_i < 1\} \cup \{ \min_i X_i \geq 1, \underbrace{e^{n(1-\min_i X_i)} < C}_{(1-\min_i X_i) < \frac{1}{n}\log C} \} =$$

$$= \{\min_i X_i < 1\} \cup \{ \min_i X_i > 1 + \tfrac{1}{n}\log(C^{-1}) \}, \quad (>1)$$

ed ora ci servirebbe la densità di $\min_i X_i$ (sotto ogni P^θ)

....

$\boxed{\text{EX}}$ Sotto P^ϑ, $\vartheta>0$, la densità di $\min_{i=1,\dots,n} X_i$ è

$$f^\vartheta(x) = n\, e^{-n(x-\vartheta)}\, \mathbb{1}_{[\vartheta,+\infty)}(x), \quad x>0.$$

$$\forall t\geq 0,\quad P^\vartheta\Big[\min_{i=1,\dots,n} X_i \geq t\Big] = P^\vartheta[X_1\geq t,\dots,X_n\geq t] \overset{\text{INDIP.+ISO.}}{=}$$

$$= \left(P^\vartheta[X_1\geq t]\right)^n. \quad \text{Ora, } P^\vartheta[X_1\geq t] =$$

$$= \int_t^{+\infty} e^{-(x-\vartheta)}\, \mathbb{1}_{[\vartheta,+\infty)}(x)\, dx$$

$$= \int_{t\vee\vartheta}^{+\infty} e^{-(x-\vartheta)}\, dx =$$

$$= \begin{cases} \displaystyle\int_t^{+\infty} e^{-(x-\vartheta)}\, dx \ \text{ se } t\geq\vartheta, & = \left[e^{-(x-\vartheta)}\right]_{+\infty}^{t} = e^{-(t-\vartheta)} \\ \displaystyle\int_\vartheta^{+\infty} e^{-(x-\vartheta)}\, dx \ \text{ se } t\leq\vartheta, & = \left[e^{-(x-\vartheta)}\right]_{+\infty}^{\vartheta} = 1 \end{cases},$$

da cui
$$P^\vartheta\Big[\min_{i=1,\dots,n} X_i \geq t\Big] = \begin{cases} 1 & \text{se } t\leq\vartheta \\ e^{-n(t-\vartheta)} & \text{se } t\geq\vartheta \end{cases}.$$

Ma, d'altra parte, anche
$$\int_t^{+\infty} f^\vartheta(x)\, dx = \int_t^{+\infty} n\, e^{-n(x-\vartheta)}\, \mathbb{1}_{[\vartheta,+\infty)}(x) =$$
$$= \int_{t\vee\vartheta}^{+\infty} n\, e^{-n(x-\vartheta)}\, dx = \left[e^{-n(x-\vartheta)}\right]_{x=+\infty}^{x=t\vee\vartheta} \text{ che è la stessa cosa.}$$

Cenni di ANALISI DELLA VARIANZA. (ANOVA)

Supponiamo di avere k campioni gaussiani, $k \geq 2$, fra loro indipendenti, ciascuno con le sue taglie, le sue medie e le sue varianze:

$$\mathbb{X}_1 = (X_{1,1}, \dots, X_{1,n_1}) \text{ campione di legge } N(m_1, \sigma_1^2)$$
$$\mathbb{X}_2 = (X_{2,1}, \dots, X_{2,n_2}) \text{ campione di legge } N(m_2, \sigma_2^2)$$
$$\vdots$$
$$\mathbb{X}_k = (X_{k,1}, \dots, X_{k,n_k}) \text{ campione di legge } N(m_k, \sigma_k^2)$$

INDIPENDENTI mutuamente

$\boxed{\text{NOTA}}$ (Ormai sappiamo bene come costruire un modello stat. nel quale tali oggetti esistono davvero!)

L'ANALISI DELLA VARIANZA si propone di risolvere problemi di test che riguardano i parametri dei diversi campioni $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_k$ al tempo stesso, come per esempio H_0: $m_1 = m_2 = \dots = m_k$ contro $\exists\, i \neq j$ ($i,j \in \{1,\dots,k\}$) tali che $m_i \neq m_j$.

Complicato ..... Vediamo due esempi classici (comunque non immediati).

ES. 1 (PROBLEMA DI BEHRENS – FISHER) Mettiamoci nella situazione precedente con $k=2$:

$$\begin{cases} X=(X_1,\dots,X_p) \text{ campione } N(m_1,\sigma_1^2) \\ Y=(Y_1,\dots,Y_q) \text{ campione } N(m_2,\sigma_2^2) \end{cases}, \; X,Y \text{ INDIPENDENTI}.$$

Problema di test a $m_1=m_2$ contro $m_1\neq m_2$ → Si suppone che $\boxed{\sigma_1^2=\sigma_2^2=:\sigma^2}$ (sempre sconosciuta).

Se $\overline{X}=\frac{1}{p}\sum_{i=1}^p X_i$ e $\overline{Y}=\frac{1}{q}\sum_{i=1}^q Y_i$, allora una elementare generalizzazione di Cochran ci direbbe che la v.a.

$$Z_{p,q} \doteq \frac{\sqrt{p+q-2}\,(\overline{X}-\overline{Y})}{\sqrt{\frac{1}{p}+\frac{1}{q}}\sqrt{\sum_{i=1}^p (X_i-\overline{X})^2+\sum_{i=1}^q (Y_i-\overline{Y})^2}}$$ che

legge di Student a $p+q-2$ gradi di libertà decentrata di $\frac{m_1-m_2}{\sigma\sqrt{\frac{1}{p}+\frac{1}{q}}}$. Abbiamo detto che tali leggi sono a rapporto di verosimiglianza crescente in m_1-m_2, per cui per il problema di test d'ipotesi nulla $m_1=m_2$ possiamo usare regioni del tipo $\{|Z_{p,q}|>c\}$, $c>0$, e osservare che sotto $m_1=m_2$ $Z_{p,q}\sim t(p+q-2)$ (anche simmetrica!). ✓

ES. 2 ("quello vero") Partiamo ora dal <u>TEST DI OMOGENEITÀ</u>, ma premettendo alcune cose...

→ Dati $n_1, n_2 \in \mathbb{N}_*$, <u>LA LEGGE DI FISHER-SNEDECOR</u> "$F(n_1,n_2)$" è la legge della v.a.r. $\dfrac{Z_1/n_1}{Z_2/n_2}$ dove

Z_1, Z_2 sono v.a.r. INDIPENDENTI con $Z_1 \sim \chi^2(n_1)$, $Z_2 \sim \chi^2(n_2)$.

ritorno l' "Esempio." Sia $X=(X_1,\dots,X_n)$ un vettore gaussiano di legge $N_n(m, \sigma^2 I_n)$ (è un campione ~~...~~),

dove supponiamo $\boxed{m \in E \text{ con } E < \mathbb{R}^n \text{ avente } \dim E =: k < n}$,

ossia con $1 \le k < n$, e fissiamo un altro sottospazio $\boxed{H < E \text{ con } 1 \le \dim H =: r < k}$. per fare il problema di test

$\boxed{\ll m \in H \text{ contro } m \in E \setminus H \gg}$. Vogliamo provare

ad usare il $\boxed{\text{teorema di Cochran}}$ ...

Posto $Y \doteq X - m$, Y è un vettore gaussiano $\boxed{\sim N_n(0, \sigma^2 I_n)}$ e vale

$\boxed{X = m + Y}$ (tutto OK) ; in particolare, $Y = (Y_1, \dots, Y_n)$

è un campione di legge $N(0, \sigma^2)$, e possiamo applicare Cochran

rispetto alla decomposizione ortogonale di $\mathbb{R}^n$ data da

$H \oplus$ ortogonale di H in E $\oplus$ ortogonale di E , alla quale

corrisponde la scrittura $Y_H + (Y_E - Y_H) + (Y - Y_E) = Y$ (

se ad esempio $Y_V \doteq$ la proiezione ortogonale di Y su V, dato

$V \leq \mathbb{R}^n$) : allora, per Cochran, in

particolare, $Y - Y_E$ e $Y_E - Y_H$ sono indipendenti

e di leggi
$$\begin{cases} \dfrac{\|Y - Y_E\|^2}{\sigma^2} \sim \chi^2(n-k) \\ \dfrac{\|Y_E - Y_H\|^2}{\sigma^2} \sim \chi^2(k-r) \end{cases} ,$$

per cui
(PER DEFINIZIONE)
$$\boxed{\frac{\|Y_E - Y_H\|^2 / (k-r)}{\|Y - Y_E\|^2 / (n-k)} \sim F(k-r, n-k)} :$$

Adesso, tornando al nostro $X = (X_1, \dots, X_n) = m + Y$, avendo che $m \in E$ si ha $m_E = m$, per cui

$$\begin{cases} X_E = Y_E + m_E = Y_E + m \\ X_H = Y_H + m_H \end{cases}$$ e allora

$$\begin{cases} X - X_E = (Y+m) - (Y_E + m) = Y - Y_E \\ X_E - X_H = (Y_E + m) - (Y_H + m_H) = (Y_E - Y_H) + (m - m_H), \end{cases}$$

dunque riassumendo

$$\boxed{\begin{cases} X - X_E = Y - Y_E \\ X_E - X_H = (Y_E - Y_H) + (m - m_H) \end{cases} \quad \text{(e restano quindi indipendenti)}}$$

pertanto, concludiamo osservando che ~~tanto~~ quanto "più" $m \neq m_H$ (cioè $m \notin H$), ossia ~~tanto~~ quanto più $\|m - m_H\|^2$ è grande, tanto più $\|X_E - X_H\|^2$ sarà grande, e allora potremmo prendere come regione critica

$$\left\{ \frac{\|X_E - X_H\|^2 / k-h}{\|X - X_E\|^2 / n-k} > C \right\}, \; C > 0, \text{ da determinarsi}$$

che ... hanno legge e rapporto di occorrenza/fisher crescente in $\|m - m_H\|^{(2)}$, ed avere così le

Concludete che, sotto $m = m_H$, $X_E - X_H = Y_E - Y_H$

e allora $\dfrac{\|X_E - X_H\|^2 / k-r}{\|X - X_E\|^2 / n-k} \sim F(k-r, n-k)$!! ☺

Vediamo finalmente il $\boxed{\text{test di omogeneità}}$: siano

$$\begin{cases} X_1 = (X_{1,1}, \ldots, X_{1,m_1}) \text{ campione } N(m_1, \sigma^2) \\ \vdots \\ X_k = (X_{k,1}, \ldots, X_{k,m_k}) \text{ campione } N(m_k, \sigma^2) \end{cases}, \quad k \geq 2, \quad (m_i \geq 2)$$

$X_1, \ldots, X_k$ indipendenti, e poniamo il problema del test di ipotesi nulla $\boxed{H_0: m_1 = m_2 = \ldots = m_k}$. (La comune varianza σ^2 è sconosciuta! (Parametro di disturbo).)

Sia $X \doteq (X_1, \ldots, X_k)'' = (X_{1,1}, \ldots, X_{1,m_1}, X_{2,1}, \ldots, X_{2,m_2}, \ldots$

$\ldots, X_{k,1}, \ldots, X_{k,m_k})$, per cui X ha componenti indipendenti di certe medie ("medie m_i a gruppi di n_i elementi") e di varianza σ^2, e ci stiamo dunque chiedendo se X è un campione di legge $N(m, \sigma^2)$, $m \doteq m_1 = m_2 = \ldots = m_k$.

Proviamo ad usare quanto detto finora per

$$X \, : \begin{cases} \overline{X}_i = \dfrac{1}{m_i} \displaystyle\sum_{j=1}^{m_i} X_{i,j} \quad , \; i=1,\dots,k & \text{(media i-esimo campione)} \\[2ex] \overline{X} = \dfrac{1}{m_1+\dots+m_k} \displaystyle\sum_{\substack{i=1,\dots,k \\ j=1,\dots,m_i}} X_{i,j} & \text{(media "globale")} \end{cases}$$

La media del vettore X è, quindi,

$$(\underbrace{\mu_1,\dots,\mu_1}_{m_1 \text{ volte}}, \underbrace{\mu_2,\dots,\mu_2}_{m_2 \text{ volte}}, \dots, \underbrace{\mu_k,\dots,\mu_k}_{m_k \text{ volte}})^t =: m \quad ,$$

per cui $m \in \mathbb{R}^n$, $n \doteq m_1+m_2+\dots+m_k$, che in realtà' in ($n>k$)

$E < \mathbb{R}^n$, $E \doteq \mathrm{Span}_{\mathbb{R}} \langle \eta_1, \dots, \eta_k \rangle$ dove (per prendenti innanzi)

$$\begin{cases} \eta_1 \doteq \dfrac{1}{\sqrt{m_1}}(\underbrace{1,\dots,1}_{m_1 \text{ volte}}, \underbrace{0,\dots,0}_{m_2}, \dots, \underbrace{0,\dots,0}_{m_k}) \\[2ex] \eta_2 \doteq \dfrac{1}{\sqrt{m_2}}(\underbrace{0,\dots,0}_{m_1}, \underbrace{1,\dots,1}_{m_2 \text{ volte}}, \dots, \underbrace{0,\dots,0}_{m_k}) & \text{(base ortonormale!)} \\ \vdots \\ \eta_k \doteq \dfrac{1}{\sqrt{m_k}}(0,\dots,0,0,\dots,0,\dots,\underbrace{1,\dots,1}_{m_k\text{-volte}}) \end{cases} \quad , \text{perché}$$

essendo $m = \sqrt{m_1}\,\mu_1\eta_1 + \sqrt{m_2}\,\mu_2\eta_2 + \dots + \sqrt{m_k}\,\mu_k\eta_k$ e

H$_0$ è che se $m \in H \doteq \mathrm{Span}_{\mathbb{R}} \langle \overbrace{\tfrac{1}{\sqrt{n}}(1,\dots,1)}^{=:\eta} \rangle$:

riassumendo, $\mu \in E < \mathbb{R}^n$ con $\dim E = k < m$ e ci chiediamo se sia $\mu \in H < E$ con $\dim H = 1 < k$. L'$\boxed{\text{IDEA}}$, allora, resta quella di basare il test su regioni della forma

$$\left\{ \underbrace{\frac{\|X_E - X_H\|^2 / _{k-1}}{\|X - X_E\|^2 / _{m-k}}}_{} > c \right\}, \quad c > 0 \,. \checkmark$$

$\left(\sim F(k-1, m-k) \text{ sotto } H_0 \right)$ Concludiamo l'esempio esplicitando le v.a. ≥ 0 $\|X - X_E\|^2$ e $\|X_E - X_H\|^2$:

$$\boxed{\boxed{X_E} = \overline{X}_1 \sqrt{m_1}\, \eta_1 + \dots + \overline{X}_k \sqrt{m_k}\, \eta_k .} \checkmark$$

$\Big[X_E = \sum_{i=1}^{k} \langle X; \eta_i \rangle \eta_i = \sum_{i=1}^{k} \frac{\langle X; \eta_i \rangle}{\sqrt{m_i}} \sqrt{m_i}\, \eta_i$, ed ora

$$\frac{1}{\sqrt{m_i}} \langle X; \eta_i \rangle \overset{(\text{imm.})}{=} \frac{1}{m_i} \sum_{j=1}^{m_i} X_{i,j} \overset{(\text{def.})}{=} \overline{X}_i, \quad \forall i = 1, \dots, k. \checkmark \Big]$$

$$\boxed{\boxed{X_H} = \overline{X} \sqrt{m}\, \eta} \quad . \checkmark \quad \Big[\text{Ovvio}: X_H = \cancel{X} \langle X; \eta \rangle \eta =$$

$$= \frac{\langle X; \eta \rangle}{\sqrt{m}} \sqrt{m}\, \eta, \text{ e } \frac{1}{\sqrt{m}} \langle X; \eta \rangle \overset{\text{ovvio}}{=} \overline{X} . \checkmark \Big]$$

Scritto in modo ancora più esplicito,

$$\begin{cases} X_E = (\underbrace{\overline{X}_1, \dots, \overline{X}_1}_{m_1 \text{ volte}}, \underbrace{\overline{X}_2, \dots, \overline{X}_2}_{m_2 \text{ volte}}, \dots\dots, \underbrace{\overline{X}_k, \dots, \overline{X}_k}_{m_k \text{ volte}}) \\ X_H = (\underbrace{\overline{X}, \overline{X}, \dots\dots, \overline{X}}_{m \text{ volte}}) \end{cases} (\in \mathbb{R}^m)$$

e allora

$$\begin{cases} X_E - X_H = (\overbrace{\overline{X}_1 - \overline{X}, \dots, \overline{X}_1 - \overline{X}}^{m_1 \text{ volte}}, \overbrace{\overline{X}_2 - \overline{X}, \dots, \overline{X}_2 - \overline{X}}^{m_2 \text{ volte}}, \dots \\ \qquad \dots, \underbrace{\overline{X}_k - \overline{X}, \dots, \overline{X}_k - \overline{X}}_{m_k \text{ volte}}) \\ X - X_E = (\overbrace{X_{1,1} - \overline{X}_1, \dots, X_{1,m_1} - \overline{X}_1}^{m_1 \text{ volte}}, \overbrace{X_{2,1} - \overline{X}_2, \dots, X_{2,m_2} - \overline{X}_2}^{m_2 \text{ volte}}, \\ \qquad \dots\dots, \underbrace{X_{k,1} - \overline{X}_k, \dots, X_{k,m_k} - \overline{X}_k}_{m_k \text{ volte}}) \end{cases}$$

$$\Rightarrow \boxed{\begin{aligned} \|X - X_E\|^2 &= \sum_{i=1}^{k} \sum_{j=1}^{m_i} (X_{i,j} - \overline{X}_i)^2 \\ \|X_E - X_H\|^2 &= \sum_{i=1}^{k} m_i (\overline{X}_i - \overline{X})^2 \end{aligned}}$$

VARIAZIONE INTERNA (somma variazioni di ogni "gruppo" rispetto alla media e. del gruppo stesso)

VARIAZIONE ESTERNA (somma variazioni delle medie c. di ogni gruppo rispetto alla media c. di tutto il campione)

da cui insieme c'interessa "quasi" il quoziente

$$\frac{\sum_{i=1}^{k} m_i (\overline{X}_i - \overline{X})^2 \Big/ k-1}{\sum_{i=1}^{k} \sum_{j=1}^{m_i} (X_{i,j} - \overline{X}_i)^2 \Big/ n-k} \quad \text{def} \quad \frac{\text{varianza esterna}}{\text{varianza interna}} \quad \checkmark$$

$\boxed{\text{NOTA}}$ $X - X_H = \underset{(\in E^{\perp})}{(X - X_E)} \overset{\text{(ortogonali)}}{+} \underset{(\in E)}{(X_E - X_H)} \Rightarrow$

$\|X - X_H\|^2 \overset{\text{(Pitagora)}}{=} \|X - X_E\|^2 + \|X_E - X_H\|^2$. Ma

$X - X_H = (\overbrace{X_{1,1} - \bar{X}, \dots, X_{1,n_1} - \bar{X}}^{n_1 \text{ volte}}, \dots\dots, \overbrace{X_{k,1} - \bar{X}, \dots, X_{k,n_k} - \bar{X}}^{n_k \text{ volte}})$

$\Rightarrow \|X - X_H\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X})^2$ = VARIAZIONE TOTALE = somma delle deviazioni di ciascun gruppo rispetto alla media globale, da cui la formula di HUYGENS

$$\boxed{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2 + \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2}$$

(variazione totale = variaz. interna + variaz. esterna).

~~Ultimi argomenti "quelle là"~~ (MA il modello stat. BAYESIANO non è tanto "in" come argomento.....).

$\boxed{\text{IL TEST DEL } \chi^2 \text{, O TEST DI ADATTAMENTO.}}$

Consideriamo uno sp. di probabilità $(\Omega, \mathcal{F}, P)$ e, su di esso, un campione di v.a. $X=(X_1,\dots,X_n)$ a valori finiti, nel senso che ogni X_j assume un n° finito di valori: poniamo che, $\forall j=1,\dots,n$, X_j sia a valori nell'insieme $\{1,2,\dots,k\}$, $k\in\mathbb{N}_*$. Dunque, la legge di ogni X_j è data dei k numeri $q_i \doteq P[X_j = i]$ ($i=1,\dots,k$, $j=1,\dots,n$) tali che (come sempre) $q_1,\dots,q_k \underset{(>0)}{\geq 0}$ e $\sum_{i=1}^{k} q_i = 1$.

(REMARK: X è un campione, per cui $X_1,\dots,X_n$ sono indipendenti e hanno la stessa legge, per cui $q_1,\dots,q_k$ NON dipendono da $j=1,\dots,n$!!)

Insomma le X_j sono a valori in $\{1,2,\dots,k\}$ e hanno $P[X_j=i]=q_i$ per ogni $i=1,2,\dots,k$.

Sia $q\doteq(q_1,q_2,\dots,q_k)$ $(\in[0,1]^k)$. Il test del χ^2, o di adattamento, è un problema di test nel quale ci chiediamo se q sia una (nuova legge discreta) data, oppure no: cioè, dato $p\doteq(p_1,p_2,\dots,p_k)\in[0,1]^k$ con $\sum_{i=1}^{k} p_i = 1$, il problema di test è $(p_i>0\ \forall i=1,\dots,k)$

$\boxed{\text{« } q=p \text{ contro } q\neq p \text{ »}}$, cioè che ipotesi nulla

$\boxed{H_0 : q_1 = p_1,\ q_2 = p_2,\ \ldots\ldots,\ q_k = p_k}$. (Ad esempio, potremmo chiederci se esiste una situazione di distribuzione uniforme, cioè se $q_i = \frac{1}{k}$ per ogni $i = 1, \ldots, k$.)

Ad essere più precisi, il test del χ^2, o di adattamento, è il test che si propone di risolvere tale problema basandosi sulla statistica

$$\boxed{T^{(n)} \doteq \sum_{i=1}^{k} \frac{\left(N_i^{(n)} - np_i\right)^2}{np_i}}, \text{ dove}$$

$$N_i^{(n)} \doteq \sum_{j=1}^{n} \mathbb{1}_{\{X_j = i\}}$$

(EFFETTIVO EMPIRICO DEL VALORE i)

(Cochran...)

($\forall i = 1, \ldots, k$, N_i è il numero delle X_j (fra 1 e n) che hanno assunto il valore i)

(T è una norma al quadrato in $\mathbb{R}^k$....)

e dove in particolare T è così scelta in base a calcoli che dimostrano il teorema fondamentale a riguardo, che è il seguente.

Teorema (PEARSON). Su $(\Omega, \mathcal{A}, P)$, sia $(X_n)_{n \geq 1}$ una successione a valori in $\{1, 2, \ldots, k\}$, $k \in \mathbb{N}_k$ dato, di v.a.(r.) i.i.d., e per esse consideriamo

$$N_i^{(n)} \doteq \sum_{j=1}^{n} \mathbb{1}_{\{X_j = i\}} \qquad \forall n \geq 1, \forall i = 1, \dots, k$$

(delle prime n X_j, quante X_j fanno i)

$$T^{(n)} \doteq \sum_{i=1}^{k} \frac{(N_i^{(n)} - np_i)^2}{np_i} \qquad \forall n \geq 1 \quad \text{(specie di media "con" le } N_i^{(n)}\text{)}$$

,

dove introduciamo "fissati" $p_1, p_2, \dots, p_k > 0$ aventi $\sum_{i=1}^{k} p_i = 1$. ~~[illegible]~~ Se $q_1, \dots, q_k$ sono $q_i \doteq P[X_j = i]$ per ogni $i = 1, \dots, k$ (e j qualsiasi, $j = 1, \dots, n$), allora si hanno i due, seguenti casi:

(a) se $q \neq p$, allora $T_n \xrightarrow[n \to +\infty]{P\text{-q.c.}} +\infty$;

(b) $\boxed{\text{se } q = p}$, allora $T_n \xrightarrow[n \to \infty]{\mathcal{L}^P} \chi^2(k-1)$.

$\boxed{\text{NOTA 1}}$ Mentre $N_i^{(n)}$ è l'effettivo EMPIRICO del valore i all'"istante n", $T^{(n)}$ potrebbe esser considerato una specie di ~~[illegible]~~ "scarto" con l'effettivo TEORICO np_i nel caso $q \cong p$:

infatti, per Kolmogorov, $\forall i$, $\frac{1}{n} N_i^{(n)} \xrightarrow[n \to \infty]{P\text{-q.c.}} \mathbb{E}^P[\mathbb{1}_{\{X_j = i\}}] = q_i \overset{(SE)}{\cong} p_i$,

ossia $\forall i = 1, \dots, k$, $N_i^{(n)} \underset{n \to +\infty}{\approx} np_i \; (q_i)$

Dando un attimo per buona la dimostrazione, vediamo dell'importanza del teorema per il nostro test:

dato il punto (a), è chiaro che come regione critica (del campione di n elementi) prenderemo gli eventi della forma $\{T>c\}$, $c>0$,

e per il calcolo di c tale che $\underline{P[T>c]=\alpha}$ ($\in(0,1)$, $\ll 1$, assegnato)

sotto l'ipotesi che $q=p$, possiamo usare per (b) che

$$\alpha = P[T>c] = 1-P[T\le c] \cong 1-\underbrace{F_{k-1}(c)}_{\text{f. di r. della } \chi^2(k-1)},$$

e così che $c \cong \chi^2_{1-\alpha,\,k-1}$ √ ☺

Passiamo alla dimostrazione del TEO. di Pearson

Dim. Il punto (a) è "rapido", mentre per il punto (b) occorrono dei risultati preliminari.

(a) Supponiamo che $\exists\, i_0 \in \{1,2,\dots,k\}$ tale per cui $q_{i_0} \neq p_{i_0}$;

ma, per Kolmogorov, $\frac{1}{n} N^{(n)}_{i_0} \xrightarrow[n\to+\infty]{P\text{-q.c.}} q_{i_0}$ $(\neq p_{i_0})$

e allora
$$T^{(n)} \overset{\text{(ovvio)}}{\geq} \frac{\left(N^{(n)}_{i_0}-np_{i_0}\right)^2}{np_{i_0}} \overset{\text{(calc.)}}{=} n\,\frac{\left(\frac{N^{(n)}_{i_0}}{n}-p_{i_0}\right)^2}{p_{i_0}} \xrightarrow[n\to+\infty]{P\text{-q.c.}} +\infty. \ \surd$$

$\to q_{i_0}-p_{i_0} \neq 0$

(b) Richiamiamo/usiamo tre risultati.

(1) $\boxed{\text{LEMMA}}$ Siano $p_1, p_2, \dots, p_k > 0$ con $\sum_{i=1}^{k} p_i = 1$, e

(Corollario COCHRAN...)

consideriamo la matrice reale $A \in \mathbb{R}^{k^2}$, $A = (a_{i,j})_{i,j=1,\dots,k}$, data da

$a_{i,j} \doteq \delta_{i,j} - \sqrt{p_i}\sqrt{p_j}$, e cioè

$$A = \begin{bmatrix} 1 - p_1 & -\sqrt{p_1 p_2} & \cdots & -\sqrt{p_1 p_k} \\ -\sqrt{p_1 p_2} & 1 - p_2 & \cdots & -\sqrt{p_2 p_k} \\ \vdots & \vdots & \ddots & \vdots \\ -\sqrt{p_1 p_k} & -\sqrt{p_2 p_k} & \cdots & 1 - p_k \end{bmatrix}$$

, che è simmetrica e (⇒ autov. reali).

~~semi-definita positiva (autov. positivi) (ad esempio, per il 1° teorema di Gershgorin~~

~~in generale, l'unione di~~ ~~Noto che $p_1 = \sqrt{p_1 p_1}$~~ ~~$1-p_1$~~ ~~1~~ ~~raggio~~

~~$\sqrt{p_1 p_2} + \dots + \sqrt{p_1 p_k} = \sqrt{p_1}\,\underbrace{\boxed{\sqrt{p_2 + \dots + p_k}}}_{<1} < \sqrt{p_1}$~~

~~0 $\sqrt{p_i}$ $1-p_i$ 1~~

~~$1 - p_i - \sqrt{p_i} > 0$~~

~~$1 > p_i + \sqrt{p_i}$~~

~~$\sqrt{p} \dots \sqrt{p}$ 0 1 p~~ ~~... problema per $p_i \approx 1$...~~

Allora A è semi-definita positiva e ...

se Y è un vettore gaussiano con $\sim N_k(0,A)$, allora $\|Y\|^2 \sim \chi^2(k-1)$.

2) $\boxed{\text{TEOREMA LIMITE CENTRALE "VETTORIALE"}}$: data una (su $(\Omega,\mathcal{A},P)$) succ. $(X_n)_{n\geq 1}$ di vettori aleatori in $\mathbb{R}^k$ i.i.d., aventi media $\mu\in\mathbb{R}^k$ e covarianza $\Gamma\in\mathbb{R}^{k^2}$

(cioè, $\mathbb{E}^P[X_1]=\begin{pmatrix}\mathbb{E}^P[X_{1,1}]\\ \vdots \\ \mathbb{E}^P[X_{1,k}]\end{pmatrix}=\mu$ e $\Gamma_{i,j}=\mathrm{Cov}^P[X_{1,i},X_{1,j}]$, $i,j=1,\dots,k$, se $X_1=(X_{1,1},\dots,X_{1,k})^t$),

vale
$$\frac{X_1+\dots+X_n-n\mu}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}^P} N_k(0,\Gamma),$$

cioè
$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right) \xrightarrow[n\to+\infty]{\mathcal{L}^P} N_k(0,\Gamma).$$

3) $\boxed{\text{ESERCIZIO.}}$ Data $A\in\mathbb{R}^{k^2}$ di proiezione ~~simm. e def. positiva~~, se $(Y_n)_{n\geq 1}$ è una succ. di v.a. tale che $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}^P} N_k(0,A)$, allora $\|Y_n\|^2 \xrightarrow[n\to+\infty]{\mathcal{L}^P} \chi^2(k-1)$.

(Tutte convergenze in legge.)

Schema della dimostrazione: usare ① + ② + ③ per dimostrare ⓑ del teorema, e poi dimostrare dopo le parti ① e ③ (dando il TLC di ② per OK, ovviamente).

Dim. ⓑ (usando ① + ② + ③): (dovrebbe essere facile...) consideriamo

$Y = (Y_1, \dots, Y_M)$" dove, $\forall j = 1, \dots, M$, Y_j è e' "me volte un vettore aleatorio, dato da

$$\boxed{Y_j \doteq \left(\frac{1}{\sqrt{p_1}} 1\!\!1_{\{X_j = 1\}}, \frac{1}{\sqrt{p_2}} 1\!\!1_{\{X_j = 2\}}, \dots, \frac{1}{\sqrt{p_k}} 1\!\!1_{\{X_j = k\}} \right) \in \mathbb{R}^k},$$

Cosicché! (per costruzione)

- $Y_1, \dots, Y_M$ sono $\boxed{\text{i.i.d.}}$ (in $\mathbb{R}^k$)
- $\boxed{\sum_{j=1}^{M} Y_j = \left(\frac{1}{\sqrt{p_1}} N_1^{(M)}, \dots, \frac{1}{\sqrt{p_k}} N_k^{(M)} \right)}$ (OK).

D'altra parte, chiaramente, **nell'ipotesi $q = p$**, ($\forall j$)

$$\rightarrow \quad \mathbb{E}^p[Y_j] = (\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_k}) \qquad (" = \sqrt{p}\,")$$

$$\rightarrow \quad \mathbb{E}^p[Y_{j,m_1} Y_{j,m_2}] = \begin{cases} 0 & \text{se } m_1 \neq m_2 \text{ (ovvio)} \\ \frac{1}{\sqrt{p_{m_1}}} \frac{1}{\sqrt{p_{m_1}}} \cdot p_{m_1} = 1 \end{cases}$$

(componenti m_1-esima e m_2-esima di Y_j)

(m_1, m_2 in $\{1, 2, \dots, k\}$)

da cui facilmente $\underset{(\forall m_1, m_2 = 1,\dots,k)}{}$ $\mathrm{Cov}^P[Y_j]_{m_1,m_2} \stackrel{\Delta}{=}$ (matrice reale simm. semi-def. pos. in $\mathbb{R}^{k^2}$)

$$\doteq \mathrm{Cov}^P[Y_{j,m_1}, Y_{j,m_2}] = \mathbb{E}^P[Y_{j,m_1} \cdot Y_{j,m_2}] -$$

$$- \mathbb{E}^P[Y_{j,m_1}]\,\mathbb{E}^P[Y_{j,m_2}] = \begin{cases} 1 - p_{m_1} & \text{se } m_1 = m_2 \\ -\sqrt{p_{m_1}}\sqrt{p_{m_2}} & \text{se } m_1 \neq m_2 \end{cases} =$$

$\doteq a_{m_1,m_2}$ (dal lemma), il che dimostra anche

che $A \doteq (a_{i,j})_{i,j=1,\dots,k}$ è semi-def. positiva! Adesso, allora,

grazie al TLC "vettoriale", vale

$$\frac{Y_1 + \dots + Y_n - n(\sqrt{p_1}, \dots, \sqrt{p_k})^t}{\sqrt{n}} \xrightarrow[n \to +\infty]{\mathcal{L}^P} N_k(0, A).$$

Ma
$$\frac{Y_1 + \dots + Y_n - n(\sqrt{p_1}, \dots, \sqrt{p_k})}{\sqrt{n}} =$$

$$= \frac{\left(\frac{1}{\sqrt{p_1}} N_1^{(n)} - n\sqrt{p_1}, \dots, \frac{1}{\sqrt{p_k}} N_k^{(n)} - n\sqrt{p_k}\right)}{\sqrt{n}} \stackrel{(\text{ovv.})}{=}$$

$$= \left(\frac{N_1^{(n)} - np_1}{\sqrt{np_1}}, \dots, \frac{N_k^{(n)} - np_k}{\sqrt{np_k}}\right), \xrightarrow{\mathcal{L}^P} N_k(0, A),$$

e quindi per l'ESERCIZIO segue subito che

$$\left\| \left(\frac{N_1^{(n)} - np_1}{\sqrt{np_1}}, \dots, \frac{N_k^{(n)} - np_k}{\sqrt{np_k}} \right) \right\|^2 \overset{(def.)}{=} T^{(n)} \xrightarrow[n\to+\infty]{\mathcal{L}^P} \chi^2(k-1). \;\square$$

Passiamo finalmente alla dim. del Lemma e dell'EX.

ESERCIZIO [Sia $Y \sim N_k(0, A)$ tale che $Y_n \xrightarrow[n]{\mathcal{L}^P} Y$; allora, per continuità di $\|\cdot\|^{(2)}$, risulta che $\|Y_n\|^2 \xrightarrow[n]{\mathcal{L}^P} \|Y\|^2$. Ma, ora,

$\|Y\|^2 \sim \chi^2(k-1)$ per il lemma. ✓]

DIM. LEMMA [Partiamo da Cochran: se (in generale)

$Z = (Z_1, \dots, Z_k)$, $k \geq 2$, è un vettore gaussiano di legge

$N_k(0, I_k)$, cioè è un campione di legge $N(0,1)$,

e se $E \overset{\cdot}{=} \mathrm{Span}\langle \underbrace{(\sqrt{p_1}, \dots, \sqrt{p_k})}_{=:\sqrt{p}} \rangle$ (versore!), allora $Z - Z_E$ e

Z_E sono indipendenti ed inoltre $\|Z - Z_E\|^2 \sim \chi^2(k-1)$.

Vediamo che $\boxed{Z - Z_E \sim N_k(0, A)}$ per concludere.

Ma, infatti, $Z_E = \langle Z; \sqrt{p} \rangle \sqrt{p} = \left(\sum_{i=1}^k Z_i \sqrt{p_i} \right) \sqrt{p}$;

e allora $Z - Z_E = Z - \sqrt{p} \left(\sum_{i=1}^k Z_i \sqrt{p_i} \right)$, ed

è evidente (essendo) che, $\forall u \in \mathbb{R}^k$, $\langle Z - Z_E; u \rangle$ è gaussiana. ✓

$\left(Z_1(\dots) + Z_2(\dots) + \dots + Z_k(\dots) \right)$

Per ciò, basta ottenere che $\mathbb{E}^P[Z - Z_E] = 0$ e

$\mathbb{E}^P[(Z - Z_E)_i (Z - Z_E)_j] = a_{ij}$ per concludere:

→ $\boxed{\text{medie}}$ ($\forall j = 1, \dots, k$)

$$\mathbb{E}^P\left[Z_j - \sqrt{p_j}\sum_{i=1}^k Z_i\sqrt{p_i}\right] = \underbrace{\mathbb{E}^P[Z_j]}_{=0} - \sqrt{p_j}\sum_{i=1}^k \sqrt{p_i}\,\overbrace{\mathbb{E}^P[Z_i]}^{=0} = 0 - 0 = 0 \checkmark$$

→ $\boxed{\text{Covarianze}}$

$$\mathbb{E}^P\left[\left(Z_i - \sqrt{p_i}\sum_{h=1}^k Z_h\sqrt{p_h}\right)\left(Z_j - \sqrt{p_j}\sum_{h=1}^k Z_h\sqrt{p_h}\right)\right] =$$

$$= \underbrace{\mathbb{E}^P[Z_iZ_j]}_{\text{OK}} \left(\overset{=}{\delta_{ij}}\right) - \underbrace{\sqrt{p_i}\sqrt{p_j}}_{\substack{(\text{INDIP.})\\(\text{e centrate})}}\underbrace{\mathbb{E}^P[Z_i^2]}_{=1} - \sqrt{p_i}\sqrt{p_j}\underbrace{\mathbb{E}^P[Z_j^2]}_{=1} +$$

$\leftarrow$ (VAR 1) $\rightarrow$

$$+ \sqrt{p_i}\sqrt{p_j}\underbrace{\sum_{h=1}^k p_h \underbrace{\mathbb{E}^P[Z_h^2]}_{=(1)}}_{=1} =$$

$\left(\sum_{h=1}^k \sqrt{p_h} Z_h \sim N(0,1)!\right)$

$$= \delta_{ij} - \sqrt{p_i}\sqrt{p_j} \overset{(\text{def.})}{=} a_{ij}, \quad \text{OK} :) \quad \square$$

LA FUNZIONE DI RIPARTIZIONE EMPIRICA ED IL METODO DEI MOMENTI

Su $(\Omega, \mathcal{F}, P)$, partiamo da un campione $X =$

$= (X_1, \dots, X_n)$ di legge μ $(= P(X_1))$ sconosciuta, anzi che voler conoscere (μ incognita). (La funzione di ripartizione "usuale" di X_i sarebbe $F(x) = P[X_i \le x] =$

$= \mu((-\infty, x])$, $x \in \mathbb{R}$;) LA FUNZ. DI RIPARTIZIONE EMPIRICA in $x \in \mathbb{R}$ del campione è data da

$$\omega \mapsto \boxed{F_n(x, \omega) \doteq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}}(\omega)}, \quad \omega \in \Omega,$$

con questo $x \in \mathbb{R}$ fissato. Dunque, $\forall x \in \mathbb{R}$,

$F_n(x, \omega)$ è una v.a.r. (su Ω), a valori in

$\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$, ed è la percentuale

delle X_i che stanno sotto x.

Dato che $X_1, \dots, X_n$ sono indipendenti, per Kolmogorov:

$$\boxed{\forall x \in \mathbb{R}, \quad F_n(x, \cdot) \doteq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} \xrightarrow[n \to \infty]{P\text{-q.c.}} F(x)}$$

($\Rightarrow$ $\forall x \in \mathbb{R}$, $F_n(x, \cdot)$ approssima P-q.c. $F(x)$, da cui il nome!).

Non solo: $\forall x \in \mathbb{R}$ e $\forall i=1,\dots,n$ $\mathbb{1}_{\{X_i \le x\}} \begin{cases} \nearrow 0 & 1-F(x) \\ \searrow 1 & F(x) \end{cases}$, e così

$\mathbb{1}_{\{X_i \le x\}} \sim \mathcal{B}(1, F(x))$, per cui che media $F(x)$ (ok) e varianza $F(x)(1-F(x))$ (comunque limitata), e che ora inoltre (per il TLC Lévy)

$$\forall x \in \mathbb{R},\ \sqrt{n}\left(F_n(x,\cdot) - F(x)\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(0, F(x)(1-F(x))\right). \checkmark$$

! $\forall x \in \mathbb{R}$, $\exists A_x \in \mathcal{F}$ con $P(A_x)=1$ tale che, $\forall \omega \in A_x$,

$$F_n(x,\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}}(\omega) \xrightarrow[n\to\infty]{} F(x).$$

Però, anche se fosse $\bigcap_{x\in\mathbb{R}} A_x \in \mathcal{F}$, potrebbe non essere $P\left(\bigcap_{x\in\mathbb{R}} A_x\right)=1$!!

Il fatto è che vorremmo avere

« $F_n(x,\cdot) \xrightarrow[n\to\infty]{P\text{-q.c.}} F(x)$ per ogni $x \in \mathbb{R}$ », e cioè

avere una convergenza uniforme in x ... rispetto a ω

— — — — — — — — —

Il risultato che raffina tutto questo è il seguente, che dà anche la desiderata convergenza uniforme in x.

Teorema (di GLIVENKO – CANTELLI). Su $(\Omega, \mathcal{F}, P)$, sia $(X_n)_{n\geq 1}$ una succ. di v.a.r. i.i.d., e si siano $F_n(x,\omega) \doteq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}(\omega)$, $\forall \omega\in\Omega$, come altre v.a.r., $(\forall n \geq 1)$ per ogni $x \in \mathbb{R}$. Se $(\forall n \geq 1)$

$$\boxed{D_n(\omega) \doteq \sup_{x\in\mathbb{R}} \left| F_n(x,\omega) - F(x) \right|},$$

$\omega \in \Omega$, allora

(1) $\forall n \geq 1$, D_n è una v.a.r.;

(2) $D_n \xrightarrow[n\to+\infty]{P\text{-q.c.}} 0$.

Corollario $\Rightarrow$ $\exists A \in \mathcal{F}$ con $P[A] = 1$ tale che, $\forall \omega \in A$,

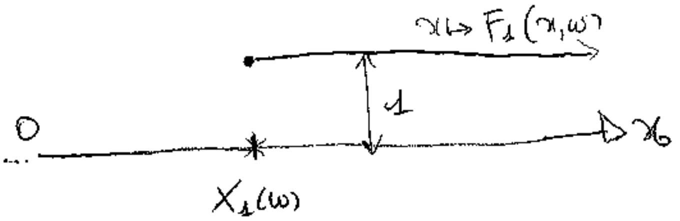
$D_n(\omega) \xrightarrow[n\to+\infty]{} 0$, ossia $\sup_{x\in\mathbb{R}} \left| F_n(x,\omega) - F(x) \right| \xrightarrow[n\to+\infty]{} 0$,

cioè per ω $F_n(x,\omega) \to F(x)$ $\forall x \in \mathbb{R}$ e, ovvero, si ha la conv. uniforme in x.

Prima della dimostrazione (non breve...), alcuni ragionamenti collaterali ....

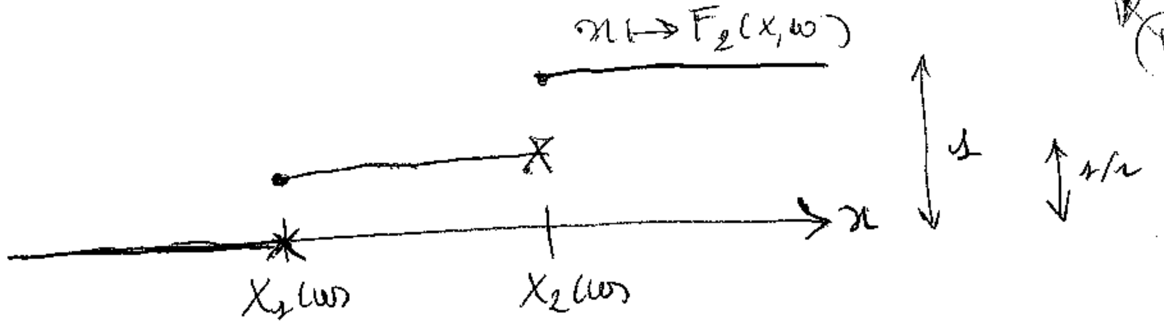
Andiamo

OSS. • Se avessimo solo X_1 (su $(\Omega,\mathcal{A},P)$), allora $F_1(x,\cdot) = 1_{\{X_1\le x\}}$: $\forall\omega\in\Omega$, si ha



. Dunque, mentre $\omega\mapsto F_1(\bar{x},\omega)$ ($\bar{x}\in\mathbb{R}$ fissato) è "solo" una f. di rip. empirica, la $x\mapsto F_1(x,\bar{\omega})$ ($\bar{\omega}\in\Omega$ fissato) è una "vera" f. di rip.: è chiaramente la f. di rip. corrispondente alla probabilità $\delta_{X_1(\bar{\omega})}$ (su $\mathcal{B}(\mathbb{R})$) (delta di Dirac nel punto $X_1(\bar{\omega})$)!! (MASSA 1 SUL PUNTO $X_1(\omega)$)

• Se avessimo solo X_1 e X_2, allora $F_2(x,\cdot)=\frac{1}{2}1_{\{X_1\le x\}}+\frac{1}{2}1_{\{X_2\le x\}}$: $\forall\omega\in\Omega$, immaginando $X_1(\omega)\le X_2(\omega)$ (Non occorre!), si ha



, per cui è chiaro che $x\mapsto F_2(x,\omega)$ sia la f. di rip. corrispondente alla probabilità (su $\mathcal{B}(\mathbb{R})$) $\frac{1}{2}\delta_{X_1(\omega)}+\frac{1}{2}\delta_{X_2(\omega)}$! (MASSA 1/2 SU $X_1(\omega)$ e SU $X_2(\omega)$)

• In generale, $\forall\omega\in\Omega$, $x\mapsto F_n(x,\omega)$ è la f. di rip. corrispondente alla probabilità (reale)

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(\omega)}}$$ (MASSA $\frac{1}{n}$ nei PUNTI $X_1(\omega),\dots,X_n(\omega)$)

METODO DEI MOMENTI (cenni): stimare i momenti di (ogni) X_i usando F_n anziché F!! ($\Rightarrow$ la loro legge μ!) Precisamente:

abbiamo che $X_1,\dots,X_n$ sono (indip. e) identiche, di legge μ; supponiamo che $\int_{\mathbb{R}}|x|\,d\mu(x)<+\infty$, cioè che ($i\in\{1,\dots,n\}$)

$$E^P[|X_i|]=\int_{\mathbb{R}}|x|\,d\mu(x)<+\infty \quad : \quad \text{vogliamo stimare}$$

$\int_{\mathbb{R}}x\,d\mu(x)$. Se $F(x)\doteq P[X_i\le x]=\mu((-\infty,x])$,

allora possiamo scrivere $\int_{\mathbb{R}}x\,d\mu(x)=\int_{\mathbb{R}}x\,dF(x)$:

vogliamo "sostituirlo" con $\int_{\mathbb{R}}x\,dF_n(x,\omega)$, ottenendo una "stima aleatoria". (Ora, effettivamente $\forall\omega\in\Omega$, $F_n(\cdot,\omega)$ è la f. di rip. della legge $\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(\omega)}$)

per cui $\int_{\mathbb{R}}x\,dF_n(x,\omega)=\underbrace{\int_{\mathbb{R}}x\,d\Big(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(\omega)}\Big)(x)}_{(\text{"somma finita"})}\overset{(\text{imm.})}{=}$

$$\overset{(\text{iws})}{=} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \quad ! \quad \text{In effetti}, \ \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow[n\to+\infty]{P\text{-q.c.}} \mathbb{E}^P[X_1]!$$

$$\left(= \overline{X}(\omega)\right)$$

Più in generale, se $\mathbb{E}^P[|X_1|^m] = \int_{\mathbb{R}} |x|^m \, d\mu_{X_1} =$ ($m \in \mathbb{N}_*$)

$= \int_{\mathbb{R}} |x|^m \, dF(x) \overset{<+\infty}{\checkmark}$, allora sappiamo approssimare $\mathbb{E}^P[X_1^m]$ con le v.e.r. (X_i)

$$\boxed{\omega \mapsto \int_{\mathbb{R}} x^m \, dF_n(x,\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega)^m}$$

, che infatti (sempre per Kolmogorov) $\xrightarrow[n\to+\infty]{P\text{-q.c.}}$ momento m-esimo. //

Ad esempio, se $m=2$, allora $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$

$$\xrightarrow[n\to+\infty]{P\text{-q.c.}} \mathbb{E}^P[(X_1)^2] - \mathbb{E}^P[X_1]^2 = \mathrm{Var}^P[X_1].$$

Ma $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\overline{X}}\right)^2 \underset{(\text{soliti conti})}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 =$

$$\left(= \frac{n-1}{n} S^2 \right). \quad //$$

Lasciamo perdere la dim. di G.-C., e vediamo invece una applicazione di G.-C. stesso alla $\boxed{\text{MEDIANA}}$, per pensare infine all'ultimo argomento: i modelli di regressioni!

→ Data una legge di prob. ν su $\mathcal{B}(\mathbb{R})$, un reale $m \in \mathbb{R}$ è <u>UNA MEDIANA DI ν</u> se

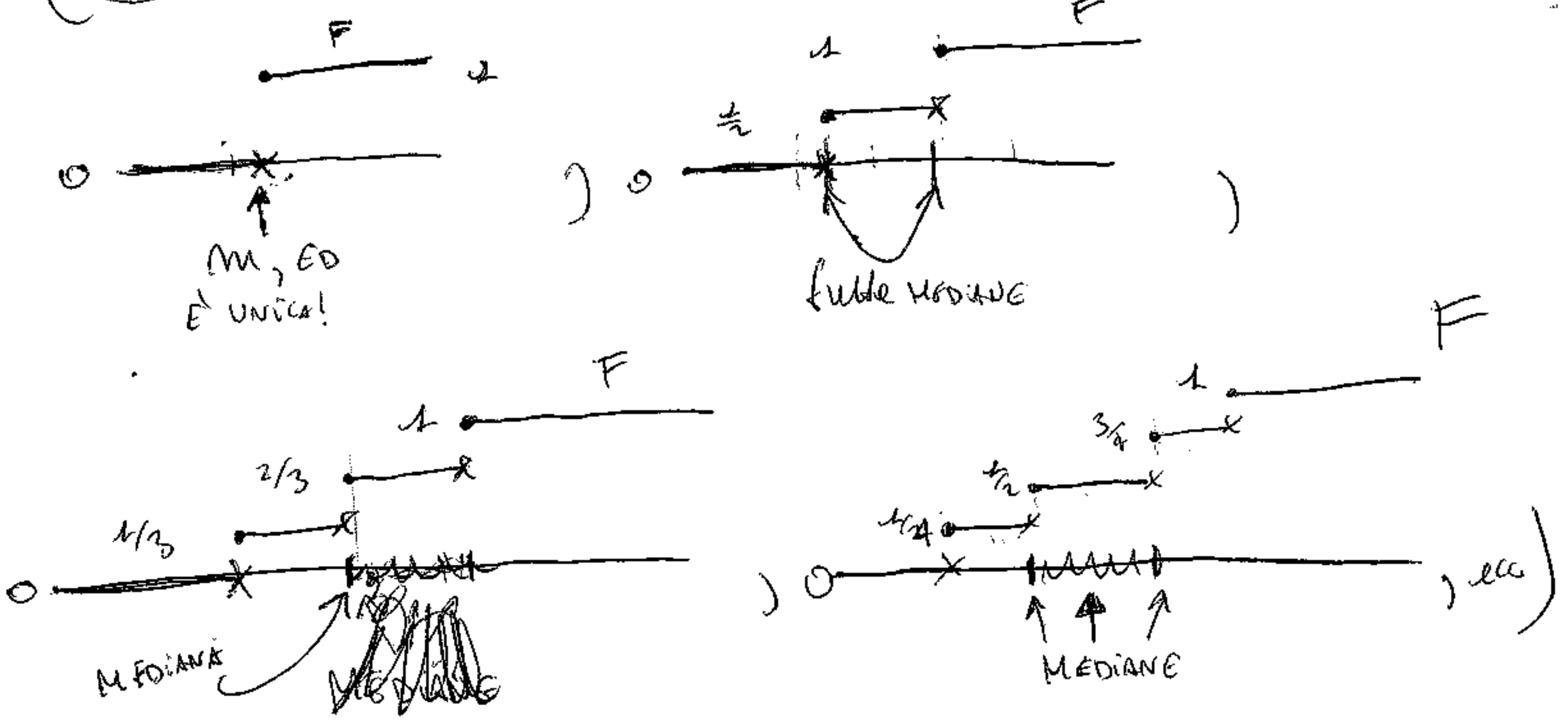
$$\begin{cases} \nu((-\infty, m]) \geq \frac{1}{2} \\ \nu([m, +\infty)) \geq \frac{1}{2} \end{cases} \qquad \left(1 - \nu((-\infty, m)) \geq \frac{1}{2}, \text{ cioè } \nu((-\infty, m)) \leq \frac{1}{2} \right):$$

cioè, se F è la f. di rip. di ν, allora

$$\boxed{F(m^-) \leq \frac{1}{2} \leq F(m)}.$$

(Dunque m è la "<u>centralità</u>" di ν) $\left(\boxed{\nu((-\infty, m)) \leq \frac{1}{2} \leq \nu((-\infty, m])} \right)$ (→ ESISTE SEMPRE!)

((Es) A proposito di $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i(\omega)}$, avevamo che



M dovrebbe essere $m \in \left[\sup\{x \in \mathbb{R} \mid F(x) < \frac{1}{2}\}, \sup\{x \in \mathbb{R} \mid F(x) \leq \frac{1}{2}\} \right]$.

Vantaggio della $\boxed{\text{mediana}}$ (come concetto) rispetto alla $\boxed{\text{media}}$: (oltre al fatto che esiste sempre)
poter lavorare con dati NON numerici!!

$\boxed{\text{NOTA}}$ (1) Se ν è simmetrica con $\int_{\mathbb{R}} |x| \, d\nu(x) < +\infty$,

allora $\mu \doteq \int_{\mathbb{R}} x \, d\nu(x) = m$, che però è unica.
(=0)

(2) Se $\int_{\mathbb{R}} |x|^2 \, d\nu(x) < +\infty$ e $\mu \doteq \int_{\mathbb{R}} x \, d\nu(x)$, $\sigma \doteq \sqrt{\int_{\mathbb{R}} x^2 \, d\nu(x)}$,

allora vale $\boxed{|\mu - m| \leq \sigma}$. (...)
(Poi......)

Solo (1) Sia $(\Omega, \mathcal{F}, P) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$ e sia $X = id$, cosicché
$X : \Omega \to \mathbb{R}$ sia v.a. di legge ν. Allora, per ipotesi
X è integrabile ed ha $\mathbb{E}[X] = \mu$. D'altra parte, per
ipotesi, X è simmetrica, ossia anche $-X$ ha legge ν.

Allora $P[X \leq \mu] = P[-X \leq \mu] = P[X \geq -\mu] =$
$= 1 - P[X < -\mu]$. Ma $\mu = 0$, perché
$\mu = \mathbb{E}[X] \overset{\text{simm.}}{=} \mathbb{E}[-X] = -\mathbb{E}[X] = -\mu$, da cui allora
$1 = P[X \leq 0] + P[X < 0]$, da cui ovviamente
$P[X < 0] \leq \frac{1}{2} \leq P[X \leq 0]$ ✓

Cosa c'entra la mediana con G.-C. ?!

Torniamo a $(\Omega, \mathcal{A}, P)$ e al campione $X = (X_1, \dots, X_n)$ di legge μ (sconosciuta): vogliamo stimare "la" (una) mediana m_μ di μ!!

In che modo: tenendo di conto che la f. di rip. (in x) $F_n(x, \omega)$ di $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$, $\omega \in \Omega$!!

~~Infatti~~ In effetti, per Glivenko-Cantelli, per q.o. $\omega \in \Omega$,

$$F_n(x, \omega) \xrightarrow[n \to +\infty]{\text{P-q.c.}} F(x)$$ (convergenza uniforme!).

Sia dunque $\mu_n(\omega) \doteq \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$, $\omega \in \Omega$, legge di f. di rip. $F_n(x, \omega)$, appunto, e consideriamo

$$\boxed{m_{\mu_n(\omega)} \doteq \inf\left\{ x \in \mathbb{R} \,\middle|\, \mu_n(\omega)((-\infty, x]) \geq \tfrac{1}{2} \right\}} :$$ (min) (*)

allora $m_{\mu_n(\omega)}$ è una mediana per μ, nel senso che

$$\mu_n^{(\omega)}((-\infty, m_{\mu_n(\omega)}]) \geq \frac{1}{2} \geq \mu_n^{(\omega)}((-\infty, m_{\mu_n(\omega)}))$$.

[Se $x_m \downarrow m_{\mu_n(\omega)}$ (nell'insieme!), allora $\mu_n(\omega)((-\infty, x]) \downarrow \mu_n(\omega)((-\infty, m_{\mu_n(\omega)}])$ (per non-decrescenza e continuità a destra di una f. di rip.); ma, appunto, Ogni $x_n \geq m_{\mu_n(\omega)}$ che $\mu_n(\omega)((-\infty, x_n]) \geq 1/2$, da cui

$\mu_n(\omega)((-\infty, m_{\mu_n(\omega)}]) \geq \frac{1}{2}$ e l'inf è un min. ✓

Ma, se l'inf è un min, allora è OVVIO che

si avrà $\mu_n(\omega)((-\infty, m_{\mu_n(\omega)})) \leq \frac{1}{2}$, perché

se $x_n \uparrow m_{\mu_n(\omega)}$ $\left((-\infty, x_n] \uparrow (-\infty, m_{\mu_n(\omega)})\right)$, allora
$\Updownarrow$
con $x_n < m_{\mu_n(\omega)}$

$\mu_n(\omega)((-\infty, x_n]) \uparrow \mu_n(\omega)((-\infty, m_{\mu_n(\omega)}))$, ma

$\mu_n(\omega)((-\infty, x_n]) < \frac{1}{2}$ $\forall n$, perché appunto $x_n < m_{\mu_n(\omega)}$.

Teorema (APPROX. MEDIANA) Nelle notazioni precedenti,

$\boxed{1}$ $\omega \mapsto m_{\mu_n(\omega)}$ è una v.a.r. ($\forall n \geq 1$).

$\boxed{2}$ Se m_μ è l'unica mediana di μ, allora

$$m_{\mu_n(\cdot)} \xrightarrow[n\to+\infty]{P\text{-q.c.}} m_\mu .$$

$\boxed{1}$ Per def., $m_{\mu_n(\omega)} = \inf\{x \in \mathbb{R} \mid F_n(x,\omega) \geq \frac{1}{2}\}$,

da cui facilmente la tesi (...).

$\boxed{2}$ Fissiamo $\varepsilon > 0$: allora, dato che m_μ e' L'UNICA mediana di μ, esiste $\delta \equiv \delta(\varepsilon) > 0$ tale che

$$F(m_\mu - \varepsilon) < \tfrac{1}{2} - \delta \; (<) \; \tfrac{1}{2} + \delta < F(m_\mu + \varepsilon),$$

nel senso
$$\begin{cases} F(m_\mu - \varepsilon) < \tfrac{1}{2} - \delta \\ F(m_\mu + \varepsilon) > \tfrac{1}{2} + \delta \end{cases}$$
, perche' ESISTE

$$\delta < \underbrace{\left(\tfrac{1}{2} - F(m_\mu - \varepsilon)\right)}_{>0} \wedge \underbrace{\left(F(m_\mu + \varepsilon) - \tfrac{1}{2}\right)}_{>0}.$$

D'altra parte, per Glivenko - Cantelli, esiste $A \in \mathcal{F}$ con $P[A] = 1$ tale che, $\forall \omega \in A$,

$$\sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| \xrightarrow[n \to \infty]{} 0.$$ Dunque

$\forall \omega \in A$ fissato, esiste $N \equiv N(\omega) \in \mathbb{N}_*$ tale che

$\forall n \geq N$, sia $\sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| \leq \delta$:

dimostriamo $\boxed{2}$ verificando che $\forall n \geq N$ (dunque $\equiv N(\omega)$),

$|m_{\mu_n}(\omega) - m_\mu| < \varepsilon$

ossia $\begin{cases} m_{\mu_n}(\omega) - m_\mu < \varepsilon \\ m_{\mu_n}(\omega) - m_\mu > -\varepsilon \end{cases}$, cioe' $\begin{cases} m_{\mu_n}(\omega) < m_\mu + \varepsilon \\ m_{\mu_n}(\omega) > m_\mu - \varepsilon \end{cases}$

Allora :

(Supremo LOGA)

→ se, per assurdo, ci fosse $m \geq N$ tale che

$\boxed{m_{\mu_m}(\omega) \geq m_\mu + \varepsilon}$, allora sarebbe

$$\left(\tfrac{1}{2}+\delta \checkmark\right)\; F(m_\mu + \varepsilon) \leq F_m(m_\mu + \varepsilon, \omega) + \delta \overset{\text{def. } m_{\mu_m}}{\leq} \tfrac{1}{2}+\delta$$

$\left(\forall x \in \mathbb{R},\ \text{essendo } m \geq N = N(\omega),\ \begin{cases} F_m(x,\omega) - F(x) \leq \delta \\ F_m(x,\omega) - F(x) \geq -\delta \end{cases}\right)$

$\left(F_m(x,\omega) - \delta \leq F(x) \leq F_m(x,\omega) + \delta\right)$

$\leq \tfrac{1}{2}+\delta$, che è assurdo. ✓

→ Analog., se per assurdo $\exists\, m \geq N$ tale che

$\boxed{m_{\mu_m}(\omega) \leq m_\mu - \varepsilon}$, allora avremmo

$$\left(\tfrac{1}{2} - \delta >\right)\; F(m_\mu - \varepsilon) \overset{(OK)}{\geq} F_m(m_\mu - \varepsilon, \omega) - \delta \geq$$

$$\overset{\text{def. } m_{\mu_m}}{\geq} \geq \tfrac{1}{2} - \delta \quad \square$$

MODELLI STAT. BAYESIANI

L'idea di modello stat. BAYESIANO è quella di modello stat. "usuale" MA tale per cui si ponga una sorta di preferenza A PRIORI sui parametri. Per definire meglio, facciamo proprio il Bayes della probabilità.

Illustriamo. ES. (INTRODUTTIVO) $\boxed{A_1}$ ⤳ un segnale proviene il 40% delle volte da $\boxed{A_2}$ ⤳

da $\boxed{A_1}$, il 60% delle volte da $\boxed{A_2}$ (questo lo diamo per buono e fisso); un segnale può essere o di tipo "lungo" L o di tipo "breve" B, il segnale che A_1 trasmette B il 48% delle volte, A_2 63%. Riceviamo un segnale breve B: qual è la probabilità che venga da A_1? (" " da A_2?)

⟶ (L'esito dell'esperimento trasforma le probabilità e fusioni! / rivoluta)

Ci serve la formula di Bayes (con le probabilità condizionata):

~~sotto che~~ ~~...~~ ~~... o che ... con un effetto di~~ (il segnale viene da A_1) (il segnale viene da A_2)

~~...~~ dato che $P[A_1] = 0.4$, $P[A_2] = 0.6$,

$P[B|A_1] = 0.48$, $P[B|A_2] = 0.63$, otteniamo

$$P[A_1|B] = \frac{P[B|A_1]P[A_1]}{P[B|A_1]P[A_1] + P[B|A_2]P[A_2]} \simeq 0.3368$$

$$\Rightarrow P[A_2|B] \simeq 0.6632 .$$ ✓

Per ciò, è come se avessimo $\boxed{\Omega \doteq \{B, L\}}$ (con $\mathcal{Y} \doteq \mathcal{P}(\Omega)$) ~~[illegible]~~ e $\Theta \doteq \{1, 2\}$, "ϑ INDICE DELL'APPARECCHIATURA", cioè come se avessimo il modello stat.

$$\left(\Omega = \{B, L\},\ \mathcal{Y} = \mathcal{P}(\Omega),\ \{P^1, P^2\}\right)$$

(interamente dominato), dove ovviamente

$$\begin{cases} P^1[\underset{(L)}{B}] \doteq P[A_1 | \underset{(L)}{B}] \\ P^2[\underset{(L)}{B}] \doteq P[A_2 | \underset{(L)}{B}] \end{cases}$$

, scegliendo "ϑ" (nell'es. prec., sceglievamo P^2, cioè A_2, dato B!)

"prior" che $\begin{cases} 1 \text{ ha preferenza } 0.4 \\ 2 \text{ ha preferenza } 0.6 \end{cases}$ ($\rightarrow$ 2 è più probabile...).

Def. Un modello stat. $(\Omega, \mathcal{Y}, \{P^\vartheta | \vartheta \in \Theta\})$ è BAYESIANO se esiste una σ-algebra $\mathcal{T}$ su Θ tale che, $\forall A \in \mathcal{Y}$, la funzione $\Theta_\vartheta \rightarrow [0, 1]$, $\vartheta \mapsto P^\vartheta[A]$, sia $\mathcal{T}$-misurabile. In tal caso, ci ricordiamo che $\mathcal{D}$ sarebbe

$\left(\Omega, \mathcal{F}, \{P^\vartheta \mid \vartheta \in (\Theta, \mathcal{T})\}\right)$. Ogni misura di probabilità $\checkmark$ su $\mathcal{T}$ è una legge a priori su Θ. $\checkmark$

→ Conviene ricordare il concetto di NUCLEO (DI TRANSIZIONE) fra due spazi misurabili:

1° spazio: CI INTERESSANO I PUNTI

2° spazio: CI INTERESSANO GLI EVENTI

dati $(E_x, \mathcal{E})$, $(F_{\mathcal{F}}, \mathcal{F}_A)$ due spazi misurabili, un NUCLEO DI E SU F "$(E,\mathcal{E}) \overset{N}{\rightsquigarrow} (F,\mathcal{F})$" è una funzione $N(x,A) : E_x \times \mathcal{F}_A \to [0,1]$ t.c.

1) $\forall A \in \mathcal{F}$, $\begin{array}{l} E \to [0,1] \\ x \mapsto N(x,A) \end{array}$ è $\mathcal{E}$-misurabile;

2) $\forall x \in E$, $\begin{array}{l} \mathcal{F} \to [0,1] \\ A \mapsto N(x,A) \end{array}$ è una probabilità su $\mathcal{F}$.

Brevemente, $N(x, A)$

in x, è $\mathcal{E}$-misurabile

in A, è una probabilità su $\mathcal{F}$.

$\boxed{\text{OSS}}$ $(\Theta_\vartheta, \mathcal{T}) \overset{N(\vartheta,A)}{\rightsquigarrow} (\Omega, \mathcal{F}_A)$, dove

$$N(\vartheta, A) \doteq P^\vartheta[A]$$, è un nucleo di $(\Theta, \mathcal{T})$ su $(\Omega, \mathcal{F})$ se, e solo se, $(\Omega, \mathcal{F}, \{P^\vartheta \mid \vartheta \in (\Theta, \mathcal{T})\})$ (mis.) è un modello stat. Bayesiano.
(questo è DEF.)

$\boxed{\text{NOTA}}$ Sfruttando il fatto che $N(x, \cdot)$ sia, $\forall x \in E$, una probabilità, potremmo scrivere anche

$$N(x, A) =: N^x(A) \quad (o\ N_x(A)).$$

Il primo risultato "vero", importante sui nuclei è il seguente (che ricordiamo.)

$\boxed{\text{Teorema (di FUBINI, generalizzato).}}$ Sia $N(x, A)$ un nucleo di $(E_x, \mathcal{E})$ su $(F, \mathcal{F}_A)$, e sia P una probabilità su $\mathcal{E}$: $(E_x, \mathcal{E}, P) \overset{N(x,A)}{\rightsquigarrow} (F, \mathcal{F}_A)_{N^x(\cdot)}$

Allora

(a) data $f : E \times F \to \mathbb{R}$ limitata che sia $\mathcal{E} \otimes \mathcal{F}$-mis.,

la funzione $x \mapsto \int_F f(x, y)\, \underbrace{dN(x, y)}_{dN^x(y),\ o\ N(x, dy)}$, $x \in E$

$\mathcal{E}$ - misurabile, allora esiste unica probabilità Q su $\mathcal{E}\otimes\mathcal{F}$ (indip. da f)

tale che valga la formula

$$\iint_{E\times F} f(x,y)\, dQ(x,y) = \int_E dP(x) \int_F f(x,y) N(x,dy).$$

$(Q(dx,dy))$

(b) Ciò vale anche per $f \geq 0$ non necess. limitata, e allora per f arbitraria ($\mathcal{E}\otimes\mathcal{F}$-mis.) f è Q-int.

se, e solo se, $\iint_{E\times F} |f(x,y)|\, Q(dx,dy) \overset{\infty}{=} \int_E P(dx) \int_F |f(x,y)| N(x,dy)$, $< +\infty$

allora nel tal caso resta vero che

$$\iint_{E\times F} f(x,y)\, Q(dx,dy) = \int_E P(dx) \int_F f(x,y) N(x,dy).$$

Torniamo al nostro modello stat. bayesiano, che immaginiamo come munito di una legge a priori sui parametri: sia dunque

$\left(\Omega, \mathcal{Y}, \{P^\vartheta \mid \vartheta \in (\Theta, \mathcal{T}, \nu)\}\right)$, con effetto $\vartheta \mapsto P^\vartheta[A]$ $\mathcal{T}$-mis. per ogni fissato $A \in \mathcal{Y}$.

Dunque, ~~[illegible]~~ (equiv.)

$$\boxed{(\Theta, \mathcal{T})_{\nu(\cdot)} \overset{N(\vartheta, A)}{\leadsto} (\Omega, \mathcal{Y}_A)_{N^\vartheta(\cdot)}}$$ dove

$N(\vartheta, A) \doteq N^\vartheta[A] \doteq P^\vartheta[A]$ è un <u>nucleo</u> di $(\Theta, \mathcal{T})$ su $(\Omega, \mathcal{Y})$, e grazie a FUBINI sappiamo che

$\exists!$ probabilità $Q \doteq Q_{\nu, N}$ su $\mathcal{T} \otimes \mathcal{Y}$ tale che, per ogni $f : \Theta \times \Omega \to \mathbb{R}$ $\mathcal{T} \otimes \mathcal{Y}$-misurabile e Q integrabile, valga la formula

$$\boxed{\iint_{\Theta \times \Omega} f(\vartheta, \omega)\, Q(d\vartheta, d\omega) = \int_{\Theta} \nu(d\vartheta) \int_{\Omega} f(\vartheta, \omega)\, N(\vartheta, d\omega) \doteq \int_{\Theta} \nu(d\vartheta) \int_{\Omega} f(\vartheta, \omega)\, P^\vartheta(d\omega)}$$

Ma in particolare, per ogni $T \in \mathcal{T}$ e $A \in \mathcal{Y}$,

$f(\vartheta, \omega) \doteq \mathbb{1}_{T \times A}(\vartheta, \omega)$ $\left(= \mathbb{1}_T(\vartheta) \cdot \mathbb{1}_A(\omega)\right)$ è abbondantemente $\mathcal{T}\otimes\mathcal{Y}$ - mis. e limitata, per cui vale

$$\underbrace{\iint_{\Theta\times\Omega} 1\!\!1_{T\times A}(\vartheta,\omega)\, Q(d\vartheta,d\omega)}_{\substack{\| \text{ (ovvio)} \\ Q(T\times A)}} = \int_{\Theta} \nu(d\vartheta) \int_{\Omega} \underbrace{1\!\!1_{T\times A}(\vartheta,\omega)}_{\substack{\| \\ 1\!\!1_T(\vartheta)\, 1\!\!1_A(\omega)}} P^{\vartheta}(d\omega) =$$

$$= \int_T \nu(d\vartheta)\, P^{\vartheta}[A] \quad : \quad \text{riassumendo,}$$

$\forall T \in \mathcal{T}$ e $\forall A \in \mathcal{Y}$, vale la formula

$$\boxed{Q(T\times A) = \int_T P^{\vartheta}[A]\, d\nu(\vartheta)}$$

$$\left(\begin{array}{l} Q(\Theta\times A) = \int_{\Theta} P^{\vartheta}[A]\, d\nu(\vartheta) \\ Q(T\times\Omega) = \nu(T) \\ \qquad\qquad (\text{anche } \leq) \end{array} \right)$$

Guardiamo meglio il prodotto

$$(\Theta\times\Omega, \mathcal{T}\otimes\mathcal{Y}).$$

→ Cominciamo guardando $\boxed{\Theta \times A}$, al variare di $A \in \mathcal{Y}$:

è chiaro che $\mathcal{F}_{\mathcal{Y}} \doteq \{\Theta \times A \mid A \in \mathcal{Y}\}$ sia una sotto-σ-algebra di $\mathcal{T} \otimes \mathcal{Y}$, e abbiamo visto

che (grazie a FUBINI) la probabilità Q univocamente associata alla "situazione" $(\Theta, \mathcal{T})_{\nu} \overset{P^{\vartheta}[\cdot]}{\cdots\cdots\rightharpoonup} (\Omega, \mathcal{Y}_A)_{P^{\vartheta}[\cdot]}$ (ha

$$Q(T \times A)\Big|_{T=\Theta} = \int_T P^{\vartheta}[A]\, d\nu(\vartheta)\Big|_{T=\Theta} =$$

$$= \int_{\Theta} P^{\vartheta}[A]\, d\nu(\vartheta) \quad , \forall A \in \mathcal{Y}.$$

Consideriamo inoltre una funzione $U(\vartheta, \omega) : \Theta \times \Omega \to \mathbb{R}$:

allora $U(\vartheta, \omega)$ è $\mathcal{F}_{\mathcal{Y}}$-misurabile (e quindi misurabile)

se, e solo se, $U(\vartheta, \omega)$ è in realtà funzione della sola ω,

nel senso che $U(\vartheta, \omega) \equiv \tilde{U}(\omega)$ $\forall \vartheta \in \Theta$, $\tilde{U} : \Omega \to \mathbb{R}$ mis.

[Per quanto possa essere facile la dimostrazione di tale equivalenza, osserviamo subito che è un'immediata conseguenza del fondamentale criterio di misurabilità di Doob, in quanto $\mathcal{F}_{\mathcal{Y}}$ è la σ-algebra generata dalla funzione $\Theta \times \Omega \to \Omega$, $(\vartheta, \omega) \mapsto \omega$ sul secondo fattore.]

→ Analogamente, e riguardo di $T \times \Omega$ al ovvero se $T \in \mathcal{O}$, $\tilde{\mathcal{T}} \doteq \{T \times \Omega \mid T \in \mathcal{O}\}$ è una sotto-σ-algebra di $\mathcal{T} \otimes \mathcal{Y}$, quella generata dalla proiezione $\Theta \times \Omega \to \Theta$, $(\theta,\omega) \mapsto \theta$, per cui una funzione $V(\theta,\omega) : \Theta \times \Omega \to \mathbb{R}$ è $\tilde{\mathcal{T}}$-mis. se, e solo se, esiste (unica) funzione $\mathcal{V} : \Theta \to \mathbb{R}$ mis. tale che

$$V(\theta,\omega) = \mathcal{V}(\theta) \quad \forall (\theta,\omega) \in \Theta \times \Omega.$$

(o per quasi qui $(\theta,\omega) \in \Theta \times \Omega$) Inoltre,

$$Q(T \times \Omega) = \nu(T) \quad \forall T \in \mathcal{O}.$$

(Dunque $Q\big|_{\tilde{\mathcal{T}}} = \nu$ ")

$\boxed{\text{Teorema}}$. Se $X(\theta,\omega) : \Theta \times \Omega \to \mathbb{R}$ è una v.a. $\mathcal{T} \otimes \mathcal{Y}$-misurabile ed è limitata, allora la sua parte $\tilde{\mathcal{T}}$-misurabile è data da

$$\mathbb{E}^Q[X \mid \tilde{\mathcal{T}}](\theta) \overset{Q\text{-q.c.}}{=} \int_\Omega X(\theta,\omega)\, dP^\theta(\omega), \quad \forall \theta \in \Theta.$$

(~~[illegible]~~ Sarebbe $\mathbb{E}^Q[X \mid \tilde{\mathcal{T}}](\omega,\theta)$.)

[Dim. Grazie al teorema di FUBINI,

sappiamo benissimo che $\vartheta \mapsto \int_\Omega X(\vartheta,\omega)\, dP^\vartheta(\omega)$ è Θ-misurabile

(X è limitata!, ~~[illegible]~~), per cui basta

verificare che, $\forall T \in \Theta$,

$$\iint_{T\times\Omega} X(\vartheta,\omega)\, Q(d\vartheta, d\omega) = \iint_{T\times\Omega} \left(\int_\Omega X(\vartheta,\omega)\, dP^\vartheta(\omega)\right)(\vartheta)\, Q(d\vartheta,d\omega).$$

Ma, infatti, sempre per FUBINI vale

$$\iint_{T\times\Omega} X(\vartheta,\omega)\, Q(d\vartheta,d\omega) = \int_T d\nu(\vartheta) \int_\Omega X(\vartheta,\omega)\, dP^\vartheta(\omega),$$

e sempre per FUBINI vale

$$\iint_{T\times\Omega} \left(\int_\Omega X(\vartheta,\omega)\, dP^\vartheta(\omega)\right)(\vartheta)\, Q(d\vartheta,d\omega) =$$

$$= \int_T d\nu(\vartheta) \int_\Omega \underbrace{\left(\int_\Omega X(\vartheta,\omega)\, dP^\vartheta(\omega)\right)(\vartheta)}_{\text{(NON dipende da } \omega\text{!)}}\, dP^\vartheta(\omega) =$$

$$= \int_T d\nu(\vartheta) \left(\int_\Omega X(\vartheta,\omega)\, dP^\vartheta(\omega)\right)\cdot 1 . \qquad \square$$

$\boxed{\text{OSS.}}$ Dalle precedenti formule, otteniamo subito che: se $\bar{\vartheta} \in \Theta$ è tale che $\{\bar{\vartheta}\} \in \mathcal{T}$ e $\nu[\{\bar{\vartheta}\}] > 0$, allora risulta

$$\boxed{P^{\bar{\vartheta}}[A] = \underbrace{Q[\Theta \times A \mid \{\bar{\vartheta}\} \times \Omega]}_{\text{(probabilità condizionata)}}}$$

per ogni $A \in \mathcal{F}$. $(\mathcal{T} \otimes \mathcal{F})$

[Infatti, ovviamente, $Q[\{\bar{\vartheta}\} \times \Omega] = \nu[\{\bar{\vartheta}\}] > 0$, e allora

$$Q[\Theta \times A \mid \{\bar{\vartheta}\} \times \Omega] = \frac{Q[(\Theta \times A) \cap (\{\bar{\vartheta}\} \times \Omega)]}{\nu[\{\bar{\vartheta}\}]},$$

quindi $Q[\Theta \times A \cap \{\bar{\vartheta}\} \times \Omega] = Q[\{\bar{\vartheta}\} \times A] =$

$$= \int_{\{\bar{\vartheta}\}} P^{\vartheta}[A] \, d\nu(\vartheta) = P^{\bar{\vartheta}}[A] \, \nu[\{\bar{\vartheta}\}],$$

da cui la tesi. ✓]

Morale: Com'era previsto, abbiamo $P^{\vartheta}[A]$ per ogni $A \in \mathcal{F}$ <u>noto il valore di ϑ</u>, e questo grazie ad un diagramma del tipo

$$(\Theta_{\vartheta}, \mathcal{T})_{x_j} \xrightarrow{N(\vartheta, A) = P^{\vartheta}[A]} (\Omega, \mathcal{F}_A)_{P^{\vartheta}[A]}.$$

L'idea allora di "aggiornare"/risolvere la legge a priori ν su $\mathcal{T}$ noto il valore di ω (come nell'Es. iniziale delle formule di Bayes) ci spinge a volere che $P^{\omega}[A]$ non sia solo un numero: ma che sia "misurabile", nel senso che esiste nucleo di $(\Omega, \mathcal{G})$ su $(\Theta, \mathcal{T})$

$$(\Omega_{\omega}, \mathcal{G}) \overset{N(\omega,T)}{\cdots\cdots\rightsquigarrow} (\Theta, \mathcal{T})_{N(\omega,\cdot)}$$

per il quale valgono comunque delle formule "tipo le precedenti", per cui tale che per ogni

$X : \Theta \times \Omega \to \mathbb{R}$ $\mathcal{T}\otimes\mathcal{G}$-mis. e limitate risulti

$$\boxed{\mathbb{E}^{Q}[X \mid \mathcal{G}](\theta,\omega) \overset{Q\text{-q.c.}}{=} \int_{\Theta} X(\theta,\omega)\, N(\omega, d\theta)}$$

In tal caso $N(\omega,\cdot)$ è legge a posteriori condizionale a ω (su $\mathcal{T}$)

(→ che aggiorna ν!)

~~Come ripromesso, dalle nostre ipotesi, possiamo assumere~~

~~che $N(\omega, T) = Q[\Theta \times \{\omega\} \mid T \times \Omega]$,~~

~~se $\nu(T) > 0$ e $\{\omega\} \in \Omega$: $Q[T \times \{\omega\}] =$~~

~~$= \int p^{\omega}[\{\omega\}] \, d\nu(\theta)$~~

(Y misurabile $\Omega \times \Theta \to \mathbb{R}$)

$\boxed{\text{Nota.}}$ In tal caso, un FUBINI ci direbbe che

$\omega \mapsto \int_{\Theta} Y(\omega,\theta) \, N(\omega, d\theta)$ è $\mathcal{F}$-misurabile, e che

per una ed una sola probabilità $\tilde{Q}$ su $\mathcal{F} \otimes \mathcal{T}$ vale

$$\iint_{\Omega \times \Theta} Y(\omega,\theta) \, \tilde{Q}(d\omega, d\theta) = \int_{\Omega} dP(\omega) \int_{\Theta} Y(\omega,\theta) \, N(\omega, d\theta)$$

(per una fissata probabilità P su $\mathcal{F}$)

...

Boh... Più o meno parrebbe ok come soluzione...

$\triangleright$ $\tilde{Q}(A \times \Theta) = P[A] \quad \forall A \in \mathcal{F}$

$\tilde{Q}(\Omega \times T) = \int_{\Omega} N(\omega, T) \, dP(\omega) \quad \forall T \in \mathcal{T}$

⟶ Quale che sia P (per es., P^{θ^*}) e sia $\tilde{Q}$ corrispondente, vale come volevamo che

$$N(\bar{\omega}, T) = \tilde{Q}\left[\Omega \times T \mid \{\bar{\omega}\} \times \Theta\right],$$

per ogni $\bar{\omega} \in \Omega$ e $T \in \mathcal{T}$ tale che $\{\bar{\omega}\} \in \mathcal{F}$ e $P[\{\bar{\omega}\}] > 0$:

(infatti, essendo $\tilde{Q}[\{\bar{\omega}\} \times \Theta] = P[\{\bar{\omega}\}] > 0$,

vale che $\tilde{Q}[\text{"intersezione"}] = \tilde{Q}[\{\bar{\omega}\} \times T] =$

$$= \int_{\{\bar{\omega}\}} N(\omega, T)\, dP(\omega) = N(\bar{\omega}, T)\, P[\{\bar{\omega}\}], \text{ ok.})$$

☺

L'esistenza di una legge condizionale a posteriori è un problema piuttosto delicato, in generale... Ma mettiamoci nel caso di modello dominato

$(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in (\Theta, \mathcal{T}, \nu)\}, \mu)$ con

$L^\theta(\omega) = \frac{dP^\theta}{d\mu}(\omega)$ che sia $\mathcal{T} \otimes \mathcal{F}$-misurabile

$\Rightarrow$ ($\forall A \in \mathcal{Y}$) $\vartheta \mapsto P^\vartheta[A]$ è $\mathcal{T}$-misurabile, perché $\forall A \in \mathcal{Y}$

$$P^\vartheta[A] = \int_A L^\vartheta(\omega)\, d\mu(\omega)$$

, ed ora

$$\vartheta \mapsto \int_A L^\vartheta(\omega)\, d\mu(\omega) = \int_\Omega 1_A(\omega) L^\vartheta(\omega)\, d\mu(\omega) \text{ è}$$

$\mathcal{T}$-mis. per Fubini "classico". ✓)

Dunque $(\Theta, \mathcal{T}) \xrightarrow{P^\vartheta[A]} (\Omega, \mathcal{Y}_A)_{P^\vartheta[A]}$ è un nucleo, e allora $\exists!$ prob. Q su $\mathcal{T} \otimes \mathcal{Y}$

tale che
$$\iint_{\Theta \times \Omega} X(\vartheta, \omega)\, Q(d\vartheta, d\omega) = \int_\Theta d\nu(\vartheta) \int_\Omega X(\vartheta, \omega)\, dP^\vartheta(\omega)$$

per ogni $X(\vartheta, \omega) : \Theta \times \Omega \to \mathbb{R}$ mis. limitata (nel senso che

$\vartheta \mapsto \int_\Omega X(\vartheta, \omega)\, dP^\vartheta(\omega)$ è Θ-mis. e che vale la formula),

ed in particolare
$$\begin{cases} Q(T \times \Omega) = \nu(T) & \forall T \in \mathcal{T} \\ Q(\Theta(T) \times A) = \int_{\Theta(T)} P^\vartheta[A]\, d\nu(\vartheta) & \end{cases}$$
✓

$\longrightarrow$ Risultato principale:

Teorema. Nelle situazioni or ora descritte, poniamo

$$A \doteq \left\{\omega\in\Omega \;\middle|\; \int_{\Theta} L^{\vartheta}(\omega)\,d\nu(\vartheta) = 0\right\}$$

$$M \doteq \left\{\omega\in\Omega \;\middle|\; \int_{\Theta} L^{\vartheta}(\omega)\,d\nu(\vartheta) = +\infty\right\}$$

$\in \mathcal{Y}$ (per FUBINI!)

e (convenendo che (costante reale)$/{+\infty} = 0$) poniamo

$$\boxed{q(\vartheta,\omega) \doteq \begin{cases} \dfrac{L^{\vartheta}(\omega)}{\int_{\Theta} L^{\tau}(\omega)\,d\nu(\tau)} & \text{se } \omega\notin A \\ 1 & \text{se } \omega\in A \end{cases}, \quad \vartheta\in\Theta.}$$

Allora:

(a) $\Theta\times A$ che $Q[\Theta\times A]=0$. ($\Rightarrow \forall T\in\mathcal{T}, Q[T\times A]=0$!)

(b) $\mu[M]=0$.

(tanto è μ-trasc.!) (misurabile ed è)

(c) $\forall\omega\in\Omega\setminus M$, $\vartheta\mapsto q(\vartheta,\omega)$ è una densità di probabilità rispetto a ν nello spazio $(\Theta,\mathcal{T})$.

(d) La funzione $N: \underset{\omega}{(\Omega\setminus M)}\times\underset{T}{\mathcal{T}}\to[0,1]$,

$$N(\omega,T) \doteq \int_T q(\vartheta,\omega)\,d\nu(\vartheta), \quad \omega\in\Omega\setminus M \text{ e } T\in\mathcal{T},$$

è un nucleo di $(\Omega_\omega,\mathcal{Y})$ su $(\Theta,\mathcal{T}_T)$, cioè (una) legge a posteriori su $(\Theta, \mathcal{T})$ condizionale a ω

Dim. (a) $Q[\Theta \times A] \overset{(\text{visto})}{=} \int_\Theta P^\theta[A]\, d\nu(\theta) =$ (Tonelli e FUBINI)

$$= \int_\Theta d\nu(\theta) \underbrace{\int_\Omega 1_A(\omega)\, dP^\theta(\omega)}$$

$$\int_\Omega 1_A(\omega) L^\theta(\omega)\, d\mu(\omega) = \int_A L^\theta(\omega)\, d\mu(\omega)$$

$$\overset{(\text{FUBINI})}{=} \int_A d\mu(\omega) \int_\Theta L^\theta(\omega)\, d\nu(\theta) = 0$$

(che bellezza di avere μ!)

per def. di A. ✓

(b) "Di nuovo," $\int_\Omega d\mu(\omega) \int_\Theta L^\theta(\omega)\, d\nu(\theta) \overset{(\text{FUBINI})}{=}$

$$= \int_\Theta d\nu(\theta) \underbrace{\int_\Omega L^\theta(\omega)\, d\mu(\omega)}_{=1} = \nu(\Theta) = 1,$$

per cui è OVVIO che $\mu[M] = \mu[\{\omega \in \Omega \mid \int_\Theta L^\theta(\omega)\, d\nu(\theta) = 0\}]$

fa 0. ✓

(c) È immediato anche questo, perché ~~[illegible]~~ $g(\theta,\omega)$ è ≥ 0 $\forall \theta \in \Theta$ e anche $\forall \omega \in \Omega$, ed inoltre che $(\forall \omega \in \Omega \setminus M)$ $\cdot$ $\int_{\Theta} g(\theta,\omega)\, d\nu(\theta) = 1$, perché

$\longrightarrow \forall \omega \in A$, $\int_{\Theta} g(\theta,\omega)\, d\nu(\theta) \overset{(g\equiv 1)}{=} \nu(\Theta) = 1$ ✓

$\longrightarrow \forall \omega \notin A$, $\int_{\Theta} \dots = \int_{\Theta} \frac{L^{\theta}(\omega)}{\int_{\Theta} L^{\tau}(\omega)\, d\nu(\tau)}\, d\nu(\theta) = 1$, ovviamente. ✓

(d) Veniamo finalmente a $N(\omega,T) \doteq \int_T g(\theta,\omega)\, d\nu(\theta)$, $T \in \mathcal{T}$ e $\omega \in \Omega(\setminus M)$. Allora, $\forall T \in \mathcal{T}$,

$\underset{(\in\,\Omega)}{\omega} \mapsto \int_T g(\theta,\omega)\, d\nu(\theta)$ è indubbiamente $\mathcal{A}$-mis. per Fubini classico, ed inoltre $\forall \omega \in \Omega$

$\underset{(\in\,\mathcal{T})}{T} \mapsto \int_T g(\theta,\omega)\, d\nu(\theta)$ è una probabilità su $\mathcal{T}$

(per il punto (c)), e pertanto N è effettivamente un nucleo da $(\Omega_{\omega}, \mathcal{A})$ su $(\Theta, \mathcal{T}_{\tau})$. ✓

$\boxed{\text{NOTA}}$ $\forall \omega \in \Omega$, $N(\omega,T) = (g(\cdot,\omega) \cdot \nu(\cdot))[T]$

La tesi è che tale N funzioni da legge condizionale a posteriori, cioè che (per definizione)

$$\boxed{\mathbb{E}^{Q}[X \mid \mathcal{Y}](\omega) \overset{(Q\text{-q.c.})}{=} \int_{\Theta} X(\theta,\omega)\, \underbrace{N(\omega, d\theta)}_{(= g(\theta,\omega)\,d\nu(\theta)\,!)}} \; :$$

$((\theta,\omega))$

$(\Theta \times B \mid B \in \mathcal{Y})$

($X(\theta,\omega)$ $\mathcal{G}\otimes\mathcal{Y}$-mis. e limitata)

visto che $\omega \mapsto \int_{\Theta} X(\theta,\omega) g(\theta,\omega)\, d\nu(\theta)$ è davvero $\mathcal{Y}$-mis. (sempre per Fubini), la tesi insomma è che, $\forall B \in \mathcal{Y}$,

$$\iint_{\Theta \times B} X(\theta,\omega)\, Q(d\theta, d\omega) = \iint_{\Theta \times B} \left(\int_{\Theta} X(\vartheta,\omega)\, g(\vartheta,\omega)\, d\nu(\vartheta) \right)(\omega)\, Q(d\theta, d\omega)$$

Ma, infatti: $\forall B \in \mathcal{Y}$,

$$\cancel{\Rightarrow \iint_{\Theta \times B} X(\theta,\omega)\, Q(d\theta,d\omega) \overset{(\text{Fubini gen.})}{=} \int_{\Theta} d\nu(\theta) \int_{B} X(\theta,\omega)\, dP^{\theta}(\omega)} =$$

$$= \cancel{\int_{\Theta} d\nu(\theta) \int_{B} X(\theta,\omega)\, L^{\theta}(\omega)\, d\mu(\omega)}$$

$\forall B \in \mathcal{Y}, \quad \iint_{\Theta \times B} X(\theta,\omega)\, Q(d\theta, d\omega) = \Big(\text{(cont (Q) !)} \Big)$

$$= \iint_{\Theta \times (B \cap A^c)} X(\theta,\omega)\, Q(d\theta, d\omega) \overset{\text{(Fu.)}}{=} \int_{\Theta} d\nu(\theta) \int_{B \cap A^c} X(\theta,\omega)\, dP^{\theta}(\omega) =$$

$$= \boxed{\int_{\Theta} d\nu(\theta) \int_{B \cap A^c} X(\theta,\omega)\, \mathcal{L}^{\theta}(\omega)\, d\mu(\omega)} \; . \quad \checkmark$$

→ D'autre part, $\displaystyle \iint_{\Theta \times B} \left(\int_{\Theta} X(\vartheta,\omega)\, g(\vartheta,\omega)\, d\nu(\vartheta) \right)(\omega)\, Q(d\theta, d\omega) \overset{(e)}{=}$

$$= \iint_{\Theta \times (B \cap A^c)} \left(\int_{\Theta} X(\vartheta,\omega)\, g(\vartheta,\omega)\, d\nu(\vartheta) \right)(\omega)\, Q(d\theta, d\omega) = \text{(Fub.)}$$

$$= \int_{\Theta} d\nu(\theta) \int_{B \cap A^c} \left(\int_{\Theta} X(\vartheta,\omega)\, \boxed{g(\vartheta,\omega)}\, d\nu(\vartheta) \right)(\omega)\, \mathcal{L}^{\theta}(\omega)\, d\mu(\omega) =$$

$\left(\omega \in B \setminus A \Rightarrow g(\vartheta,\omega) = \dfrac{\mathcal{L}^{\vartheta}(\omega)}{\int_{\Theta} \mathcal{L}^{\tau}(\omega)\, d\nu(\tau)} \right)$

$$= \int_{\Theta} d\nu(\theta) \int_{B \cap A^c} \left(\int_{\Theta} X(s,\omega) \frac{\mathcal{L}^{s}(\omega)}{\int_{\Theta} \mathcal{L}^{\tau}(\omega)\, d\nu(\tau)}\, d\nu(s) \right)(\omega)\, \mathcal{L}^{\theta}(\omega)\, d\mu(\omega)$$

$=$

$$= \int_{\Theta} d\nu(s) \int_{B\cap A^c} \left(\frac{1}{\int_{\Theta} L^{\tau}(\omega)\, d\nu(\tau)} \cdot \int_{\Theta} X(s,\omega) L^{\sigma}(\omega)\, d\nu(\sigma) \right)(\omega)\, L^{\theta}(\omega)\, d\mu(\omega) =$$

FUBINI (scambio il 1° col 3°)

$$= \int_{\Theta} d\nu(s) \int_{B\cap A^c} \left(\underbrace{\frac{1}{\int_{\Theta} L^{\tau}(\omega)\, d\nu(\tau)} \cdot \int_{\Theta} L^{\theta}(\omega)\, d\nu(\theta)}_{1} \right)(\omega)\, X(s,\omega) L^{s}(\omega)\, d\mu(\omega)$$

$$= \int_{\Theta} d\nu(s) \int_{B\cap A^c} X(s,\omega) L^{s}(\omega)\, d\mu(\omega) \quad \text{, OK. } \square$$

FORMALISMO DECISIONALE & DECISIONE BAYESIANA.

L'idea iniziale è quella di estendere quanto riguarda LA STIMA:

→ ricordiamo l'idea base della teoria della stima: abbiamo convenuto che un certo fenomeno "funzioni matematicamente" come $(\Omega, \mathcal{F}, P)$, per cui noi

Lo abbiamo come $(\Omega, \mathcal{F}, \{P^{\vartheta} | \vartheta \in \Theta\})$, Θ effettivo, dove $P = P^{\vartheta^*}$ va trovata. Allora l'idea è definire una topologia di v.a. $U(\omega): \Omega \to D \subseteq \mathbb{R}$ aperto che "stimino" probabilmente una funzione di $\vartheta \in \Theta$, cioè $g(\vartheta): \Theta \to D$, con l'idea di ottenere un risultato tipo:

→ se $\omega^* \in \Omega$ è l'esito dell'esperimento, allora "probabilmente" $U(\omega^*) \approx g(\vartheta^*)$. In altre parole l'idea è quella di prendere $U(\omega^*)$ come $g(\vartheta^*)$. Che errore commettiamo? Commettiamo

$$C(\vartheta^*, U(\omega^*)) \overset{\text{AD ES.}}{=} |U(\omega^*) - g(\vartheta^*)|^2.$$

In media, commettiamo

$$E^{P^{\vartheta^*}}\left[|U(\omega) - g(\vartheta^*)|^2\right].$$

Rispetto ad ogni $\vartheta \in \Theta$, commettiamo $R(\vartheta) = E^{P^{\vartheta}}\left[|U(\omega) - g(\vartheta)|^2\right]$.

Ci piacerebbe avere dunque $R_U(\vartheta)$ minimo possibile, il che dipende però ovviamente [illegible].

Più
In generale, in seguito all'osservazione dell'esito ω^*, potremo voler compiere un'azione diversa dalla stima ϑ^* "giusta" di $U(\omega^*)$ (fra. evento preliminarmente nella U), e per poi calcolare il costo, $\forall \vartheta \in \Theta$, $C(\vartheta, U(\omega^*))$: Più in generale, potremo voler $\boxed{\text{COMPIERE UN'AZIONE } a \in A \text{ e CALCOLARE } C(\vartheta, a), \forall \vartheta \in \Theta}$.

→ Una REGOLA DECISIONALE è una funzione

$\delta(\omega): \Omega \to A$ $(\delta(\omega) = a)$, A insieme $\neq \emptyset$ qualsiasi,

che porta $\omega \mapsto C(\vartheta, \delta(\omega))$ $\forall \vartheta \in \Theta$,

$C(\vartheta, a): \Theta \times A \to [0, +\infty)$ funzione costo scelta.

(ES Uno stimatore con $A = D$ è una regola decisionale!)

Supponiamo di lavorare con insieme

"delle azioni" A che sia misurabile, sia $(A, \mathcal{A})$ mis.,

e supporremo quando che $\delta(\omega) : \Omega \to A$ sia mis. e

che $\omega \mapsto C(\theta, \delta(\omega))$ sia mis. (per ogni $\theta \in \Theta$),

ANZI: che $\begin{array}{c} A \longrightarrow [0,+\infty) \\ a \mapsto C(\theta, a) \end{array}$ sia mis. $\forall \theta \in \Theta$.

Allora definiremo il RISCHIO della regola $\delta(\omega)$

$$R_\delta(\theta) \doteq \mathbb{E}^{P^\theta}\left[C(\theta, \delta(\cdot)) \right] \quad , \forall \theta \in \Theta .$$

Confrontiamo le decisioni $\delta(\omega)$ in base al minor rischio in modo ~~del tutto~~ quasi analogo e come sempre detto, nel caso: adesso stiamo lavorando con un modello bayesiano

$$\left(\Omega, \mathcal{A}, \{ P^\theta \mid \theta \in (\Theta, \mathcal{T}, \nu) \} \right),$$

($\theta \mapsto P^\theta[A]$ $\mathcal{T}$-mis. per ogni $A \in \mathcal{A}$)

LEGGE A PRIORI

dunque l'idea è minimizzare $R_\delta(\theta)$, $\theta \in \Theta$, NON necessariamente θ per θ, ma in ν-media:

dovremo minimizzare

$$\boxed{\int_{\Theta} R_\delta(\theta)\, d\pi(\theta) = \int_{\Theta} E^{p_\theta}[C(\theta, \delta(\cdot))]\, d\pi(\theta)}$$, nell'ipotesi:

naturalmente che $\theta \mapsto R_\delta(\theta)$ sia $\mathcal{T}$-mis..

$\boxed{\text{IDEA}}$ più precise nel contesto bayesiano: date ρ prob su $(\Theta, \mathcal{T})$, abbiamo una funzione costo $C(\theta, a): \Theta \times A \to [0, +\infty)$ misurabile, e avremmo "minimizzare in tutti i sensi"!

Allora, considerando le probabilità medie come $\int_{\Theta} C(\theta, a)\, d\rho(\theta)$ $\forall a \in A$, scelte a ρ, dovremo

$$\boxed{\inf_{a \in A} \int_{\Theta} C(\theta, a)\, d\rho(\theta)} \xrightarrow{\text{(def.)}} \boxed{\text{IL RISCHIO BAYESIANO (relativo a } \rho)}.$$

L'ideale, invece, è che la situazione sia tale per cui esiste $a^* \in A$ tale che

$$\int_{\Theta} C(\theta, a^*)\, d\rho(\theta) = \inf_{a \in A} \int_{\Theta} C(\theta, a)\, d\rho(\theta)$$

(costo medio / errore / perdita media dell'azione a^* → il minimo possibile!),

cioè avere $\delta(\omega) : \Omega \to A$ tale che $\delta(\omega) \approx a^*$.

→ a^* è $\forall A$ DECISIONE BAYESIANA relativa a ρ.

Nel caso ciò accade, potrebbe definire

$\rho \longmapsto a^*$, chiamandola

$d : \rho \longmapsto d(\rho) = a^*$ che minimizza $a \mapsto \int_{\Theta} C(\theta, a)\, d\rho(\theta)$.

— — — — — —

Ebbene, il meccanismo decisionale bayesiano cioè lo possiamo immaginare, ovvero: certo nella scelta $\delta(\omega) \doteq d(\underbrace{N^{\omega}[\cdot]}_{(\rho^{\cdot})})$ se

$N^{\omega}[T]$, $T \in \mathcal{T}$, è la legge a posteriori su $(\Theta, \mathcal{T})$ condizionale a ω. ☺

$\boxed{\text{RIASSUNTO}}$ Se $(\Omega, \mathcal{F}, \{P^{\theta} \mid \theta \in (\Theta, \mathcal{T}, \overset{\text{a priori}}{\nu})\}, \mu))$ è un modello stat. (dominato) bayesiano che ammette

$(\Omega_{\omega}, \mathcal{F}) \overset{\nu^{\omega}(T)}{\dashrightarrow} (\Theta, \mathcal{T}_{\tau})$ legge a posteriori condizionale

(per es., $\nu^{\omega}(T) = \int_T g(\theta, \omega)\, d\nu(\theta)$ $\forall T \in \mathcal{T}, \forall \omega \in \Omega$, come visto ultimamente), allora in seguito all'esito ω^* dobbiamo intraprendere un'azione a in A, con costo $C(\theta, a)$ per ogni $\theta \in \Theta$: scegliamo una regola "ottimale" $\delta(\omega): \Omega \to A$, con costo $\omega \mapsto C(\theta, \delta(\omega))$ per ogni $\theta \in \Theta$. Allora, $\forall a \in A$, il costo medio è $\int_{\Theta} C(\theta, a)\, d\rho(\theta)$, e

ρ prob. su $(\Theta, \mathcal{G})$ (fissata): sappiamo che $\exists!\, e^* \in A$

tale che $\int_\Theta C(\vartheta, e^*)\, d\rho(\vartheta) = \inf_{e \in A} \int_\Theta C(\vartheta, e)\, d\rho(\vartheta)$,
($\max_{e\in A}$)

con il valore di "fare" "$\delta(\omega) \simeq e^*$". Per avere "$\delta(\omega)$ preciso", è naturale pensare a porre

$$\delta(\omega) \doteq d(\nu^\omega) \text{, se } d : \rho \mapsto e^* = d(\rho) \text{ :}$$

cioè $\boxed{\delta : \omega \mapsto \text{ nell' "}e^*\text{" relativo a } \nu^\omega}$! $\checkmark$ (→ dovrebbe essere il δ "migliore"...)

~~Tutto questo che senso:~~

$\boxed{\text{Teorema.}}$ Nella precedente situazione, supponiamo che sia tutto OK: dunque, ad esempio, che esista(!) $d(\nu^\omega)$ $\forall \omega \in \Omega$, e che $\delta : \omega \mapsto d(\nu^\omega)$ sia $\mathcal{F}$-misurabile. Allora, $\forall$ regola δ,

$$\boxed{\int_\Theta R(\vartheta, \delta(\cdot))\, d\nu(\vartheta) \geq \int_\Theta R(\vartheta, \delta^*(\cdot))\, d\nu(\vartheta).}$$

[Dim. Conseguenza del teorema visto all'inizio (per cui $L^\rho(\omega)$ è $\mathcal{G}\otimes\mathcal{F}$-mis., ecc. ....)

Dunque, se $A \doteq \{\omega \in \Omega \mid \int_{\Theta} \mathcal{L}^{\vartheta}(\omega)\, d\nu(\vartheta) = 0\}$ $(\in \mathcal{S})$,

allora $Q[\Theta \times A] = 0$. / Calcoliamo:

$$\int_{\Theta} R_{\delta}(\vartheta)\, d\nu(\vartheta) \overset{\text{def.}}{=} \int_{\Theta} \mathbb{E}^{P^{\vartheta}}[C(\vartheta, \delta(\cdot))]\, d\nu(\vartheta) \overset{\text{(OVVIO)}}{=}$$

$$= \int_{\Theta} d\nu(\vartheta) \int_{\Omega} C(\vartheta, \delta(\omega))\, dP^{\vartheta}(\omega) = \text{(FUBINI (GEN))}$$

$$= \iint_{\Theta \times \Omega} C(\vartheta, \delta(\omega))\, Q(d\vartheta, d\omega) \overset{Q[\Theta \times A] = 0}{=} \iint_{\Theta \times A^{c}} C(\vartheta, \delta(\omega))\, Q(d\vartheta, d\omega) =$$

$$\overset{\text{(FUBINI)}}{=} \int_{\Theta} d\nu(\vartheta) \underbrace{\int_{A^{c}} C(\vartheta, \delta(\omega))\, dP^{\vartheta}(\omega)}_{= \int_{A^{c}} C(\vartheta, \delta(\omega)) \mathcal{L}^{\vartheta}(\omega)\, d\mu(\omega)} \quad \text{(FUBINI "NORMALE")}$$

$$= \int_{A^{c}} d\mu(\omega) \int_{\Theta} C(\vartheta, \delta(\omega))\, \mathcal{L}^{\vartheta}(\omega)\, d\nu(\vartheta) \overset{\text{(OVVIO)}}{=}$$

$$= \int_{A^{c}} d\mu(\omega) \left(\int_{\Theta} \mathcal{L}^{\tau}(\omega)\, d\nu(\tau)\right)(\omega) \int_{\Theta} C(\vartheta, \delta(\omega)) \cdot \underbrace{\frac{\mathcal{L}^{\vartheta}(\omega)}{\int_{\Theta} \mathcal{L}^{\tau}(\omega)\, d\nu(\tau)}\, d\nu(\vartheta)}_{= d\nu^{\omega}(\vartheta),\ \text{(essendo } \omega \notin A!!)}$$

$$= \int_{A^{c}} d\mu(\omega) \left(\int_{\Theta} \mathcal{L}^{\tau}(\omega)\, d\nu(\tau)\right)(\omega) \int_{\Theta} C(\vartheta, \delta(\omega))\, d\nu^{\omega}(\vartheta)$$

Adesso, per definizione di $d(\nu^\omega) = \delta^*(\omega)$, vale

$$\boxed{\int_\Theta \mathcal{C}(\vartheta, \delta(\omega))\, d\nu^\omega(\vartheta) \geq \int_\Theta \mathcal{C}(\vartheta, \delta^*(\omega))\, d\nu^\omega(\vartheta) \quad \forall \omega \in \Omega},$$

da cui in definitiva $\int_\Theta R_\delta(\vartheta)\, d\nu(\vartheta) \geq$

$$\geq \int_{A^c} d\mu(\omega) \left(\int_\Theta \mathcal{L}^\tau(\omega)\, d\nu(\tau) \right)(\omega) \int_\Theta \mathcal{C}(\vartheta, \delta^*(\omega))\, d\nu^\omega(\vartheta),$$

ed ora è OVVIO che tale RHS sia =

$= \int_\Theta R_{\delta^*}(\vartheta)\, d\nu(\vartheta)$, ripercorrendo a ritroso. $\square$

OSS Se, nella situazione prec., $T : \Omega \to (E, \mathcal{E})$ è fine sufficiente, o cioè (Neyman-Fisher) $\mathcal{L}^\vartheta(\omega) = g^\vartheta(T(\omega))\, h(\omega)$,

allora $\forall \omega \in \Omega \setminus A$ "$\frac{d\nu^\omega}{d\nu}(\vartheta)$" $= g(\vartheta, \omega) =$

$$= \frac{\mathcal{L}^\vartheta(\omega)}{\int_\Theta \mathcal{L}^\tau(\omega)\, d\nu(\tau)} = \frac{g^\vartheta(T(\omega))}{\int_\Theta g^\tau(T(\omega))\, d\nu(\tau)},$$ e

per ciò ν^ω ha densità risp. a ν T-misurabile.

$\boxed{\text{STIMATORI BAYESIANI.}}$ Continuando dall'contesto precedente, supponiamo stimare $g(\theta): \Theta \to \mathbb{R}$ mis. limitata, con $C(\theta, a) \doteq |g(\theta) - a|^2$, $\theta \in \Theta$ e $a \in A \doteq \mathbb{R}$:

dunque, come visto, ci domandiamo anzitutto se esiste "unico"

$$\min_{a \in \mathbb{R}} \int_\Theta |g(\theta) - a|^2 \, d\rho(\theta),$$

quale che sia ρ probabilità su $(\Theta, \mathcal{T})$.

$\boxed{\text{Lemma}}$ Se $\int_\Theta g^2(\theta) \, d\rho(\theta) < +\infty$,

allora $\boxed{a^* \doteq d(\rho) \doteq \int_\Theta g(\theta) \, d\rho(\theta)}$ è tale che

$$\int_\Theta |g(\theta) - a|^2 \, d\rho(\theta) \geq \int_\Theta |g(\theta) - a^*|^2 \, d\rho(\theta).$$

Dim. $\int_\Theta |g(\theta) - a|^2 d\rho(\theta) \underset{(|(g(\theta)-a^*)+(a^*-a)|^2)}{=} (\text{OK})$ $\quad \pm a^*$

$$= \int_\Theta |g(\theta) - a^*|^2 + (a^* - a)^2 + 2(a^* - a) \underbrace{\int_\Theta (g(\theta) - a^*) \, d\rho(\theta)}_{(=0 \text{ def. } a^*)}$$

$$= \int_\Theta |g(\theta) - a^*|^2 + \underbrace{(a^* - a)^2}_{\geq 0}. \quad \square$$

$\Rightarrow$ Lo stimatore bayesiano di $\varphi(\vartheta): \Theta \to \mathbb{R}$ è

$\delta^*(\omega) = d(\nu^{\omega}) : \Omega \to \mathbb{R}$, dato da

$$\boxed{\delta^*(\omega) = \int_{\Theta} \varphi(\vartheta)\, d\nu^{\omega}(\vartheta)} = \text{(TEOREMA FONDAMENTALE)}$$

$$= \begin{cases} \int_{\Theta} \varphi(\vartheta)\, d\nu(\vartheta) & \text{se } \omega \in A \\ \dots & \\ \int_{\Theta} \dfrac{\varphi(\vartheta) L^{\vartheta}(\omega)}{\int_{\Theta} L^{\tau}(\omega)\, d\nu(\tau)}\, d\nu(\vartheta) & \text{se } \omega \notin A \end{cases}$$

$\left(\omega \in A \overset{(\text{def.})}{\Longleftrightarrow} \int_{\Theta} L^{\tau}(\omega)\, d\nu(\tau) = 0\right)$. Chiamando

$T(\omega) \doteq \int_{\Theta} \varphi(\vartheta)\, d\nu^{\omega}(\vartheta)$. Allora

$$\boxed{T(\omega) \overset{(Q\text{-q.c.})}{=} \mathbb{E}^{Q}[\varphi(\vartheta) \mid \mathcal{F}](\omega)}$$ (: infatti, come

visto, $\mathbb{E}^{Q}[\varphi(\vartheta) \mid \mathcal{F}](\omega) \overset{(Q\text{-q.c.})}{=} \int_{\Theta} \varphi(\vartheta)\, \underbrace{d\nu^{\omega}(\vartheta)}_{\text{"}N(\omega, d\vartheta)\text{"}} = T(\omega)$)

Dunque T è una speranza condizionale.

$\boxed{\text{Teorema.}}$ Se lo stimatore bayesiano $T(\omega)=\int_{\Theta} g(\vartheta)\,d\pi^{\omega}(\vartheta)$ di $g(\vartheta)$ $\left(\text{avente } \int_{\Theta} g(\vartheta)^2\,d\pi(\vartheta) < +\infty\right)$ e' corretto $\left(\text{cioè } \mathbb{E}^{P_\vartheta}[T]=g(\vartheta)\ \forall \vartheta\in\Theta\right)$, allora vale

$$\iint_{\Theta\times\Omega} |T(\omega)-g(\vartheta)|^2\, Q(d\vartheta,d\omega)=0.$$

Dim. $X(\vartheta,\omega) \doteq T(\omega)-g(\vartheta)$ sta in $L^2(Q)$ ed è tale che

$$\mathbb{E}^{Q}[X\mid\tilde{\mathcal{F}}] = \underbrace{\mathbb{E}^{Q}[T\mid\tilde{\mathcal{F}}]}_{\substack{\parallel\ (\text{perché } T \text{ è } \tilde{\mathcal{F}}\text{-mis.})\\ T}} - \underbrace{\mathbb{E}^{Q}[g\mid\tilde{\mathcal{F}}]}_{\substack{\parallel\ (\text{visto ora})\\ T}} = 0,$$

dunque per ogni U che sia $\tilde{\mathcal{F}}$-misurabile vale

$\mathbb{E}^{Q}[XU]=0$: infatti,

$$\mathbb{E}^{Q}[XU] = \mathbb{E}^{Q}[\mathbb{E}^{Q}[XU\mid\tilde{\mathcal{F}}]] \underset{(\tilde{\mathcal{F}}\text{-mis.})}{=} \mathbb{E}^{Q}[U\,\underbrace{\mathbb{E}^{Q}[X\mid\tilde{\mathcal{F}}]}_{=0}]$$

$=0$. ✓

Allora, vista la def. di X, basta ottenere che anche $\boxed{\mathbb{E}^{Q}[X\mid\tilde{\mathcal{G}}]=0}$ per avere che X

l'ortogonale anche a sé stesso, che è la tesi! ✓.

Ma, infatti, $\mathbb{E}^{Q}[X|\tilde{\mathcal{G}}] \overset{\text{TEO.}}{=} \int_{\Omega} \underbrace{X(\vartheta,\omega)}_{(T(\omega)-g(\vartheta))} dP^{\vartheta}(\omega) =$

$= \mathbb{E}^{P^{\vartheta}}[T] - g(\vartheta) = 0$ per correttezza. $\square$)

$\boxed{\text{ESERCIZIO.}}$ Se le P^{ϑ} sono mutuamente equivalenti (modello regolare), allora T può ~~bastare~~ NON essere corretto:

se lo fosse, $0 = \iint_{\Theta\times\Omega} |T(\omega)-g(\vartheta)|^2\, Q(d\vartheta,d\omega) =$

$= \int_{\Theta} d\nu(\vartheta) \underbrace{\int_{\Omega} |T(\omega)-g(\vartheta)|^2\, dP^{\vartheta}(\omega)}$, per cui bisogna

$N \in \mathcal{G}$ con $\nu(N)=0$ tale che, $\forall \vartheta \in \Theta \setminus N$,

$\int_{\Omega} |T(\omega)-g(\vartheta)|^2\, dP^{\vartheta}(\omega) = 0$, ossia tale che

$\forall \vartheta \notin N \quad P^{\vartheta}[\underbrace{\{\omega\in\Omega \mid T(\omega)\neq g(\vartheta)\}}_{=:B_{\vartheta}}] = 0$. Per ciò, se

$\longrightarrow \#(\Theta\setminus N) = \# N^c \geq 2$,

$\longrightarrow g(\vartheta)$ NON è costante su $\Theta\setminus N = N^c$, allora ciò che mi ecciudo. Infatti,

se fissiamo $\vartheta_0 \neq \vartheta_1 \in \Theta \setminus N$ tali che $g(\vartheta_0) \neq g(\vartheta_1)$.

Allora $P^{\vartheta_0}[B_{\vartheta_1}^c] =$

$= P^{\vartheta_0}[\{\omega \in \Omega \mid T(\omega) \neq g(\vartheta_1)\}] \leq$ (se c'è ... in $\mathcal{T}$)

$$\leq P^{\vartheta_0}\Big[\bigcup_{\substack{\vartheta \in \Theta \\ \text{t.c. } g(\vartheta) \neq g(\vartheta_0)}} \{\omega \in \Omega \mid T(\omega) = g(\vartheta)\}\Big] \overset{\text{Giusto}}{=}$$

$= P^{\vartheta_0}[B_{\vartheta_0}^c] = 0$, da cui

$\boxed{P^{\vartheta_0}[B_{\vartheta_1}^c] = 0}$. Ma allora, essendo $P^{\vartheta_0} \sim P^{\vartheta_1}$,

anche $P^{\vartheta_1}[B_{\vartheta_1}^c] = 0$: assurdo; perché c'è anche

$P^{\vartheta_1}[B_{\vartheta_1}] = 0$! ✓

Insomma uno stimatore consistente sarà ben difficilmente corretto... Magari però nemmeno troppo non-corretto! Vediamo meglio.

ES. 1 $(X_1,\dots,X_n)$ campione di legge ~~$B(1,\theta)$, $\theta\in(0,1)$,~~ ~~per~~ cui ~~$L^\theta(x_1,\dots,x_n)=\theta^{\sum_{i=1}^n x_i}(1-\theta)^{\sum_{i=1}^n x_i - n}$, $x_i\in\{0,1\}$~~

geometrica μ^θ di parametro $\theta\in(0,1)$, ossia μ^θ concentrata su $\mathbb{N}$ e data da $\mu^\theta[\{k\}]=\theta(1-\theta)^{k-1}$, $\forall k\in\mathbb{N}_*$. $\left(\sum_{k\geq 1}(1-\theta)^{k-1}=\sum_{m\geq 0}(1-\theta)^m=\frac{1}{1-(1-\theta)}=\frac{1}{\theta}\right)$, $(1-\theta)\in(0,1)$

$\rightarrow$ Modello regolare $(\Omega,\mathcal{A},\{P^\theta\mid\theta\in(\Theta,\mathcal{T},\nu)\},\mu)=$

$=\left((\mathbb{N}_*)^n,\ \mathcal{P}(\mathbb{N}_*^n),\ \{(\mu^\theta)^{\otimes n}\mid\theta\in((0,1),\mathcal{B}(0,1)),\nu\},\ \mu\right)$

ν: di Lebesgue su $\mathcal{B}(0,1)$ $\quad$ μ: conta i punti di $\mathbb{N}_*^n$

$\mu^\theta(x)=\theta(1-\theta)^{x-1}$,

$\rightsquigarrow$ $\boxed{L^\theta(x_1,\dots,x_n)=\theta^n(1-\theta)^{\sum_{i=1}^n x_i - n}}$ per ogni $\theta\in(0,1)$, $x_i\in\mathbb{N}_*$. Allora $(\theta;(x_1,\dots,x_n))\mapsto L^\theta(x_1,\dots,x_n)$, $\Theta\times\Omega\to[0,1]$ è $\mathcal{T}\otimes\mathcal{A}$-misurabile (ed inoltre $\sum_{i=1}^n X_i$ è una statistica esaustiva, se $X_i(x_1,\dots,x_n)=x_i$), ed in particolare il modello è bayesiano. Volendo uno stimatore bayesiano di $g(\theta)=\theta:\Theta=(0,1)\to(0,1)$, visto che $L^\theta(x_1,\dots,x_n)$ è sempre >0 ed che quindi $\int_0^1 \vartheta^n (1-\vartheta)^{\sum_{i=1}^n x_i - n}\, d\vartheta \neq 0 \quad \forall x_1, \dots, x_n$ in $\mathbb{N}_*$, finalmente

$$T(x_1,\dots,x_n) = \int_0^1 \vartheta \frac{\mathcal{L}^\vartheta(x_1,\dots,x_n)}{\int_0^1 \mathcal{L}^\tau(x_1,\dots,x_n)\,d\tau}\, d\vartheta =$$

$$= \int_0^1 \frac{\vartheta^{n+1}(1-\vartheta)^{\sum_{i=1}^n x_i - n}\, d\vartheta}{\int_0^1 \tau^n (1-\tau)^{\sum_{i=1}^n x_i - n}\, d\tau}.$$

Adesso, diciamo per OK la formula

$$\int_0^1 \vartheta^{\alpha-1}(1-\vartheta)^{\beta-1}\, d\vartheta = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \alpha,\beta>0.$$

Allora ne deduciamo $(\forall N, M \in \mathbb{N})$ $\int_0^1 \vartheta^N (1-\vartheta)^M\, d\vartheta =$ $\begin{pmatrix} N=(N+1)-1 \\ M=(M+1)-1 \end{pmatrix}$

$$= \frac{\Gamma(N+1)\Gamma(M+1)}{\Gamma(N+M+2)} = \frac{N!\,M!}{(N+M+1)!},$$

da cui $$T(x_1,\dots,x_n) = \frac{\int_0^1 \vartheta^{n+1}(1-\vartheta)^{\sum_{i=1}^n x_i - n}\, d\vartheta}{\int_0^1 \vartheta^n (1-\vartheta)^{\sum_{i=1}^n x_i - n}\, d\vartheta} =$$

$$= \frac{\dfrac{(n+1)!\,\left(\sum_{i=1}^n x_i - n\right)!}{\left(\sum_{i=1}^n x_i + 2\right)!}}{\dfrac{n!\,\left(\sum_{i=1}^n x_i - n\right)!}{\left(\sum_{i=1}^n x_i + 1\right)!}} = \frac{n+1}{\sum_{i=1}^n x_i + 2},$$

ovvero $\boxed{T = \dfrac{n+1}{\sum_{i=1}^{n} X_i + 2}}$. ✓ $\boxed{\text{NOTA}}$ Esiste stimatore di $\sim T\ldots$

Stimatore verosimiglianza di ϑ, ed è $\hat{\vartheta} = \dfrac{n}{\sum_{i=1}^{n} X_i} \left(= \dfrac{1}{\overline{X}} \right)$. ✓

$$\mathcal{L}^{\vartheta}(\omega) = \vartheta^{n} (1-\vartheta)^{\sum_{i=1}^{n} X_i(\omega) - n} =$$

$$= \exp\Big(\underbrace{\log\big(\vartheta^{n} (1-\vartheta)^{\sum X_i - n}\big)}_{= n\log(\vartheta) + (\sum X_i - n)\log(1-\vartheta)} \Big)$$

Ma l'equazione con $\log \mathcal{L}^{\vartheta} = n\log(\vartheta) + \left(\sum_{i=1}^{n} X_i - n\right)\log(1-\vartheta)$

$$0 = \frac{\partial}{\partial\vartheta} \log \mathcal{L}^{\vartheta} = \frac{n}{\vartheta} - \frac{\sum_{i=1}^{n} X_i - n}{1-\vartheta}, \text{ cioè}$$

$$\frac{1}{\vartheta} - 1 = \frac{1}{n}\sum_{i=1}^{n} X_i - 1 .$$

$\boxed{\text{ES. 2}}$ $(X_1, \ldots, X_n)$ campione di legge $\mu^{\vartheta} = \mathcal{E}(\vartheta)$, $\vartheta \in (0,+\infty)$,

per cui mettiamoci in $\left(\Omega, \mathcal{A}, \{P^{\vartheta} \mid \vartheta \in \Theta\}, \mathcal{T}, \nu\}, \mu\right) =$

$= \left((0,+\infty)^{n}, \mathcal{B}(0,+\infty)^{n}, \{\mu^{\vartheta, \otimes n} \mid \vartheta \in (0,+\infty)\}, \mathcal{B}(0,+\infty), \nu\}, \mu\right)$

di Lebesgue n-dim. su $\mathcal{B}(0,+\infty)^{n}$

$d\nu(\vartheta) \doteq a\, e^{-a\vartheta}\, d\vartheta$, $a > 0$ dato (l'un'altra esponenziale!),

$X_i(x_1, \ldots, x_n) = x_i$.

Allora $\mathcal{L}^{\vartheta}(x_1,\dots,x_n) = \vartheta^n e^{-\vartheta \sum_{i=1}^{n} x_i}$ (misurabile e >0)

$\sum_{i=1}^{n} X_i(x_1,\dots,x_n)$

Lo stimatore bayesiano di $g(\vartheta)=\vartheta$ è

$$T(\cdot) = \frac{\int_{\Theta} g(\vartheta)\mathcal{L}^{\vartheta}(\cdot)\, d\nu(\vartheta)}{\int_{\Theta} \mathcal{L}^{\vartheta}(\cdot)\, d\nu(\vartheta)} = \quad (d\nu(\vartheta) = a e^{-a\vartheta} d\vartheta)$$

$$= \frac{\int_0^{+\infty} \vartheta^{n+1} e^{-\vartheta\left(\sum_{i=1}^{n} x_i + a\right)} d\vartheta}{\int_0^{+\infty} \vartheta^{n} e^{-\vartheta\left(\sum_{i=1}^{n} x_i + a\right)} d\vartheta}.$$

"Ricordiamo" che $\boxed{\int_0^{+\infty} \vartheta^N e^{-\vartheta b} d\vartheta = \frac{N!}{b^{N+1}}}$

$$\left[\begin{cases} \int_0^{+\infty} e^{-\vartheta b} d\vartheta = \left[\frac{1}{b} e^{-\vartheta b}\right]_{\vartheta=+\infty}^{\vartheta=0} = \frac{1}{b} \quad \text{ok} \\ \int_0^{+\infty} \vartheta e^{-\vartheta b} d\vartheta = \left[\frac{1}{b}\vartheta e^{-\vartheta b}\right]_{+\infty}^{0} + \frac{1}{b}\int_0^{+\infty} e^{-\vartheta b} d\vartheta = \frac{1}{b^2} \quad \text{ok} \end{cases}\right. \qquad (0! = 1! = 1)$$

induzione

$$\int_0^{+\infty} \vartheta^{N+1} e^{-\vartheta b} d\vartheta = \left[\frac{1}{b}\vartheta^{N+1} e^{-\vartheta b}\right]_{+\infty}^{0} + \frac{(N+1)}{b}\underbrace{\int_0^{+\infty} \vartheta^N e^{-\vartheta b} d\vartheta}_{\frac{N!}{b^{N+1}}} =$$

$$= \frac{(N+1)!}{b^{N+2}} \quad \text{ok.} \Big]$$

Per ciò otteniamo immediatamente

$$T = \frac{(n+1)! \Big/ \left(\sum x_i + a\right)^{n+2}}{n! \Big/ \left(\sum x_i + a\right)^{n+1}} = \frac{n+1}{\sum_{i=1}^{n} x_i + a}, \text{ di}$$

messo non molto diverso dallo stimatore di max. verosimiglianza

$$\hat{\vartheta} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

[→ NOTA $\log L^\vartheta =$

$$= n \log \vartheta - \vartheta \sum_{i=1}^{n} x_i \rightarrow \frac{\partial}{\partial \vartheta} \log L^\vartheta =$$

$$= \frac{n}{\vartheta} - \sum_{i=1}^{n} x_i. \quad \checkmark \,]$$

Torniamo al concetto di MEDIANA di una legge ν su $\mathcal{B}(\mathbb{R})$: uno di questi scalari $m \in \mathbb{R}$ tali che

$$\boxed{\nu((-\infty, m)) \leq \frac{1}{2} \leq \nu((-\infty, m])}.$$

Legato sempre agli stimatori bayesiani, si rivela utile fare il seguente ex.

$\boxed{\text{ESERCIZIO}}$ Se ν è dotata di momento secondo, e se quando

$$\begin{cases} m \rightarrow \text{una mediana di } \nu \\ \mu \rightarrow \text{la media di } \nu \\ \sigma \rightarrow \text{la deviazione standard di } \nu \end{cases}$$

, allora vale $|m - \mu| \leq \sigma$.

$\boxed{\text{Dim.}}$ Sia $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ v.a.r. avente legge ν:

Mostriamo che ~~[illegible]~~ valgono

$$\boxed{1} \qquad \inf_{a \in \mathbb{R}} \mathbb{E}^P[|X-a|] = \mathbb{E}^P[|X-m|] .$$

$$\boxed{2} \qquad \inf_{a \in \mathbb{R}} \mathbb{E}^P[|X-a|^2] = \mathbb{E}^P[|X-\mu|^2] \underset{(\text{def.})}{=} \sigma^2 .$$

Allora $\boxed{1} + \boxed{2}$ danno la tesi, perché se poniamo

$$\begin{cases} \alpha \doteq \mathbb{E}^P[(X-m)^+] \\ \beta \doteq \mathbb{E}^P[(X-m)^-] \end{cases} \qquad \left(\alpha, \beta \geq 0 \text{ tali che } \begin{cases} \alpha - \beta = \mathbb{E}^P[X-m] \\ \alpha + \beta = \mathbb{E}^P[|X-m|] \end{cases} \right).$$

allora abbiamo

$$|\mu - m| = |\mathbb{E}^P[X-m]| = |\alpha - \beta| \overset{(\text{Dis. tr.})}{\leq} \alpha + \beta =$$

$(\mathbb{E}^P[X])$ Direttamente ... = $\left(\mathbb{E}^P[|Y|] \overset{(\text{Hölder})}{\leq} \{\mathbb{E}^P[|Y|^2]\}^{1/2} \right)$

$$= \mathbb{E}^P[|X-m|] \overset{\boxed{1}}{=} \inf_{a \in \mathbb{R}} \mathbb{E}^P[|X-a|] \leq \inf_{a \in \mathbb{R}} \sqrt{\mathbb{E}^P[|X-a|^2]} =$$

$$\overset{\boxed{2}}{=} \sqrt{\sigma^2} = \sigma . \quad \checkmark \quad (\underline{\text{Bene}}.)$$

$\boxed{\text{Dim. } \boxed{1}}$ → Possiamo supporre $m = 0$, nel senso che ~~[illegible]~~ basta ~~[illegible]~~ dimostrare che $\boxed{\inf_{a \in \mathbb{R}} \mathbb{E}^P[|Y-a|] = \mathbb{E}^P[|Y|]}$ per ogni v.a.r. Y ~~[illegible]~~ ~~avente~~ ~~$\mathbb{E}^P[Y] = 0$~~ ($\because$ infatti, allora, poi

Preso $Y \doteq X - m$ ed osservo che

$$\inf_{a \in \mathbb{R}} \mathbb{E}^P[|X-m-a|] = \inf_{a' \in \mathbb{R}} \mathbb{E}^P[|X-a'|] \ \checkmark).$$

Inoltre, per Y, basta dimostrare che

$$\boxed{\inf_{a<0} \mathbb{E}^P[|Y-a|] = \mathbb{E}^P[|Y|]}$$

(perché poi

$$\inf_{a>0} \mathbb{E}^P[|Y-a|] \overset{(\text{ovvio})}{\underset{(|z|=|-z|)}{=}} \inf_{a>0} \mathbb{E}^P[|\underbrace{(-Y)-(-a)}_{(-Y+a)}|] =$$

$$= \inf_{a'<0} \mathbb{E}^P[|(-Y)-a'|] = \mathbb{E}^P[|-Y|] = \mathbb{E}^P[|Y|] \ \checkmark).$$

Dunque, vogliamo ottenere che per ogni v.a. Y vale $\mathbb{E}^P[|Y|] \le \mathbb{E}^P[|Y-a|]$ $\forall a<0$. Oss. che, $\forall a<0$ e $\forall x \in \mathbb{R}$ vale

$$|x| \le |x-a| + a\left(\mathbb{1}_{[0,+\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)\right) \quad (\text{ok}),$$

da cui

$$|Y| \le |Y-a| + a\left(\mathbb{1}_{[0,+\infty)}(Y) - \mathbb{1}_{(-\infty,0)}(Y)\right)$$

$$\Rightarrow \mathbb{E}^P[|Y|] \le \mathbb{E}^P[|Y-a|] + a\left(P[Y \ge 0] - P[Y<0]\right),$$

da cui ora (essendo $a<0$) basta dimostrare che

$P[Y\geq 0] - \underbrace{P[Y<0]}_{1-P[Y\geq 0]} \geq 0$. Ma, "infatti",

$P[X-m\geq 0] - P[X-m<0] =$

$= \underbrace{P[X\geq m]}_{\nu([m,+\infty))} - \underbrace{P[X<m]}_{P[X\geq m]-1} = 2\underbrace{P[X\geq m]}_{\geq \frac{1}{2}\ (\text{def. } m)} - 1 \geq 0.$ ✓

Dim. 2 $\forall e\in\mathbb{R}$, $(X-e)^2 = [(X-\mu)+(\mu-e)]^2 =$

$= (X-\mu)^2 + (\mu-e)^2 + 2(\mu-e)(X-\mu)$, da cui

$E^P[(X-e)^2] \overset{(\text{def. }\mu)}{=} E^P[(X-\mu)^2] + (\mu-e)^2 \geq$

$\geq E^P[(X-\mu)^2]$, ed è $=$ $\Leftrightarrow e=\mu$. ✓ □

Quindi, tornando agli stimatori bayesiani, se consideriamo come costo $C(\vartheta,a) = |\vartheta-a|$, nel caso $g(\vartheta)=\vartheta\in\mathbb{R}\ni a$, allora dovremmo quindi trovare

$$\inf_{a\in\mathbb{R}}\int_{\mathbb{R}} C(\vartheta,a)\,d\nu(\vartheta) = \boxed{\inf_{a\in\mathbb{R}}\int_{\mathbb{R}}|\vartheta-e|\,d\nu(\vartheta)}.$$

Ora, ispirati dal punto $\boxed{1}$ dell'Ex. precedente, poniamo

$$(\Omega, \mathcal{F}, P) \equiv (\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu) \xrightarrow{X = id} (\mathbb{R}, \mathcal{B}(\mathbb{R}))_\nu$$

(nelle notazioni dell'esempio) $\quad (\Theta = \mathbb{R} = A)$,

supponiamo che ν abbia momento primo e sia m una mediana di ν: allora, come visto,

$$\inf_{a \in \mathbb{R}} \mathbb{E}^P[|X-a|] = \mathbb{E}^P[|X-m|] \quad \text{o} \ \underline{\text{cioè}}$$

$$\inf_{a \in \mathbb{R}} \int_{\mathbb{R}} |\theta - a| \, d\nu(\theta) = \int_{\mathbb{R}} |\theta - m| \, d\nu(\theta), \text{ o cioè}$$

$$\boxed{\int_{\mathbb{R}} |\theta - m| \, d\nu(\theta) \leq \int_{\mathbb{R}} |\theta - a| \, d\nu(\theta) \quad \forall a \in \mathbb{R}}$$ (ed, in

particolare, $\forall m_1, m_2$ mediane di ν, $\int_{\mathbb{R}} |\theta - m_1| d\nu(\theta) = \int_{\mathbb{R}} |\theta - m_2| d\nu(\theta)$).

(In questo caso, lo stimatore bayesiano di $\theta \in \mathbb{R}$ è una mediana di ν!)

(nel caso: $\omega \mapsto$ le mediane di ν^ω, ripete invece)"

$\boxed{\text{ESERCIZIO.}}$ Mettiamoci in $(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in (\Theta, \mathcal{T}, \nu)\}, \mu) =$

$= ((0,+\infty)^{\infty}, \mathcal{B}((0,+\infty)), \{\mu^\theta \mid \theta \in ((0,+\infty), \mathcal{B}((0,+\infty)), \nu)\}, \mu)$

$L^\theta(x)$ $\|$ uniforme su $[0,\theta]$, $\theta > 0$:

$$\boxed{\frac{d\mu^\theta}{d\lambda}(x) = \frac{1}{\theta} 1_{[0,\theta]}(x) = \frac{1}{\theta} 1_{[x,+\infty)}(\theta)} \quad (\forall x > 0)$$

(di Lebesgue su $\mathcal{B}((0,+\infty))$)

$$\boxed{d\nu(\theta) = \theta e^{-\theta} d\theta} \quad \text{(di nuovo)}$$

Sia $X(x) \doteq x$ $(X = id)$.

Per $f(\vartheta)=\vartheta\ \in(0,+\infty)$, calcolare lo "stimatore bayesiano" rispetto a

$$\begin{cases} \longrightarrow C(\vartheta,a)=(\vartheta-a)^2 \\ \longrightarrow C(\vartheta,a)=|\vartheta-a| \end{cases}$$

[Svolgiamo. $\boxed{C(\vartheta,a)=(\vartheta-a)^2}$. Osserviamo che

$$\int_0^{+\infty} L^\vartheta(x)\,d\nu(\vartheta) = \int_0^{+\infty}\left(\frac{1}{\vartheta}\mathbb{1}_{[x,+\infty)}(\vartheta)\right)\cdot\left(\vartheta e^{-\vartheta}\,d\vartheta\right) =$$

$$= \int_x^{+\infty} e^{-\vartheta}\,d\vartheta = e^{-x} > 0,$$ da cui

$$T = e^{x}\cdot\int_0^{+\infty}\vartheta L^\vartheta(x)\,d\nu(\vartheta).$$ Ma, analogamente,

$$\int_0^{+\infty}\vartheta L^\vartheta(x)\,d\nu(\vartheta) = \int_x^{+\infty}\vartheta e^{-\vartheta}\,d\vartheta = \underbrace{\left[\vartheta e^{-\vartheta}\right]^{x}_{+\infty}}_{(=xe^{-x})} +$$

$$+\underbrace{\int_x^{+\infty} e^{-\vartheta}\,d\vartheta}_{(=e^{-x})} = e^{-x}(x+1),$$ e allora

$\boxed{T=1+x}$. ✓ $\boxed{C(\vartheta,a)=|\vartheta-a|}$ Vogliamo verificare che esiste uno ed un solo $\overset{m(\omega)=m(x)}{m}\in\mathbb{R}$ tale che

$\frac{1}{2}\underset{(=)}{\leq}\nu^x([m,+\infty))$. Ma $\nu^x[T] =$
$\overset{\uparrow}{\underset{(\mathcal{T})}{(\mathcal{B}((0,+\infty)))}}$

$= e^X \int_T L^\vartheta(X)\, d\nu(\vartheta)$, come visto , perché

$$\frac{d\nu^X}{d\nu}(\vartheta) = e^X \underbrace{L^\vartheta(X)}_{\mathbb{1}_{[X,+\infty)}(\vartheta)}$$: dunque ~~[illegible]~~ $\exists !\, m$,

tale che / volete che $\frac{1}{2} = \nu^X([m,+\infty)) = e^X \int_m^{+\infty} \mathbb{1}_{[X,+\infty)}(\vartheta)\, e^{-\vartheta} d\vartheta =$

$$= e^X \int_{m \vee X}^{+\infty} e^{-\vartheta} d\vartheta = e^{X-m}$$, in quanto non

$$\to \begin{cases} e^{-m} & \text{se } m > X \\ e^{-X} & \text{se } m \le X \end{cases}$$

può essere $m \le X$ (per ω in quantità non trasc.) , altrimenti $\frac{1}{2} = 1$. Dunque $m = X + \log 2$. $\square$

TEST del punto di vista bayesiano , Poi
ESERCIZIO FINALE RIASSUNTIVO.

Sia $(\Omega, \mathcal{A}, \{P^\vartheta \mid \vartheta \in (\Theta, \mathcal{T}, \nu)\}, \mu)$ bayesiano (standard),

e sia $\emptyset \neq \Theta_0 \subsetneq \Theta$, con $\Theta_1 \doteq \Theta \setminus \Theta_0 = \Theta_0^c$,

e poniamoci il problema di test

e $H_0 : \vartheta \in \Theta_0$ contro $H_1 : \vartheta \notin \Theta_0$ ($\vartheta \in \Theta_1$)

Consideriamo allora come insieme delle azioni

$A \doteq [0, 1]$: facciamo (NATURALMENTE)

0: accettiamo H_0

1: rigettiamo H_0

in mezzo, "via di mezzo"

$$G(\vartheta, a) \doteq \begin{cases} a\, c_0 & \text{se } \vartheta \in \Theta_0 \quad (\text{MAX } c_0) \\ (1-a)\, c_1 & \text{se } \vartheta \in \Theta_1 \quad (\text{MAX } c_1) \end{cases}, \text{ con}$$

$$c \doteq \frac{c_1}{c_0} \text{ scelto}$$

c_1 (>0): costo errore 2° specie

c_0 (>0): costo errore 1° specie

Dunque, $\forall$ prob ρ su $(\Theta, \mathcal{T})$, in media otteniamo

$$\int_{\Theta} G(\vartheta, a)\, d\rho(\vartheta) = a\, c_0\, \rho(\Theta_0) + (1-a)\, c_1\, \rho(\Theta_1) =$$

$$= a\left[c_0\, \rho(\Theta_0) - c_1\, \rho(\Theta_1)\right] + c_1\, \rho(\Theta_1),$$

e allora riusciamo a fare $\inf_{a \in [0,1]}$ (MIN),

perché! : se effettivamente $c \doteq \dfrac{c_1}{c_0}$, allora

$$\min_{a \in \{0,1\}} \int_{\Theta} C(\theta,a)\, d\rho(\theta) = \begin{cases} c_1 \rho(\Theta_1) & \text{se } \rho(\Theta_0) > c\,\rho(\Theta_1) \\ \quad \text{per } a=0 & \\ c_0 \rho(\Theta_0) & \text{se } \rho(\Theta_0) < c\,\rho(\Theta_1) \\ \quad \text{per } a=1 & \\ c_1 \rho(\Theta_1) & \text{se } \rho(\Theta_0) = c\,\rho(\Theta_1) \\ \quad \text{ogni } a & \end{cases}$$

ossia risp. $d(\rho) = \begin{cases} 0 & \text{se } \rho(\Theta_0) > c\,\rho(\Theta_1) \\ 1 & \text{se } \rho(\Theta_0) < c\,\rho(\Theta_1) \\ \text{"quel che piace"} & \text{se } \rho(\Theta_0) = c\,\rho(\Theta_1) \end{cases}$

Dunque il test bayesiano è la decisione bayesiana

$$\phi(\omega) := d(\nu^{\omega}) = \begin{cases} 0 & \text{se } \nu^{\omega}(\Theta_0) > c\,\nu^{\omega}(\Theta_1) \\ 1 & \text{"} \quad < \quad \text{"} \\ \text{"che piace"} & \text{"} \quad = \quad \text{"} \end{cases}$$

è minimale. ✓

⇝ Dato l'esito $\omega \in \Omega$, si prende l'azione 1 (rigettare H_0) esattamente quando $\nu^{\omega}(\Theta_0) < c\,\nu^{\omega}(\Theta_1)$: supponendo $\int_{\Theta} L^{\theta}(\omega)\, d\nu(\theta) > 0$, significa (come sappiamo)

$$\int_{\Theta_0} L^{\theta}(\omega)\, d\nu(\theta) < c \int_{\Theta_1} L^{\theta}(\omega)\, d\nu(\theta).$$

$\int_{\Theta} \ldots \qquad \int_{\Theta} \ldots$

Per ciò, imponendo anche che $\mu\{\omega\in\Omega \mid \nu^\omega(\Theta_0)=c\,\nu^\omega(\Theta_1)\}=0$,

abbiamo che $\Phi(\omega)$ è di regione critica

$$D \doteq \left\{\omega\in\Omega \;\middle|\; \int_{\Theta_0} L^\theta(\omega)\,d\tau(\theta) < c\int_{\Theta_1} L^\theta(\omega)\,d\tau(\theta)\right\}$$

(anche perfettamente intuitivo!). ✓

Altrimenti, sarà

$$\{\omega\in\Omega \mid \nu^\omega(\Theta_0) < c\,\nu^\omega(\Theta_1)\} \subseteq D \subseteq \{\omega\in\Omega \mid \nu^\omega(\Theta_0) \le c\,\nu^\omega(\Theta_1)\},$$

(ES.) Se $\left(\Omega, \mathcal{F}, \left\{\{P^0, P^1\} \;\middle|\; \begin{matrix}\text{dati } \nu(0),\nu(1)\ge 0\\ \text{con } \nu(0)+\nu(1)=1\end{matrix}\right\}\right)$, e se il problema di test è a $\vartheta=0$ contro $\vartheta=1$, allora le "verosimiglianze" sono date da $L^0(\omega)$ e $L^1(\omega)$ ed il test Bayesiano è

$$\Phi(\omega) \doteq \begin{cases} 0 & \text{se } L^0(\omega) > k\,L^1(\omega) \\ 1 & \text{" } < \text{ "} \\ \text{a piacere (ma costante)} & \text{" } = \text{ "} \end{cases},$$

$k \doteq \dfrac{c_1\,\nu(1)}{c_0\,\nu(0)}$, essendo $(\forall\omega\in\Omega)$ $\begin{cases} \nu^\omega(0) = L^0(\omega)\nu(0) \big/ \left(L^0(\omega)\nu(0)+L^1(\omega)\nu(1)\right) \\ \nu^\omega(1) = L^1(\omega)\nu(1) \big/ \left(L^0(\omega)\nu(0)+L^1(\omega)\nu(1)\right) \end{cases}$,

che è solamente un test di Neyman–

– Pearson. :) ✓

ESERCIZIO(NE) FINALE RIASSUNTIVO.

$(X_1, \dots, X_n)$ campione di legge μ^θ su $(0,1)$ di densità $\boxed{f^\theta(x) \doteq \theta x^{\theta-1}}$, $x \in (0,1)$, dove $\underline{\theta > 0}$, rispetto alla Lebesgue λ su $\mathcal{B}(0,1)$. Per ciò, il modello è

$$(\Omega, \mathcal{F}, \{P^\theta \mid \theta \in \Theta\}, \mu) = \left((0,1)^n, \mathcal{B}((0,1)^n), \{(\mu^\theta)^{\otimes n} \mid \theta \in (0,+\infty)\}, \lambda^{\otimes n}\right),$$

con $X_i(x_1, \dots, x_n) \doteq x_i$, e la verosimiglianza è

$$L^\theta(x_1, \dots, x_n) = \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}$$

, e quindi il modello è <u>regolare</u>.

(1) STATISTICA ESAUSTIVA E COMPLETA (?) Vale $L^\theta(x_1, \dots, x_n) =$

$$= \exp\left((\theta-1)\log\left(\prod_{i=1}^n x_i\right) + n\log\theta\right),$$

ossia se

$$\boxed{T \doteq \log \prod_{i=1}^n X_i} \quad \left(= \sum_{i=1}^n \log X_i\right)$$

allora $L^\theta = \exp((\theta-1)T + n\log\theta)$, per cui T è <u>esaustiva</u> (Neymann-Fischer), ed è anche <u>completa</u> perché è σ-finita (su $\mathcal{B}((-\infty,0))$), essendo assolutamente continua rispetto alla Lebesgue, ed il modello è "sostanzialmente" esponenziale (che densità!). ✓

(2) STIMATORI DI θ. Proviamo direttamente con la Massima verosimiglianza:

$\frac{d}{d\vartheta}\log L^{\vartheta} = 0 \iff T = -\frac{n}{\vartheta}$, da cui

$$\boxed{\hat{\vartheta} = -\frac{n}{T}} = -\frac{n}{\sum_{i=1}^{n}\log(X_i)} \quad (>0!) .$$

Adesso, $\boxed{\text{NOTA}}$ Sotto P^{ϑ}, $-\log X_i \sim \mathcal{E}(\vartheta)$. $(=\Gamma(1,\vartheta))$

[Perché $-\log X_i > 0$ e $P^{\vartheta}[-\log X_i \le \underbrace{a}_{(>0)}] =$

$= P^{\vartheta}[X_i \ge e^{-a}] = 1 - P^{\vartheta}[X_i < e^{-a}]$, ed ora

$P^{\vartheta}[X_i < e^{-a}] = \int_0^{e^{-a}} \vartheta x^{\vartheta-1}\,dx = [x^{\vartheta}]_{x=0}^{x=e^{-a}} = e^{-a\vartheta}$. ✓]

Allora, per indipendenza, $\boxed{-T = -\sum_{i=1}^{n}\log(X_i) \overset{\text{sotto } P^{\vartheta}}{\sim} \Gamma(n,\vartheta)}$,

da cui $E^{P^{\vartheta}}[-\frac{1}{T}] = \int_0^{+\infty} \frac{1}{x}\frac{1}{\Gamma(n)}\vartheta^{n} x^{n-1} e^{-\vartheta x}\,dx \underset{\uparrow}{=}$ sì ($\Gamma(n)=\Gamma(n-1)(n-1)$)

$$= \frac{1}{n-1}\vartheta \underbrace{\int_0^{+\infty} \frac{1}{\Gamma(n-1)}\vartheta^{n-1} x^{(n-1)-1} e^{-\vartheta x}\,dx}_{=1!} =$$

$= \frac{\vartheta}{n-1}$, e allora (mentre $\hat{\vartheta} = -\frac{n}{T}$ è stimatore di MAX verosimiglianza per ϑ) $\boxed{U \doteq -\frac{n-1}{T}}$ (>0) è uno ~~("[illegible]")~~

stimatore corretto di ϑ ed ottimale che tutti gli stimatori ~~di~~ corretti di ϑ di quadrato int. (per Blackwell-Rao). ✓

$\boxed{\text{③ TEST UNILATERO}}$ Ci poniamo il problema del test
« $\mathcal{H}_0 : \vartheta \leq 1$ contro $\mathcal{H}_1 : \vartheta > 1$ »,
unilatero „standard" („col ≤") con $\vartheta_0 = 1$. Il modello
è regolare, quindi possiamo fare il rapporto di verosimiglianza:
per ogni $0 < \vartheta_1 < \vartheta_2$, vale
($\forall\, x_i \in (0,1)$.)

$$\frac{L^{\vartheta_2}(x_1,\dots,x_n)}{L^{\vartheta_1}(x_1,\dots,x_n)} = \left(\frac{\vartheta_2}{\vartheta_1}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\vartheta_2-\vartheta_1 \cancel{\mp 1}} =$$

$$= \Lambda^{\vartheta_1,\vartheta_2}(y)\Big|_{y=\prod_{i=1}^{n} x_i} \rightarrow$$

$\rightarrow$ monotòno crescente rispetto alla statistica
$\prod_{i=1}^{n} X_i$. Allora poniamo col test di regione
critica $D_\alpha \doteq \left\{\prod_{i=1}^{n} X_i > C\right\}$, $C \in (0,1)$,
in fareundo $P^{1}\left[\left\{\prod_{i=1}^{n} X_i > C\right\}\right] = \alpha$, con
$\alpha \in (0,1)$ ($\alpha \ll 1$) assegnato, tipo $\alpha = 0.05$.

Ma, come già osservato, $-\sum_{i=1}^{n} \log X_i \overset{P^1}{\sim} \Gamma(n,1)$,

per cui scriviamo $\left\{\prod_{i=1}^{n} X_i > e\right\} = \left\{-\sum_{i=1}^{n} \log X_i < -\log e\right\} =$

$$= \int_0^{-\log(e)} \frac{1}{(n-1)!} x^{n-1} e^{-x}\,dx \quad : \quad \underline{\underline{OK}}\,.\ (\text{Già visto!}).$$

4 STIMATORI BAYESIANI DI θ

Su $\Theta = (0,+\infty)$, poniamo $\mathcal{T} \doteq \mathcal{B}((0,+\infty))$ e $d\nu(\theta) \doteq e^{-\theta}\,d\theta$. Date le regolarità di $L^\theta(\omega)$, possiamo usare le stesse nozioni/formule di stimatori bayesiani di θ, con per altro $\int_\Theta L^\theta(\underset{(x_1,\dots,x_n)}{\omega})\,d\nu(\theta) =$

$= \int_0^{+\infty} \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1} e^{-\theta}\,d\theta$ sicuramente > 0 $\underline{\forall \omega \in \Omega}$,

da cui lo stimatore desiderato è

$$\boxed{\frac{\int_0^{+\infty} \theta^{n+1} \left(\prod_{i=1}^{n} X_i\right)^{\theta \cancel{-1}} e^{-\theta}\,d\theta}{\int_0^{+\infty} \theta^{n} \left(\prod_{i=1}^{n} X_i\right)^{\theta \cancel{-1}} e^{-\theta}\,d\theta}}$$

. Osserviamo

che $(\forall b>0)$ $\int_0^{+\infty} \theta^n \underbrace{b^{\theta \cancel{-1}}}_{\left(\exp((\theta \cancel{-1})\log(b))\right)} e^{-\theta}\,d\theta = \int_0^{+\infty} \theta^n e^{-\theta(1-\log(b))}\,d\theta =$

$= \cancel{\ldots} \dfrac{n!}{(1-\log(b))^{n+1}}$ (VISTO IN PRECEDENZA!) , da cui in definitiva lo

stimatore bayesiano di ϑ è

$$\frac{\dfrac{(n+1)!}{\left(1-\log\left(\prod_{i=1}^{n} x_i\right)\right)^{n+2}}}{\dfrac{n!}{\left(1-\log\left(\prod_{i=1}^{n} x_i\right)\right)^{n+1}}} = \frac{n+1}{1-\log\left(\prod_{i=1}^{n} x_i\right)} =$$

$$= \frac{n+1}{1-T} .$$

5 (TEST BAYESIANO) Ci poniamo il problema di test $\vartheta = 1$ contro $\vartheta \neq 1$, e scegliamo un test bayesiano. Fissiamo i costi $\begin{smallmatrix} c_0 \\ c_1 \end{smallmatrix}$ (>0) col il rapporto $c \doteq \dfrac{c_1}{c_0}$, e scegliamo come legge a priori su $\Theta =]0,+\infty)$

$$\nu \doteq \frac{1}{2}\delta_1 + \frac{1}{2}\mu \quad , \text{ dove } d\mu(\vartheta) \doteq e^{-\vartheta} d\vartheta .$$

(Dirac in 1)

Dunque $\Theta_0 = \{1\}$, e ~~sappiamo che~~

$$\frac{d\nu^\omega}{d\nu}(\vartheta) = \frac{\mathcal{L}^\vartheta(\omega)}{\int_\Theta \mathcal{L}^t(\omega)\,d\nu(t)} = \left(\bullet \longleftarrow (\omega = (x_1,\dots,x_n))\right)$$

(VISTO ORA!)

$$= \frac{\vartheta^n\left(\prod_{i=1}^n x_i\right)^{\vartheta-1}}{\dfrac{n!}{\left(1-\sum_{i=1}^n \log x_i\right)^{n+1}}}(\omega) =$$

(ν è f de prima!!)

$$= \frac{\vartheta^n\left(\prod_{i=1}^n x_i(\omega)\right)^{\vartheta-1}\left(1-\sum_{i=1}^n \log x_i\right)^{n+1}}{n!}.$$

Scriviamo come regione critica

$$D_C \doteq \left\{\omega\in\Omega \;\middle|\; \int_{\Theta_0}\mathcal{L}^\vartheta(\omega)\,d\nu(\vartheta) < C\int_{\Theta_1}\mathcal{L}^\vartheta(\omega)\,d\nu(\vartheta)\right\} =$$

$\nu = \frac{1}{2}\delta_1 + \frac{1}{2}\mu$

$\Theta_0 = \{1\}$

$\Theta_1 = (0,1)\cup(1,+\infty)$

$$= \left\{\omega\in\Omega \;\middle|\; \frac{1}{2}\underbrace{\mathcal{L}^1(\omega)}_{1} < C\,\frac{1}{2}\int_0^{+\infty}\mathcal{L}^\vartheta(\omega)\,e^{-\vartheta}\,d\vartheta\right\} = (\text{OK})$$

$$= \left\{\omega\in\Omega \;\middle|\; 1 < C\left(\frac{n!}{\left(1-\sum_{i=1}^n \log x_i\right)^{n+1}}\right)\frac{1}{\prod_{i=1}^n x_i}\right\} =$$

$$= \left\{\omega\in\Omega \;\middle|\; \left(\prod_{i=1}^n x_i\right)\left(1-\sum_{i=1}^n \log x_i\right)^{n+1} < C\cdot n!\right\},$$ e ora

studiata $b\in(0,1)\mapsto b(1-\log b)^{n+1}$. ☺ $\square$