Inferenza statistica – Scritto 2 – 1 luglio 2024

Cognome, nome e n. di matricola: ______________________

- Si corregge solo quanto riportato in questi fogli.
- Gli studenti che hanno diritto alla riduzione del compito possono saltare i quesiti: Parte A, quesito 2; Parte B, quesiti 2-3.

Parte A - Quesiti

1. Si consideri un campione casuale $X_1, \dots, X_n$ in cui interessa effettuare confrontare le ipotesi $H_0 : \mu = \mu_0 = 0$ vs $H_1 : \mu \neq \mu_0$ sul valore atteso μ della distribuzione della singola osservazione X_i per la quale anche la varianza σ^2 è incognita. Un software fornisce il seguente output:

```
##  One Sample t-test
##
## t = 1.1328, df = 29, p-value = 0.2666
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:  -0.3268533  1.1384383
## sample estimates: mean of x = 0.4057925
```

(a) Quale ipotesi distributiva è necessaria sulle variabili aleatorie X_i per poter utilizzare il test T?

Risp. $X_i \sim N(\mu, \sigma^2)$; entrambi i parametri incogniti
iid

(b) Illustra come è possibile decidere in favore dell'accettazione o del rifiuto di H_0 in base all'output, avendo fissato probabilità di errore di prima specie $\alpha = 0.05$.

Risp. p-value $\geq 0.05 \Rightarrow$ ACCETTO H_0
OPPURE
$|t| = 1.1328 < z_{1-\frac{\alpha}{2}} = 1.96 \Rightarrow$ ACC. H_0

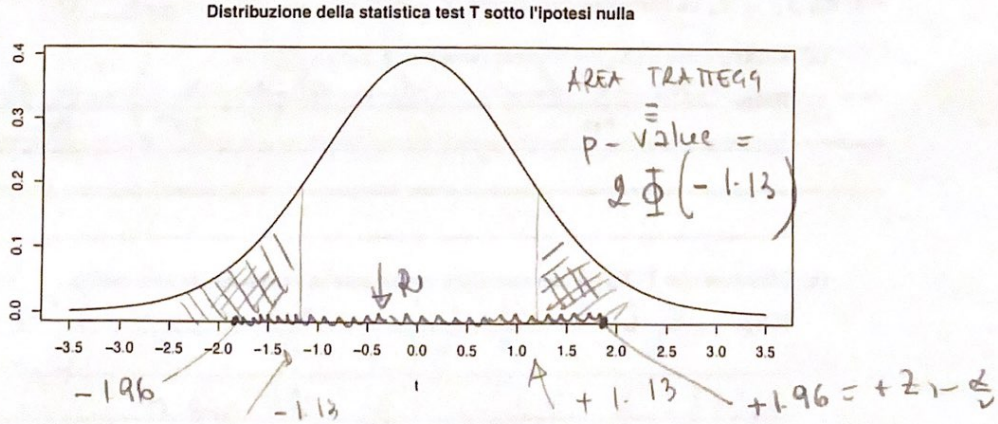
(c) Scrivere la formula della statistica test utilizzata, in corrispondenza della quale è stato ottenuto per il campione osservato il valore numerico 1.1328.

Risp. $W = \frac{\sqrt{n}}{S_n}(\bar{x}_n - \mu_0)$ con $\begin{cases} \mu_0 = 0 \\ S_n^2 = \sum (x_i - \bar{x}_n)^2 / n-1 \end{cases}$

(d) Nel seguente grafico viene riportata la funzione di densità della statistica test sotto ipotesi nulla. È possibile rappresentare graficamente il p-value riportato nell'output? In caso affermativo rappresentalo.

Risp. Sì ☒ No ☐

Distribuzione della statistica test T sotto l'ipotesi nulla



(e) Si può rappresentare sullo stesso grafico la potenza del test? In caso affermativo rappresentala.

Risp. Sì ☐ No ☒

(f) Evidenziare sullo stesso grafico l'insieme dei valori critici della statistica test T corrispondente alla regione di rifiuto del test di dimensione $\alpha = 0.05$.

2. Si consideri un campione casuale $X_1, \ldots, X_n$ con $X_i \sim \text{Pois}(\theta)$ e il seguente sistema di ipotesi sul valore atteso θ della distribuzione: $H_0 : \theta = 1$ vs $H_1 : \theta \leq 1$.

(a) Seleziona una regione di rifiuto tra le seguenti che ritieni che abbia buone proprietà inferenziali per un campione di ampiezza $n = 3$:

$$R_1 = \left\{(x_1, ..., x_n) : \sum_{i=1}^{n} x_i \leq 2\right\} ;\ R_2 \left\{(x_1, ..., x_n) : \sum_{i=1}^{n} x_i > 2\right\} ;\ R_3 \left\{(x_1, ..., x_n) : \sum_{i=1}^{n} x_i = 2\right\}.$$

Risp. R_1 ☒ R_2 ☐ R_3 ☐

(b) Fornisci una motivazione teorica per la tua scelta.

Risp. Teo di KARLIN-RUBIN: i.e. modello di Poisson ha rapporto delle veros. monotono (cresc) in $\Sigma X_i = T$

OVVERO = $\forall\ \theta_1, \theta_2,\ \theta_2 > \theta_1,\ \lambda_{21}(t)$ è ↑ in T

$\Rightarrow$ i Test UMP rifiutano H_0 con $\Sigma X_i \leq K$

(c) Spiega come è possibile derivare per la regione selezionata la probabilità di commettere un errore di prima specie e la potenza del test in corrispondenza di $\theta = 1$ e calcola numericamente le due quantità α e $1 - \beta$.

$P_\theta[Y=y] = \dfrac{e^{-n\theta}(n\theta)^y}{y!}$

Risp. $\eta(\theta)\ \ P_\theta[R] = P_\theta[\Sigma X_i \leq 2]$ con $Y = \sum_{i=1}^{3} X_i \mid \theta \sim \text{Poisson}(n\theta)$

$= P_\theta(Y \leq 2) = P_\theta[Y=0] + P_\theta[Y=1] + P_\theta[Y=2]$

$e^{-n\theta} + e^{-n\theta}(n\theta)^1 + \dfrac{e^{-n\theta}}{2}(n\theta)^2$

$\eta(\theta_0) = \alpha$ ma pozze $\theta = \theta_0 = 1$

3. Sia $X \sim f_X(x) = 2e^{-2x}$ per $x > 0$. Verificare che il quantile della distribuzione di X a livello $\alpha = 0.5$ vale 0.346.

Risp. $q : \int_0^q f_X(x)\,dx = 0.5 \leftrightarrow \int_0^q 2e^{-2x}\,dx = 0.5$

$-e^{-2x}\Big|_0^q = 1/2 \leftrightarrow 1 - e^{-2q} = 1/2 \leftrightarrow e^{-2q} = 1/2$

$-2q = -\ln 2 \leftrightarrow q = \frac{1}{2}\ln 2$

4. Sia $X_1, ..., X_n$ un campione casuale con $X_i \sim N(0, \sigma^2)$ e sia $T(\mathbf{X}_n) = \frac{1}{n+1}\sum_{i=1}^{n} X_i^2$.

(a) Mostrare che $T(\mathbf{X}_n) \sim \text{Gamma}\left(\texttt{shape} = \frac{n}{2}, \texttt{rate} = \frac{n+1}{2\sigma^2}\right)$ e

Risp. $T = \frac{n}{n+1} S_0^2 \qquad S_0^2 \sim Ga^R\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right) \Rightarrow \frac{n}{n+1} S_0^2 \sim Ga^R\left(\frac{n}{2}, \frac{n+1}{\cancel{n}} \frac{\cancel{n}}{2\sigma^2}\right)$

$\qquad \| \sim Ga^R\left(\frac{n}{2}, \frac{n+1}{2\sigma^2}\right)$

(b) Mostrare che $T(\mathbf{X}_n)$ è uno stimatore consistente in errore quadratico medio.

Risp. $E_\theta[T] = \frac{n}{2}\frac{2\sigma^2}{n+1} = \frac{n+1}{n}\sigma^2 \xrightarrow[n]{} \sigma^2 \Rightarrow$ ASINT NON DIST $\forall \sigma^2$

$V_\theta[T] = \frac{n}{2}\frac{4\sigma^4}{(n+1)^2} = \frac{n}{(n+1)^2}2\sigma^2 \xrightarrow[n]{} 0$

$\Rightarrow mse_\theta[T] \xrightarrow[n]{} 0 \quad \forall \sigma^2 \Rightarrow$ CONSIST in EQM

Parte B - Problema. Sia $\mathbf{X}_n = (X_1, \dots X_n)$ un campione i.i.d. da una popolazione con funzione di densità

$$f_X(x;\theta) = \frac{2}{\theta}\, x\, e^{-\frac{x^2}{\theta}}, \qquad x > 0, \quad \theta > 0$$

e con $\mathbb{E}[X] = a\theta$, $\mathbb{E}[X^2] = \theta$, $\mathbb{V}[X] = b\theta$, dove $a = \Gamma(3/2)$ e $b = 1 - \Gamma(3/2)^2$.

1. Verificare che il modello per la v.a. X costituisce una famiglia esponenziale.

Risp.

$f_X(x,\theta) = 2x \exp\left\{-\frac{x^2}{\theta} - \ln\theta\right\} = h(x)\exp\{\eta(\theta)T(x) - B(\theta)\}$

con

$h(x) = 2x \quad T(x) = x^2 \quad \eta(\theta) = -1/\theta \quad B(\theta) = +\ln\theta$

$\Rightarrow$ è F.E.

NB = $T_n(x_n) = \sum x_i^2$ è STAT SUFF e COMPL

2. Verificare che θ **non** è un parametro di scala per il modello.

Risp. porre $\theta = 1$

$f(x) = 2x e^{-x^2}$

$g_s(x;\theta) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right) = \frac{2}{\theta}\frac{x}{\theta} e^{-\left(\frac{x}{\theta}\right)^2}$

$= \frac{2x}{\theta^2} e^{-x^2/\theta^2} \neq f_X(x;\theta)$

3. Sia $f(x) = f_X(x;\theta = 1)$ una funzione di densità standard. Determinare $g_s(x;\sigma)$, la funzione di densità della famiglia scala ottenuta con $\sigma > 0$ da $f(x)$.

Risp.

VEDI SOPRA

$f_X(x;\sigma) = \frac{2x}{\sigma^2} e^{-\frac{x^2}{\sigma^2}}, \quad x > 0 \quad \sigma^2 > 0$

4. Considerare da ora il modello con funzione di densità $f_X(x;\theta)$. Determinare la funzione di verosimiglianza $L(\theta)$, una statistica sufficiente minimale $T(\mathbf{X}_n)$ e la stima di massima verosimiglianza $\widehat{\theta}_{mv}(\mathbf{x}_n)$ del parametro θ.

Risp. $L(\theta) = \theta^{-n} e^{-T/\theta}$ $\theta > 0$; la STAT SUFF MIN per $T(x_n) = \sum x_i^2$) il CRIT di FATT è

$\ell(\theta) = -n \ln \theta - T/\theta \Rightarrow \ell'(\theta) = -n/\theta + T/\theta^2 = 0$

$\Leftarrow\Rightarrow \hat{\theta}_{mv} = T/n$

$\ell''(\theta) = +n/\theta^2 - 2\theta T/\theta^3$ $\quad \ell''(\theta)\Big|_{\hat{\theta}_{mv}} = +\frac{n^2}{T^2} - \frac{2n^3}{T^2} < 0$

5. Determinare informazione osservata e insieme di verosimiglianza di livello q per il parametro θ.

Risp. $I_n^{oss} = -\ell''(\theta)\Big|_{\theta=\hat{\theta}_{mv}} \Rightarrow I_n^{oss} = +n^3/T^2$

$L_q = \hat{\theta}_{mv} \pm k_q \left(I_n^{oss}\right)^{-1/2}$ con $\hat{\theta}_{mv} = T/n$

$I_n^{oss} = n^3/T$

$= T/n \pm k_q \dfrac{T/n}{\sqrt{n}}$ $\quad K = \sqrt{-2\ln q}$

6. Verificare (giustificando la risposta) che la statistica campionaria $m_2(\mathbf{X}_n) = \frac{\sum_{i=1}^n X_i^2}{n}$ è UMVUE di θ.

Risp. $\mathbb{E}[m_2] = \frac{1}{n} \times n\, \mathbb{E}(X_i^2) = \theta \quad \forall\theta \Rightarrow$ m_2 è ND per θ

Quindi m_2 è

- ND per θ
- fne di STAT SUFF e COMPLETA

$\Rightarrow$ (propr delle FAM ESP)

$\Rightarrow$ m_2 è UMVUE di θ

7. Determinare l'espressione di $\widehat{\theta}_m(\mathbf{X}_n)$, stimatore dei momenti di θ, e calcolare la sua varianza.

Risp. $\mathbb{E}_\theta[X] = \bar{X}_n \Rightarrow a\theta = \bar{X}_n \Rightarrow \boxed{\hat{\theta}_n = \frac{\bar{X}_n}{a}}$

$\mathbb{V}_\theta[\hat{\theta}_n] = \frac{1}{a^2}\mathbb{V}_\theta(\bar{X}_n) = \frac{b\,\theta}{n a^2}$

8. Determinare l'intervallo di confidenza di livello $1-\alpha$ per θ che si ottiene utilizzando la distribuzione asintotica di $\widehat{\theta}_m(\mathbf{X}_n)$.

Risp. $\hat{\theta}_M = \frac{\bar{X}_n}{a} \mathrel{\dot\sim} N\left(\theta, \frac{b}{a^2}\frac{\theta}{n}\right)$

$\Rightarrow IC = \hat{\theta}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{b}}{a} \frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}}$