Approccio frequentista alla statistica

Intervalli di Confidenza

Wednesday, December 13, 2023 9:04 AM

Costruzione di Neyman degli intervalli di confidenza (CI)

↳ Proprietà statistica di questi CI : "Coverage"

↳ Ricerca di 2 estimatori (variabili casuali) $\hat{\theta}_2(X_1 \dots X_n)$ $\hat{\theta}_1(X_1 \dots X_n)$ t.c., fissato γ t.c. $0 \leq \gamma \leq 1$

$$P_\theta\left[\hat{\theta}_2(X_1 \dots X_n) \leq \theta \leq \hat{\theta}_1(X_1 \dots X_n)\right] = 1-\gamma \equiv CL$$

"Coverage"

Livello di confidenza

↳ $\forall \theta \in \Theta$

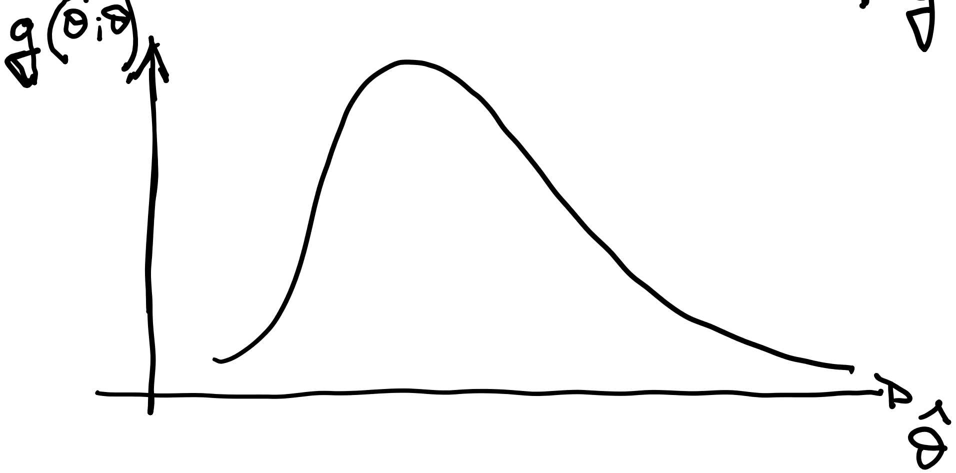
Quindi in particolare per il valore "vero" (incognito) del parametro n osservazioni sperimentali $x_1 \dots x_n$ di $X \sim f(x;\theta)$

$\hat{\theta}(\underbrace{X_1 \dots X_n}_{\text{iid } \sim f(x;\theta)})$ estimatore (ML) di θ

$\hat{\theta}_{obs} = \hat{\theta}(x_1 \dots x_n)$ stima di θ in corrisp. dell'osservazione sperim. $x_1 \dots x_n$

Supp.mo di conoscere la pdf di $\hat{\theta}$ $\forall \theta$ fissato (ma arbitrario)

$\hookrightarrow g(\hat{\theta};\theta)$



$\hat{\theta} \sim g(\hat{\theta};\underset{\substack{\downarrow \\ \text{fissato, ma arbitrario}}}{\theta})$

$\therefore\ \alpha, \beta$: probabilità

Wednesday, December 13, 2023 9:05 AM

Fissati a priori: α, β $\quad 0 \leq \alpha, \beta \leq 1 \quad 0 \leq \alpha + \beta \leq 1$

$\forall \theta \in \Theta$ sia $u_\alpha(\theta)$ il valore di $\hat{\theta}$ per cui:

$$\boxed{\alpha = P_\theta\left(\hat{\theta} \geq u_\alpha(\theta)\right)} = \int_{u_\alpha(\theta)}^{+\infty} g(\hat{\theta};\theta)\, d\hat{\theta} = 1 - G(u_\alpha(\theta);\theta)$$

$\circledast$

Supponendo che il supporto di $\hat{\theta}$ sia $(-\infty, +\infty)$

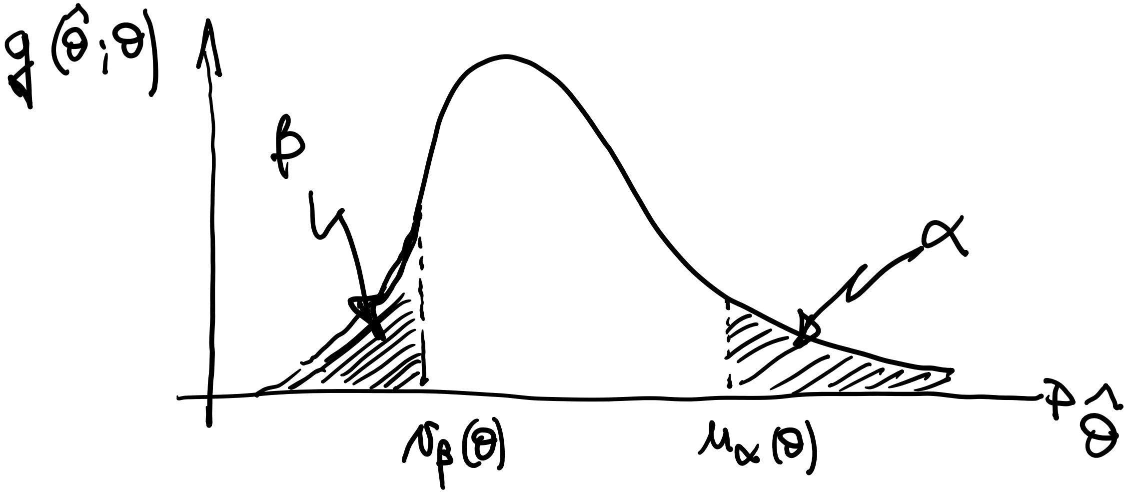
$$G(\hat{\theta};\theta) = \int_{-\infty}^{\hat{\theta}} g(\hat{\theta}';\theta)\, d\hat{\theta}'$$

$\hookrightarrow$ c.d.f. di $\hat{\theta}$ a θ assegnato

Altrettanto $(\forall \theta \in \Theta)$ sia $v_\beta(\theta)$ il valore di $\hat{\theta}$ per cui:

$$\boxed{\beta = P_\theta\left[\hat{\theta} \leq v_\beta(\theta)\right]} = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta};\theta)\, d\hat{\theta} = G(v_\beta(\theta);\theta)$$

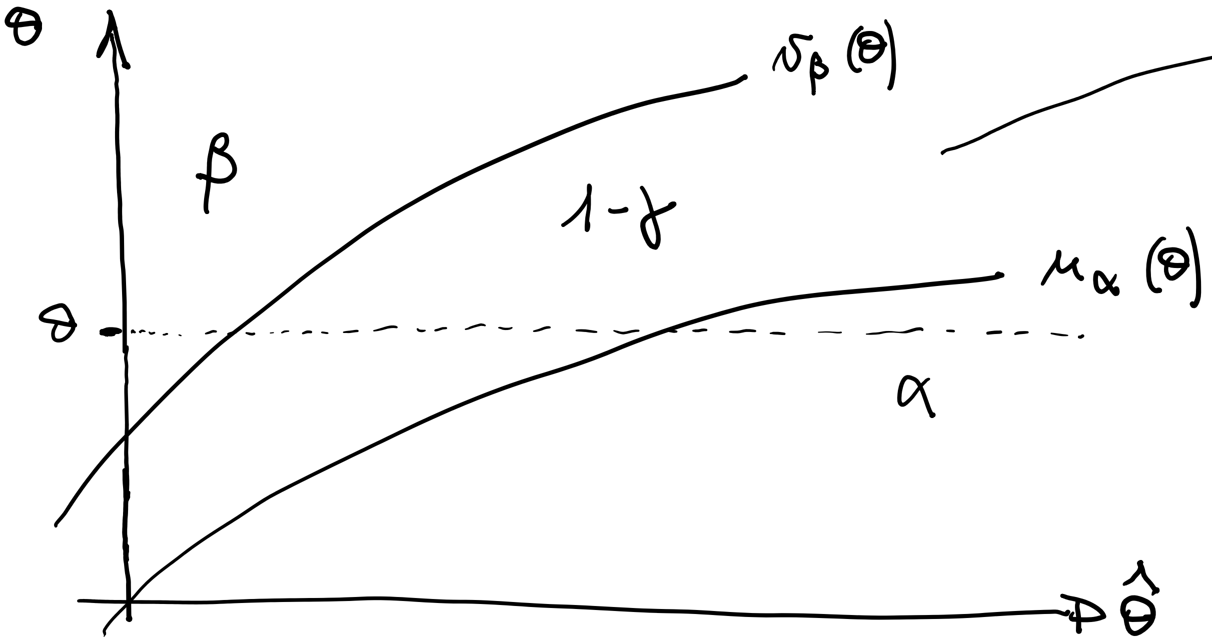
$\circledast\circledast$



θ fissato (arbitrario)

Confidence Belt
(Banda di confidenza)
con CL = $1-\gamma$

$\forall \theta$ ho un valore $u_\alpha(\theta)$ e un valore $v_\beta(\theta)$



Posto $\gamma \equiv \alpha+\beta$

$\underbrace{1-\gamma}_{CL} = 1-\alpha-\beta$

$\therefore \quad P_\theta\left(v_\beta(\theta) \le \hat{\theta} \le u_\alpha(\theta)\right) = 1-\gamma = 1-\alpha-\beta = CL$

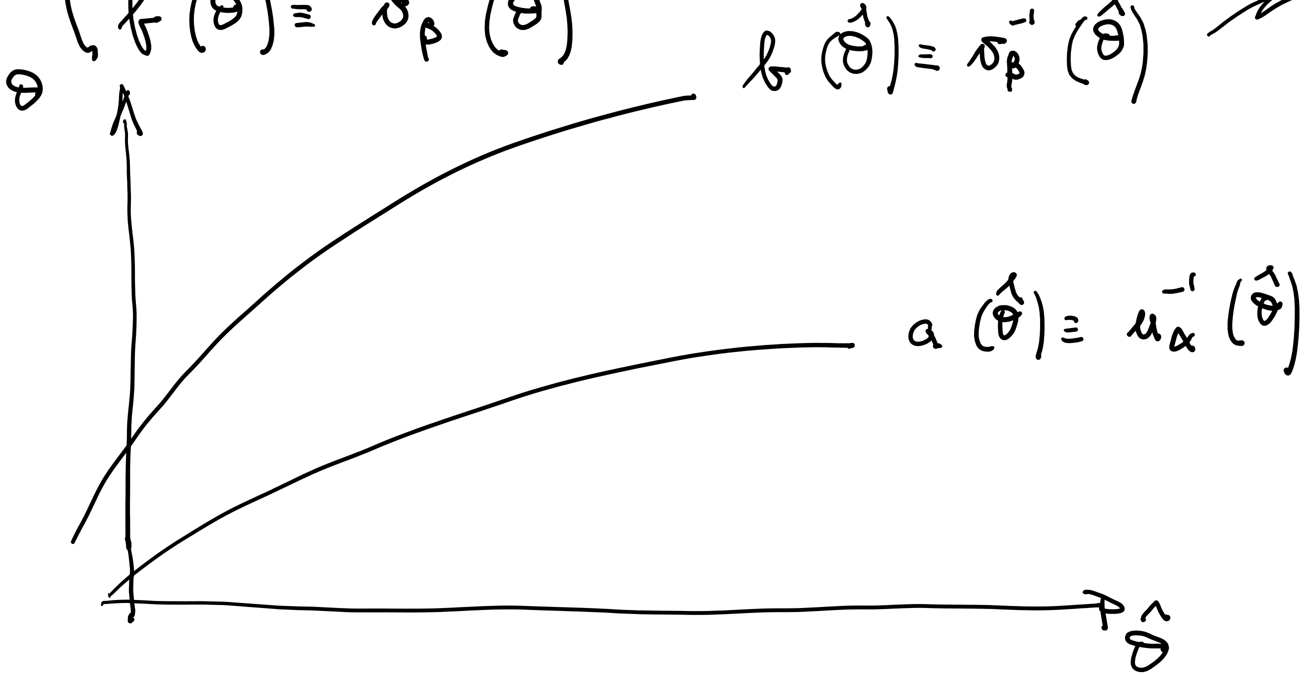
$\forall \theta \in \Theta$

NB Qui la vaiabile casuale è $\hat{\theta}$

Wednesday, December 13, 2023 9:06 AM

Se $u_\alpha(\theta)$ e $v_\beta(\theta)$ sono monotone crescenti, $\exists$ inverse:

$$\begin{cases} a(\hat{\theta}) \equiv u_\alpha^{-1}(\hat{\theta}) \\ b(\hat{\theta}) \equiv v_\beta^{-1}(\hat{\theta}) \end{cases}$$



Il grafico di $b(\hat{\theta})$ coincide con quello di $v_\beta(\theta)$

Il grafico di $a(\hat{\theta})$ coincide con quello di $u_\alpha(\theta)$

Le due diseguaglianze:

$$\hat{\theta} \geq \mu_\alpha(\theta)$$

$$\hat{\theta} \leq \nu_\beta(\theta)$$

se le due funzioni sono monotone crescenti, implicano che:

$$a(\hat{\theta}) \geq \theta$$

$$b(\hat{\theta}) \leq \theta$$

Le due relazioni probabilistiche (*) e (**) con cui ho definito $\mu_\alpha(\theta)$ e $\nu_\beta(\theta)$ diventano:

$$\begin{cases} P_\theta\left(a(\hat{\theta}) \geq \theta\right) = \alpha & (*) \\ P_\theta\left(b(\hat{\theta}) \leq \theta\right) = \beta & (**) \end{cases}$$

$\forall \theta$ fissato (ma arbitrario) $(\forall \theta \in \Theta)$

Wednesday, December 13, 2023 9:06 AM

$$\Rightarrow \boxed{P_\theta \left(a(\hat{\theta}) \le \theta \le b(\hat{\theta}) \right) = 1-\alpha-\beta = 1-\gamma = CL}$$

NB : Le variabili casuali in gioco qui sono $a(\hat{\theta})$ e $b(\hat{\theta})$

"coverage" (per costruzione) del CI costruito alla Neyman

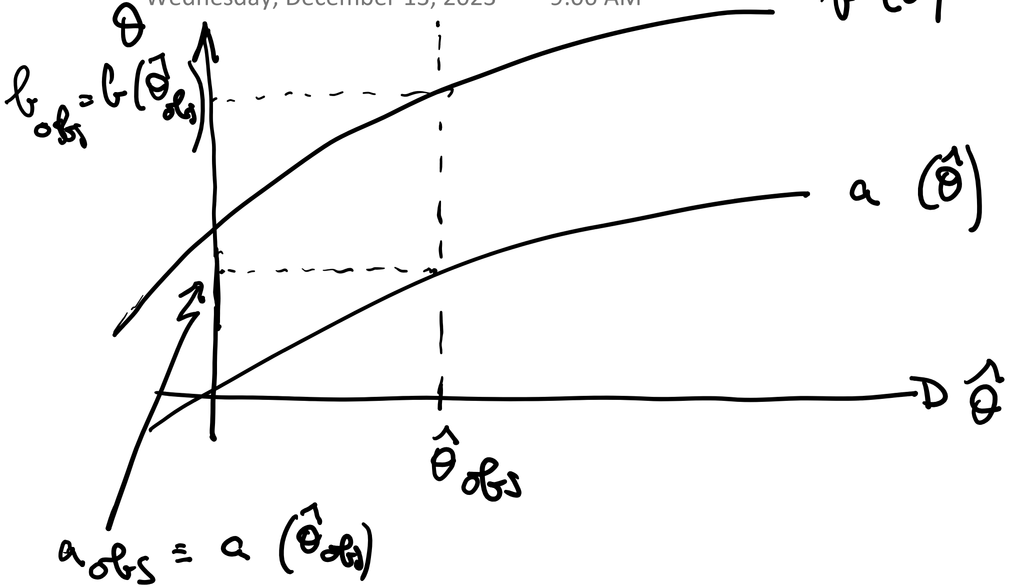
Dati $(x_1^{obs}, \dots x_n^{obs}) \rightarrow \hat{\theta}_{obs} = \hat{\theta}(x_1^{obs}, \dots x_n^{obs})$

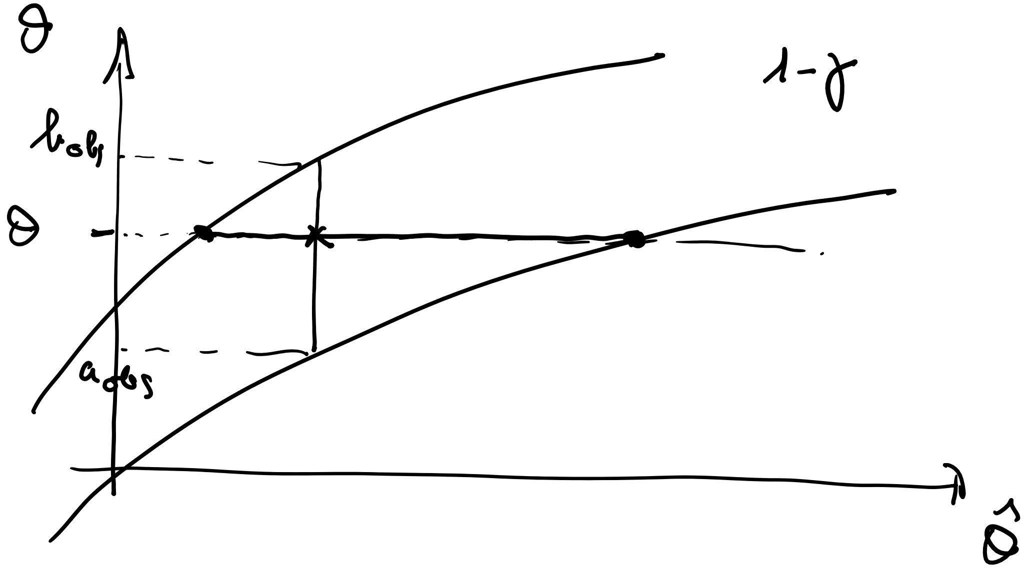
Costruzione di Neyman del CI al livello di confidenza $CL = 1-\gamma$

$$\begin{cases} a_{obs} \equiv a(\hat{\theta}_{obs}) \\ b_{obs} \equiv b(\hat{\theta}_{obs}) \end{cases}$$

L'intervallo $[a_{obs}, b_{obs}]$ si chiama CI di Neyman al $CL = 1-\gamma = 1-\alpha-\beta$

Wednesday, December 13, 2023 9:06 AM





$$P_\theta \left(a_{obs} \leq \theta \leq b_{obs} \right) = 1-\gamma = CL$$

$\forall \theta \in \Theta$

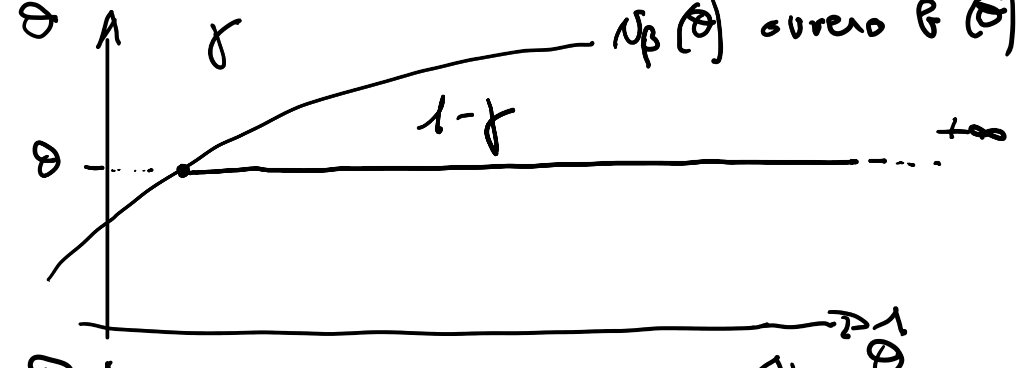
in partic. per $\theta = \theta$ "vero"

Fissato $CL = 1-\gamma$, ossia fissato $\alpha+\beta$, abbiamo diverse possibilità di fissare α, β

Fra le più comuni:

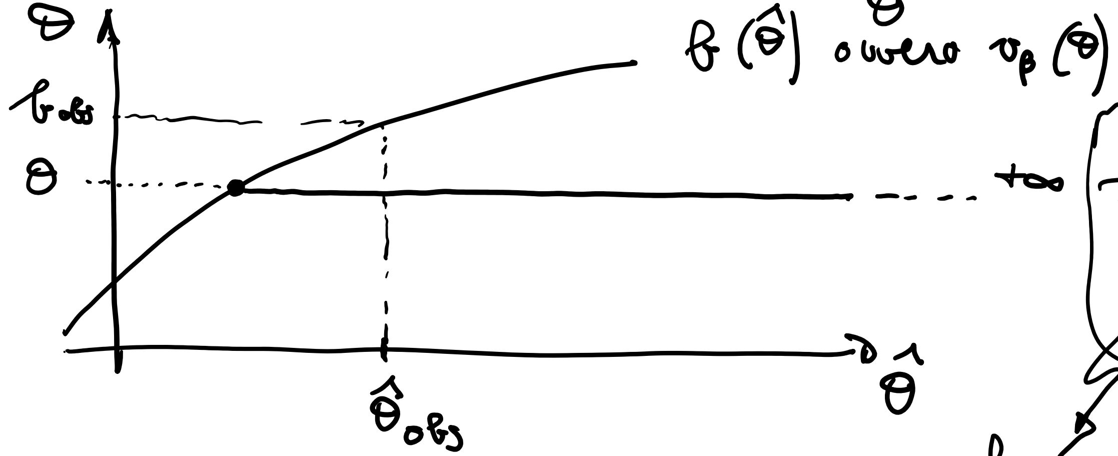
*) $\alpha = \beta = \frac{\gamma}{2} \Rightarrow$ "Intervalli di confidenza centrali" (al $CL = 1-\gamma$)

*) $\begin{cases} \beta = \gamma \\ \alpha = 0 \end{cases}$



$P_\theta(\hat{\theta} \geq v_\beta(\theta)) = 1-\gamma$

$\hookrightarrow \forall \theta \in \Theta$

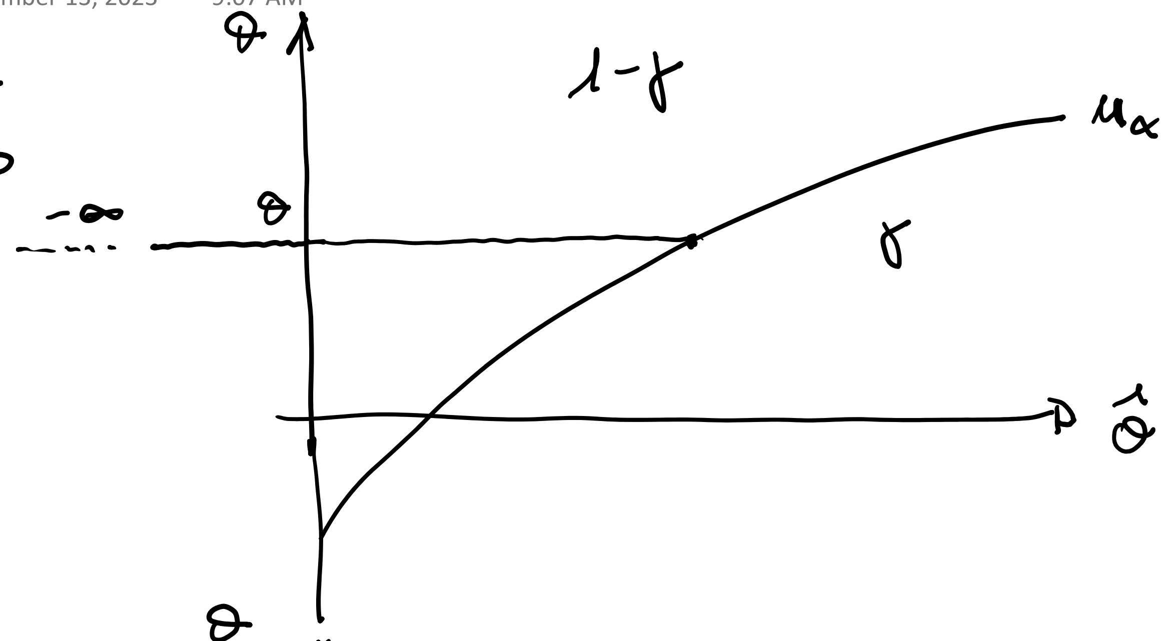


$$\boxed{P_\theta\left[\theta \leq t_{obs} = f(\hat{\theta}_{obs})\right] = 1-\gamma}$$

t_{obs} è UL su θ con $CL = 1-\gamma$

Wednesday, December 13, 2023 9:07 AM

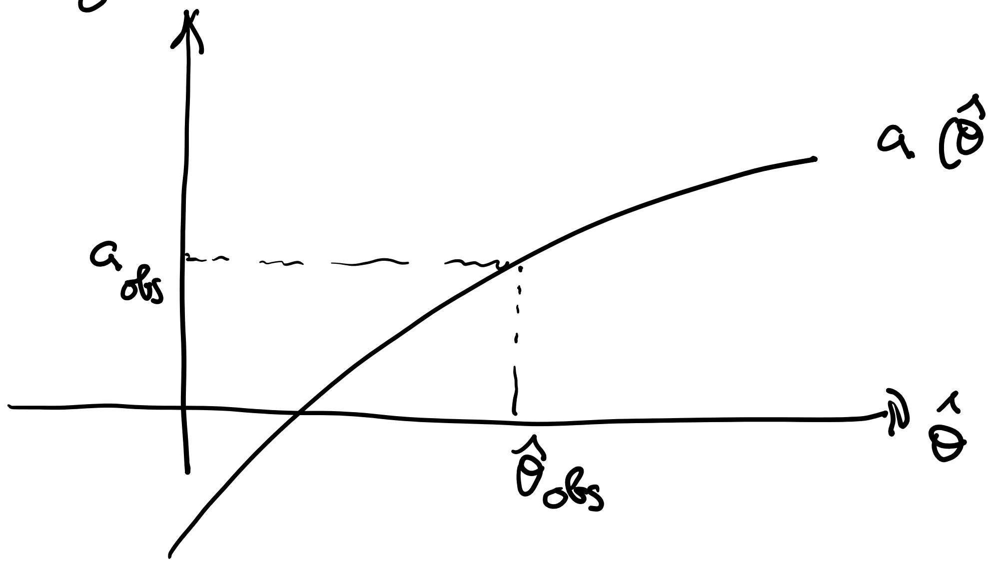
$$(*)\begin{cases}\alpha = \gamma \\ \beta = 0\end{cases}$$



$u_\alpha(\theta)$ ovvero $a(\hat\theta)$

$P_\theta\left[\hat\theta \le u_\alpha(\theta)\right] = 1-\gamma$

$\hookrightarrow \forall\, \theta \in \Theta$

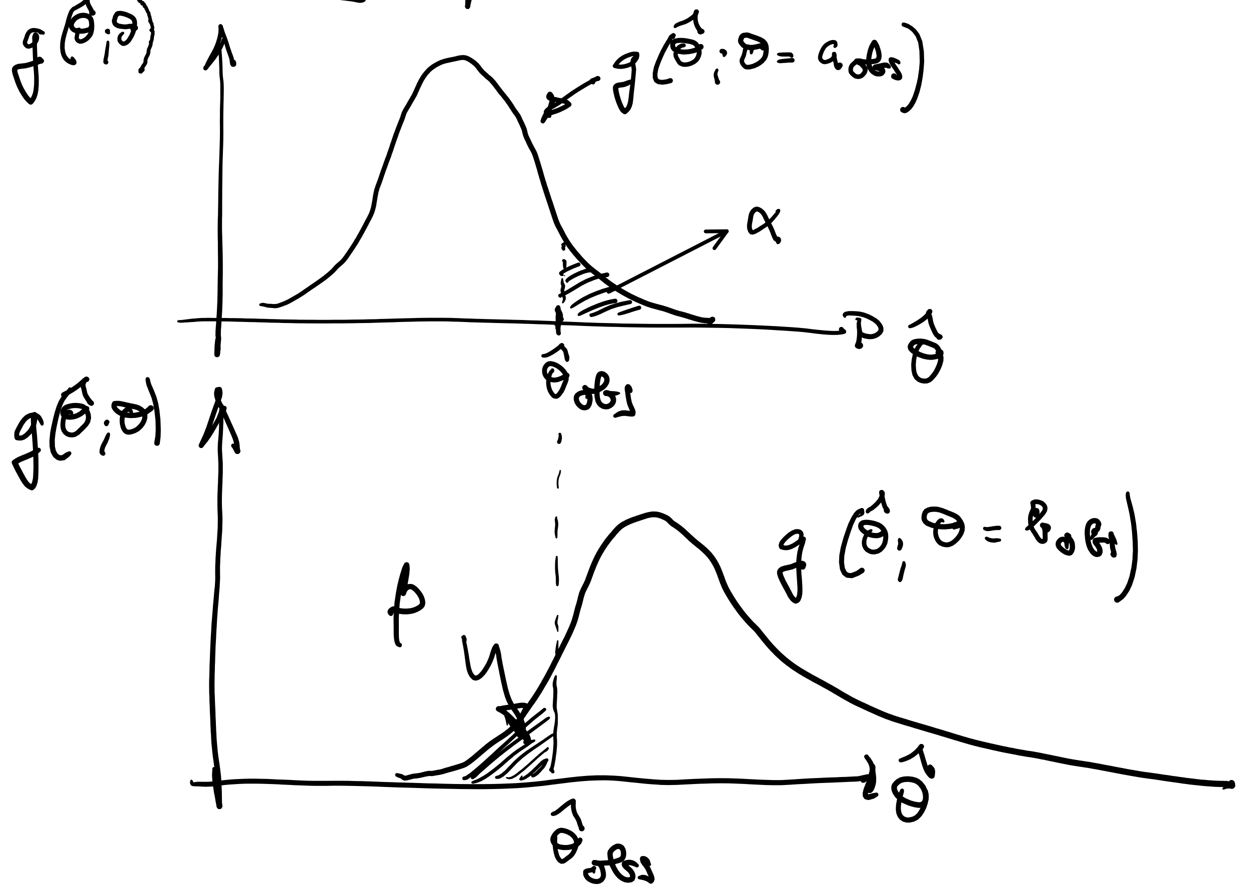


$a(\hat\theta)$ ovvero $u_\alpha(\theta)$

$P_\theta\left[\theta \ge a_{obs} \equiv a(\hat\theta_{obs})\right] = 1-\gamma$

a_{obs} è LL su θ con CL $= 1-\gamma$

Per costruzione, il valore $a_{obs} \equiv a(\hat{\theta}_{obs})$ ha anche il significato di "valore ipotetico del parametro θ della distribuzione per il quale una frazione α di stime ripetute $\hat{\theta}$ di θ risulterebbe <u>superiore</u> al valore $\hat{\theta}_{obs}$ sperimentalmente osservato"

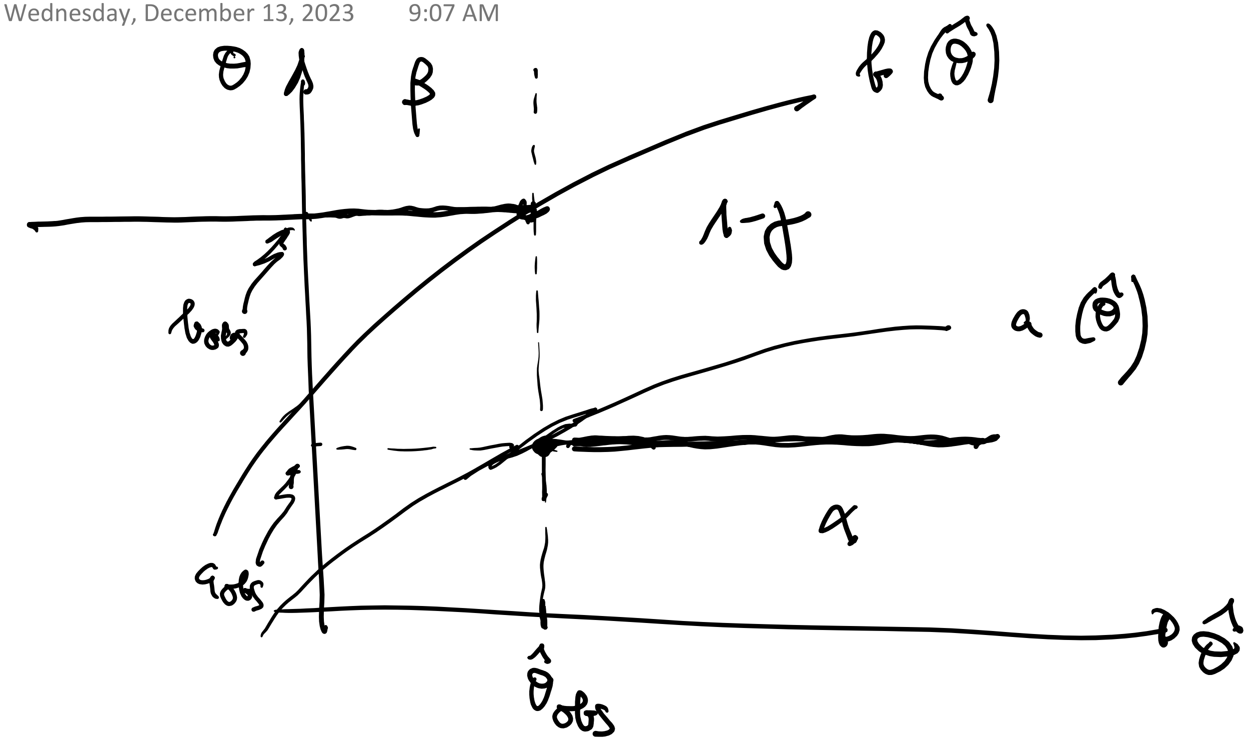


Analogamente, il valore $b_{obs} = b(\hat{\theta}_{obs})$ ha anche il significato di "valore ipotetico del parametro θ della distribuzione per il quale una frazione β di stime ripetute $\hat{\theta}$ di θ risulterebbe <u>inferiore</u> a $\hat{\theta}_{obs}$ sperim. osservato

Vale a dire, preso $\hat{\theta}_{obs} = u_\alpha(a_{obs}) = v_\beta(b_{obs})$, le equazioni (*) e (**) diventano:

$$\underset{\displaystyle P_{\theta=a_{obs}}(\hat{\theta} \geq \hat{\theta}_{obs})}{\underset{\parallel}{\alpha}} = \int_{\hat{\theta}_{obs}}^{+\infty} g(\hat{\theta}; a_{obs})\, d\hat{\theta} = 1 - G(\hat{\theta}_{obs}; a_{obs}) \quad (*)$$

$$\underset{\displaystyle P_{\theta=b_{obs}}(\hat{\theta} \leq \hat{\theta}_{obs})}{\underset{\parallel}{\beta}} = \int_{-\infty}^{\hat{\theta}_{obs}} g(\hat{\theta}; b_{obs})\, d\hat{\theta} = G(\hat{\theta}_{obs}; b_{obs}) \quad (**)$$



θ
β
b(θ̂)
1-γ
b_obs
a(θ̂)
a_obs
α
θ̂
θ̂_obs

Intervalli di confidenza per il parametro ν di una distribuzione Poissoniana

$\left(\nu = E[n] ; \quad n \sim \mathcal{P}(\nu) \right)$

$$f(n;\nu) = \frac{e^{-\nu}\nu^{n}}{n!} =$$

Fissati α, β t.c. $\alpha + \beta = \gamma$ $(1-\gamma = CL)$, data l'osservazione n_{obs} $\left(\underbrace{\hat{\nu}_{obs} = n_{obs}}_{MLE} \right)$

$$\begin{cases} \alpha = P\left(\hat{\nu} \geq \hat{\nu}_{obs} = n_{obs} ; a\right) & (*) \\ \beta = P\left(\hat{\nu} \leq \hat{\nu}_{obs} = n_{obs} ; b\right) & (**) \end{cases}$$

$\rightarrow$ Risolvendo queste 2 equazioni, si ricavano gli estremi $a(n_{obs})$, $b(n_{obs})$ del CI $[a,b]$ con $CL = 1-\gamma$

In concreto:

$$\begin{cases} \alpha = \sum_{n=n_{obs}}^{+\infty} f(n;a) = 1 - \sum_{n=0}^{n_{obs}-1} f(n;a) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{e^{-a} a^n}{n!} & (*) \\ \beta = \sum_{n=0}^{n_{obs}} f(n;b) = \sum_{n=0}^{n_{obs}} \frac{e^{-b} b^n}{n!} & (**) \end{cases}$$

$\rightarrow$ Da risolvere numericamente per a e b, dato n_{obs} e fissati α e β

NB In conseguenza della natura discreta dell'osservabile n_{obs} di questo problema, l'intervallo di confidenza $[a, b]$ che si ottiene attraverso la procedura di Neyman è un CI "conservativo"

$$\therefore \quad P_\nu\left[a(n_{obs}) \leq \nu \leq b(n_{obs})\right] \geq 1-\gamma = CL$$

anziché "=" (come sarebbe nel caso di distribuzioni con osservabili continue)