Tema di Statistica Matematica

Simulazione

1) Sia $(X_1, \ldots, X_n)$ un campione casuale estratto da una distribuzione $N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ e $\sigma^2 > 0$, quantitá finite ma incognite.

a) Determinare la distribuzione della statistica

$$T_n = \frac{1}{6}\sum_{i=1}^{3} X_i + \frac{1}{2\,(n-3)}\sum_{i=4}^{n} X_i.$$

b) Qual é il minimo valore ammissibile per l'ampiezza campionaria n?

c) La statistica T_n é stimatore efficiente per il parametro μ?

2) Sia $(X_1, \ldots, X_n)$ un campione casuale estratto da una distribuzione Normale di parametri $\mu \in \mathbb{R}$ e $\sigma^2 > 0$, con σ^2 quantitá nota.

a) Trovare lo stimatore $UMVU$ di $g(\mu) = e^{t\mu}$ con $t \neq 0$ quantitá fissata.

b) Mostrare che la varianza dello stimatore $UMVU$ trovato al punto precedente é maggiore del limite inferiore di Rao-Cramér per la varianza di un qualsiasi stimatore non disorto di $g(\mu) = e^{t\mu}$ e che il rapporto tra la varianza dello stimatore $UMVU$ trovato e il sudddetto limite inferiore di Rao-Cramér converge a 1 quando $n \to \infty$.

3) Siano $(X_1, \ldots, X_n)$ variabili casuali indipendenti e identicamente distribuite secondo una legge esponenziale (traslata) di parametri $(a, \lambda) \in \Theta = \mathbb{R} \times \mathbb{R}^+$, la cui densitá é data da

$$f_X(x; a, \lambda) = \lambda\, e^{-\lambda\,(x-a)}\, \mathbb{I}_{(a,\infty)}(x).$$

a) Dimostrare che $V_n = min(X_1, \ldots, X_n)$ ha legge di distribuzione esponenziale (traslata) di parametri $(a, n\lambda)$.

b) Fissato a, la distribuzione da cui proviene il campione casuale appartiene a famiglia esponenziale a $k = 1$ parametri? E se fosse λ a essere fissato, cosa concludere in merito all'appartenenza a famiglia esponenziale? Giustificare le risposte date.

c) Assumendo noto il parametro a, trovare una statistica sufficiente e completa per λ e determinare lo stimatore di massima verosimiglianza di λ.

d) Sempre assumendo noto il parametro a, trovare lo simatore $UMVU$ di λ. Lo stimatore $UMVU$ trovato é anche efficiente?

4) Per un campione casuale di ampiezza $n = 15$, proveniente da una distribuzione Normale, si sono ottenuti i seguenti valori di sintesi:

$$\sum_{i=1}^{15} x_i = 40.5 \quad \text{e} \quad \sum_{i=1}^{15} x_i^2 = 181.7$$

Dopo aver individuato le opportune statistiche pivot e le relative distribuzioni, costruire:

a) un intervallo di confidenza di livello 0.95 per la media μ della popolazione. Come cambierebbe l'intervallo di confidenza se la varianza della popolazione fosse nota e uguale a 9?

b) un intervallo di confidenza di livello 0.99 per la varianza σ^2 della popolazione.

5) Da una distribuzione Normale di parametri μ e $\sigma^2 = 3$ viene estratto un campione casuale di ampiezza $n = 3$ per sottoporre a verifica, sulla base dell'informazione in esso contenuta circa il parametro μ, il seguente sistema di ipotesi

$$H_0 : \mu = 2 \quad \text{vs.} \quad H_1 : \mu = 1$$

tramite un test che ha la seguente regione critica

$$C_\alpha = \{(x_1, x_2, x_3) \in \mathcal{X} : 2x_1 - 2x_2 + x_3 < 1.2\}.$$

Calcolare:

a) la probabilitá di errore di primo tipo;

b) la probabilitá di errore di secondo tipo;

c) la potenza del test.

Il test associato alla regione critica C_α di cui sopra é il test piú potente fra quelli di livello α per il dato sistema di ipotesi?

6) Nel corso di uno studio sulle proprietá di alcune leghe metalliche si é ipotizzata una relazione lineare tra la concentrazione di carbonio (x) e la tensione di snervamento (Y) delle stesse; sulla base di questa assunzione si é costruito un modello di regressione lineare $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ che é poi stato stimato sulla base di un campione di 9 osservazioni (x_i, Y_i), ottenendo

$$\hat{Y}_i = 3.858 + 1.3925\, x_i$$

Sapendo che la varianza di b_1, stimatore a minimi quadrati di β_1, é stata stimata pari a 0.025, verificare la significativitá (statistica) del coefficiente di regressione β_1.

EX.1)

a) $T_n = \frac{1}{6}\boxed{\sum_{i=1}^{3} X_i} + \frac{1}{2(n-3)}\boxed{\sum_{i=4}^{n} X_i}$

$= \underbrace{\sum_{i=1}^{3} \frac{1}{6} X_i + \sum_{i=4}^{n} \frac{1}{2(n-3)} X_i}$

$= \frac{1}{6} V_1 + \frac{1}{2(n-3)} V_2$

dove $V_1 = \sum_{i=1}^{3} X_i \sim N(3\mu, 3\sigma^2)$ e $V_2 = \sum_{i=4}^{n} X_i \sim N((n-3)\mu, (n-3)\sigma^2)$

e $V_1 \perp V_2$. Ma allora T_n è CLIN di v.c. NORMALI INDIPENDENTI rispettivamente con pesi $w_1 = \frac{1}{6}$ e $w_2 = \frac{1}{2(n-3)}$ sicché

$$T_n = \frac{1}{6} V_1 + \frac{1}{2(n-3)} V_2 \sim N\left(\mu, \frac{n\sigma^2}{12(n-3)}\right)$$

con

$$\mu_{T_n} = E(T_n) = \frac{1}{6} \cdot E(V_1) + \frac{1}{2(n-3)} E(V_2) = \frac{3\mu}{6} + \frac{(n-3)\mu}{2(n-3)}$$

$$= \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

$$\sigma^2_{T_n} = Var(T_n) \overset{V_1 \perp V_2}{=} \left(\frac{1}{6}\right)^2 Var(V_1) + \frac{1}{[2(n-3)]^2} Var(V_2)$$

$$= \frac{1}{36} \cdot 3\sigma^2 + \frac{1}{[2(n-3)]^2} \cdot (n-3)\sigma^2$$

$$= \frac{1}{12}\sigma^2 + \frac{1}{4(n-3)}\sigma^2 = \frac{(n-3)\sigma^2 + 3\sigma^2}{12(n-3)} = \frac{n\sigma^2}{12(n-3)}$$

con $n \geq 4$ condizione di esistenza della varianza di T_n

b) Per la condizione di esistenza sulla varianza, $n \geq 4$.

c) La statistica T_n è stimatore NON DISTORTO di μ; infatti

$$\mathbb{E}(T_n) = \mathbb{E}\left[\frac{1}{6}\sum_{i=1}^{3} X_i + \frac{1}{2(n-3)}\sum_{i=4}^{n} X_i\right]$$

$$= \frac{1}{6}\cdot 3\,\mathbb{E}(X_i) + \frac{1}{2(n-3)}(n-3)\,\mathbb{E}(X_i)$$

$$= \frac{1}{2}\mu + \frac{1}{2}\mu = \mu, \quad \forall \mu \in \mathbb{R}$$

d) Dato un qualsiasi stimatore NON DISTORTO di μ, V_n, si ha che

$$Var(V_n) \geq \frac{1}{I_{11}(\mu,\sigma^2)} \qquad \text{con } I(\mu,\sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{\sigma^4} \end{bmatrix}$$

matrice di INFORMAZIONE di Fisher per $N(\mu,\sigma^2)$

$$\geq \frac{\sigma^2}{n}$$

Ora

$$Var(T_n) = \frac{n\sigma^2}{12(n-3)} = \frac{n^2}{12(n-3)}\cdot\frac{\sigma^2}{n}$$

e per $n \geq 4$ si ha che

$$\frac{n^2}{12(n-3)} \geq 1$$

sicché T_n NON è STIMATORE EFFICIENTE per μ, a meno che $n=6$ nel qual caso esso COINCIDE con la MEDIA CAMPIONARIA $\bar{X}_n$.

Ex. 2)

a) Considerato che il campione proviene da una distr. $N(\mu, \sigma^2)$ con $\sigma^2 > 0$ quantità NOTA, la media CAMPIONARIA $\bar{X}_n$ è una statistica SUFFICIENTE e COMPLETA per μ.

Ora, il problema di interesse verte sul trovare uno stimatore UMVU per $g(\mu) = e^{t\mu}$ con $t \neq 0$ quantità nota.

Ricordando che $\hat{\mu}_n = \bar{X}_n$ è lo stim.re di MV di μ e, per la proprietà di INVARIANZA degli stimatori di MV, lo stim.re di MV per $g(\mu)$ è

$$g(\hat{\mu}_n) = e^{t\bar{X}_n}$$

che è FUNZIONE di STATISTICA SUFF e COMPLETA per μ data da $\bar{X}_n$. Ora si tratta di verificare se $g(\hat{\mu}_n)$ è stim.re NON DISTORTO di $e^{t\mu}$; ora

$$\mathbb{E}(e^{t\bar{X}_n}) = \mathbb{E}\left(e^{\frac{t}{n}\sum_{i=1}^{n} X_i}\right)$$

$$= \prod_{i=1}^{n} \mathbb{E}\left(e^{\frac{t}{n}X_i}\right) = \left[e^{\frac{t}{n}\mu + \frac{t^2}{n^2}\frac{\sigma^2}{2}}\right]^n$$

$$= e^{t\mu + \frac{t^2\sigma^2}{2n}} \neq e^{t\mu}$$

sicché $e^{t\bar{X}_n}$ è stimatore DISTORTO di $e^{t\mu}$ ma con una semplice trasformazione si ottiene uno stimatore NON DISTORTO da qs. ultimo:

$$T_n = e^{t\bar{X}_n} \cdot e^{-\frac{t^2\sigma^2}{2n}} = e^{t\bar{X}_n - \frac{t^2\sigma^2}{2n}}$$

che risulta essere lo stimatore UMVU di $g(\mu) = e^{t\mu}$, essendo stim.re NON DISTORTO, funzione di statistica SUFF e COMPLETA

b) Dato che il campione proviene da $N(\mu, \sigma^2)$, l'INFORMAZIONE di FISHER per μ è data da $I_n(\mu) = \frac{n}{\sigma^2}$.
La funzione $g(\mu) = e^{t\mu}$ di cui interessa uno stimatore UMVU è continua e derivabile in μ sicché il LIMITE INFERIORE di RAO-CRAMÈR per la VARIANZA di un qualsiasi stimatore NON DISTORTO V_n di $g(\mu)$ è data da

$$Var(V_n) = \frac{\left(\frac{d}{d\mu} e^{t\mu}\right)^2}{I_n(\mu)} = \frac{t^2 e^{2t\mu}}{n/\sigma^2} = \frac{\sigma^2}{n} \cdot t^2 e^{2t\mu}$$

$$= \frac{\sigma^2 t^2}{n} \cdot e^{2t\mu}$$

mentre

$$Var(T_n) = Var\left(e^{t\bar{X}_n} e^{-\frac{t^2\sigma^2}{2n}}\right)$$

$$= \left(e^{-\frac{t^2\sigma^2}{2n}}\right)^2 \cdot Var\left(e^{t\bar{X}_n}\right)$$

$$= e^{-\frac{t^2\sigma^2}{n}} \left\{ \mathbb{E}\left[\left(e^{t\bar{X}_n}\right)^2\right] - \left[\mathbb{E}\left(e^{t\bar{X}_n}\right)\right]^2 \right\}$$

$$= e^{-\frac{t^2\sigma^2}{n}} \left\{ \mathbb{E}\left(e^{2t\bar{X}_n}\right) - \left(e^{t\mu + \frac{t^2\sigma^2}{2n}}\right)^2 \right\}$$

$$= e^{-\frac{t^2\sigma^2}{n}} \left\{ e^{2t\mu + \frac{4t^2\sigma^2}{2n}} - e^{2t\mu + \frac{2t^2\sigma^2}{2n}} \right\}$$

$$= e^{-\frac{t^2\sigma^2}{n}} \left\{ e^{2t\mu + \frac{2t^2\sigma^2}{n}} - e^{2t\mu + \frac{t^2\sigma^2}{n}} \right\}$$

$$= e^{-\frac{t^2\sigma^2}{n}} \cdot e^{2t\mu + \frac{t^2\sigma^2}{n}} \left(e^{\frac{t^2\sigma^2}{n}} - 1\right)$$

$$= e^{2t\mu} \left(e^{\frac{t^2\sigma^2}{n}} - 1\right) > \frac{\sigma^2 t^2}{n} \cdot e^{2t\mu}$$

In fine

$$\frac{Var(T_n)}{\dfrac{\left(\frac{d}{d\mu} e^{t\mu}\right)^2}{I_n(\mu)}} = \frac{\cancel{e^{2t\mu}}\left(e^{\frac{t^2\sigma^2}{n}} - 1\right)}{\cancel{e^{2t\mu}} \cdot \frac{\sigma^2 t^2}{n}} \xrightarrow[n\to\infty]{} 1$$

poiché

$$\lim_{x\to 0} \frac{e^x - 1}{x} = 1.$$

Es. 3)

$$f_X(x_i; a, \lambda) = \lambda e^{-\lambda(x_i - a)} \mathbb{1}_{(a,+\infty)}(x_i), \quad \lambda > 0, \quad a \in \mathbb{R}$$

a) $V_n = \min(X_1, \dots, X_n) = X_{(1)}$

$$f_{X_{(m)}}(x) = m!\binom{n}{m} \cdot f_X(x) \cdot [F_X(x)]^{m-1} [1 - F_X(x)]^{n-m}$$

con

$$F_X(x) = F_X(x; a, \lambda) = \int_a^x \lambda e^{-\lambda(w-a)}\, dw = -e^{-\lambda(w-a)}\Big|_a^x$$

$$= 1 - e^{-\lambda(x-a)}$$

quindi, posto $m = 1$

$$f_{X_{(1)}}(x; a, \lambda) = 1!\binom{n}{1} \lambda e^{-\lambda(x-a)} \left[1 - \left(1 - e^{-\lambda(x-a)}\right)\right]^{n-1}$$

$$= n \lambda e^{-\lambda(x-a)} e^{-\lambda(x-a)(n-1)}$$

$$= (n\lambda) e^{-(n\lambda)(x-a)}$$

ossia

$$X_{(1)} \sim Exp(a, n\lambda)$$

b)

$$L(a, \lambda; \underline{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda(x_i - a)} \mathbb{1}_{(a,+\infty)}(x_i)$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i + \lambda n a} \prod_{i=1}^{n} \mathbb{1}_{(a,+\infty)}(x_i)$$

da cui

$$l(a,\lambda;\underline{x}) = \ln L(a,\lambda;\underline{x}) = n\ln(\lambda) - \lambda\sum_{i=1}^{n} x_i + \lambda n a +$$
$$+ \ln\left(\prod_{i=1}^{n} \mathbb{1}_{(a,+\infty)}(x_i)\right)$$

Assumendo a NOTO,

$$L(a,\lambda;\underline{x}) = \underbrace{\prod_{i=1}^{n} \mathbb{1}_{(a,+\infty)}(x_i)}_{C(\underline{x})} \cdot \underbrace{\lambda^n}_{D(\lambda)} \exp\Big\{\underbrace{-\lambda}_{A(\lambda)} \underbrace{\sum_{i=1}^{n} x_i}_{B(\underline{x})}\Big\}$$

sicché la dn appartiene a fam. ESPONENZIALE a $k=1$ parametri e la statistica

$$T_n(X_1,\dots,X_n) = \sum_{i=1}^{n} X_i$$

è SUFFICIENTE (minimale) e COMPLETA per λ.

Ora, noto λ, la dn dipende per quel che riguarda il SUPPORTO sul parametro a sicché risulta impossibile ricondurla alla parametrizzazione necessaria implicata dall'appartenenza a famiglia esponenziale. Nonostante questo

$$L(a,\lambda;\underline{x}) = \lambda^n \exp\Big\{-\lambda\sum_{i=1}^{n} x_i\Big\} e^{n\lambda a} \prod_{i=1}^{n} \mathbb{1}_{(a,+\infty)}(x_i)$$

$$= \underbrace{\lambda^n \exp\Big\{-\lambda\sum_{i=1}^{n} x_i\Big\} \cdot \prod_{i=1}^{n} \mathbb{1}_{\mathbb{R}}(x_i)}_{h(\underline{x})} \cdot \underbrace{e^{n\lambda a}\, \mathbb{1}_{[0,x_{(1)}]}(a)}_{g(a,\,x_{(1)})}$$

sicché la statistica

$$W_n = X_{(1)}$$

è statistica SUFFICIENTE minimale per a via Thm di FATTORIZZAZIONE.

Inoltre sia $h(W_n) = h(X_{(1)})$ una qualsiasi funzione di $X_{(1)}$.

Allora essendo $X_{(1)} \sim Exp(a, n\lambda)$

$$\mathbb{E}_a(h(X_{(1)})) = \int_a^{+\infty} h(x_{(1)}) \cdot (n\lambda) e^{-(n\lambda)(x_{(1)}-a)} dx_{(1)}$$

$$= \underbrace{(n\lambda)}_{>0} \underbrace{e^{n\lambda a}}_{>0} \int_a^{+\infty} h(x_{(1)}) \cdot \underbrace{e^{-n\lambda x_{(1)}}}_{>0} dx_{(1)}$$

sicché $\mathbb{E}_a(h(X_{(1)})) = 0$ IMPLICA che $h(X_{(1)}) = 0$ q.c.

da cui discende che $W_n = X_{(1)}$ è STATISTICA COMPLETA per il parametro a, oltre che SUFFICIENTE (minimale).

Lo stimatore di MV di λ è dato dalla SOLUZIONE della seguente equazione in λ

$$\frac{d}{d\lambda} \ell(\lambda; \underline{x}) = \frac{d}{d\lambda}\left[n \ln(\lambda) - \lambda \sum_{i=1}^n x_i + n\lambda a + \ln\left(\prod_{i=1}^n \mathbb{1}_{(a,+\infty)}(x_i)\right)\right]$$

$$= \frac{n}{\lambda} - \sum_{i=1}^n x_i + na = 0$$

$$= n - \lambda\left(\sum_{i=1}^n x_i - na\right) = 0$$

da cui

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n X_i - na} = \frac{1}{\bar{X}_n - a}$$

$$= \frac{n}{\sum_{i=1}^n (X_i - a)} = \frac{n}{V} \qquad \text{con } V = \sum_{i=1}^n (X_i - a)$$

Dallo stim.re di MV si può ricavare lo stimatore UMVU per λ trovando una FUNZIONE di $\hat{\lambda}_n$ (ovvero di STAT. SUFF. e COMPLETA) che sia stimatore NON DISTORTO di λ.

Ora, via Thm di TRASFORMAZIONE si ha che

$$W_i = X_i - a \sim Exp(\lambda) = \mathcal{G}\left(1, \frac{1}{\lambda}\right)$$

e

$$V = \sum_{i=1}^{n} W_i \sim \mathcal{G}\left(n, \frac{1}{\lambda}\right)$$

per la PROPRIETÀ di RIPRODUCIBILITÀ della distribuzione Gamma.

Ma allora

$$\mathbb{E}_\lambda(\hat{\lambda}_n) = \mathbb{E}_\lambda\left[\frac{n}{\sum_{i=1}^{n}(X_i - a)}\right] = \mathbb{E}_\lambda\left[\frac{n}{V}\right]$$

$$= \int_0^{+\infty} \frac{n}{v} \frac{1}{\Gamma(n)} \lambda^n v^{n-1} e^{-\lambda v} \, dv$$

$$= \frac{n}{n-1} \cdot \lambda \int_0^{+\infty} \underbrace{\frac{\lambda^{n-1}}{\Gamma(n-1)} v^{(n-1)-1} e^{-\lambda v}}_{\text{f densità di } \mathcal{G}(n-1, \frac{1}{\lambda})} \, dv$$

$$= \frac{n}{n-1} \lambda \neq \lambda$$

Ma

$$\mathbb{E}_\lambda\left(\frac{n-1}{n} \hat{\lambda}_n\right) = \lambda, \quad \forall \lambda > 0$$

sicché per il COROLLARIO del Thm di RAO-BLACKWELL

$$T_n = \frac{n-1}{n} \cdot \hat{\lambda}_n = \frac{n-1}{n} \cdot \frac{n}{\sum_{i=1}^{n} X_i - na} = \frac{n-1}{\sum_{i=1}^{n} X_i - na}$$

è stimatore UMVU di λ.

Ora

$$Var(\hat{\lambda}_n) = \mathbb{E}\left(\frac{n}{V}\right)^2 - \left[\mathbb{E}\left(\frac{n}{V}\right)\right]^2$$

con

$$\mathbb{E}\left(\frac{n}{V}\right)^2 = \int_0^{+\infty} \frac{n^2}{v^2} \frac{1}{\Gamma(n)} \lambda^n v^{n-1} e^{-\lambda v} \, dv$$

$$= \frac{n^2 \lambda^2}{(n-1)(n-2)} \int_0^{+\infty} \underbrace{\frac{\lambda^{n-2}}{\Gamma(n-2)} v^{(n-2)-1} e^{-\lambda v}}_{f \text{ DENSITÀ } \mathcal{G}(n-2, \frac{1}{\lambda})} \, dv$$

sicché

$$Var(\hat{\lambda}_n) = \frac{n^2 \lambda^2}{(n-1)(n-2)} - \left(\frac{n}{n-1} \lambda\right)^2$$

$$= \frac{n^2 \lambda^2}{(n-1)^2 (n-2)}$$

e

$$Var(\underbrace{T_n}_{\text{stim.re UMVU di } \lambda}) = Var\left(\frac{n-1}{n} \hat{\lambda}_n\right) = \frac{(n-1)^2}{n^2} \cdot Var(\hat{\lambda}_n)$$

$$= \frac{(n-1)^2}{n^2} \cdot \frac{n^2}{(n-1)^2 (n-2)} \cdot \lambda^2$$

$$= \frac{\lambda^2}{n-2}$$

Essendo $I_n(\lambda) = -\mathbb{E}\left[\frac{d^2}{d\lambda^2} \ell(\lambda; \underline{x})\right] = \frac{n}{\lambda^2}$ si può subito notare che

$$Var(T_n) = \frac{\lambda^2}{n-2} > \frac{1}{I_n(\lambda)} = \frac{\lambda^2}{n}$$

sicché T_n non è STIM.RE EFFICIENTE (nel senso di Bahadur) di λ.

4)

$n = 15$

$$\sum_{i=1}^{15} x_i = 40,5 \qquad \Rightarrow \qquad \bar{x}_n = \frac{1}{15} \cdot 40,5 = 2,7$$

$$\sum_{i=1}^{15} x_i^2 = 181,7 \qquad s_n^2 = \frac{1}{15-1} \cdot 181,7 - \frac{15}{15-1} \cdot (2,7)^2 = 5,168$$

a) La statistica PIVOT per la costruzione di un intervallo di confidenza per la MEDIA μ con varianza NON NOTA è

$$t = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} \sim t_{n-1}$$

sicchè

$$IC_\mu(1-\alpha) : \left[\bar{X}_n - t_{n-1;\ \alpha/2} \frac{S_n}{\sqrt{n}} ,\ \bar{X}_n + t_{n-1;\ \alpha/2} \frac{S_n}{\sqrt{n}} \right]$$

Ora $1-\alpha = 0,95$ sicchè $t_{n-1;0,025} = t_{14;0,025} = 2,145$ sicchè

$$IC_\mu(0,95) : \left[2,7 - 2,145 \cdot \frac{\sqrt{5,168}}{\sqrt{15}} ,\ 2,7 + 2,145 \cdot \frac{\sqrt{5,168}}{\sqrt{15}} \right]$$

$$= [1,441 ,\ 3,959]$$

Se la varianza σ^2 fosse nota, $\sigma^2 = 9$, allora la statistica PIVOT sarà

$$Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

sicchè in luogo di $t_{n-1;\alpha/2}$ useremo $Z_{1-\alpha/2}$. Nel nostro caso $Z_{1-0,05/2} = Z_{0,975} = 1,96$ sicchè

$$IC_\mu(0,95) : \left[2,7 - 1,96 \sqrt{\frac{9}{15}} ,\ 2,7 + 1,96 \cdot \sqrt{\frac{9}{15}} \right]$$

$$= [1,182 ,\ 4,218]$$

b) la statistica pivot per la costruzione di un intervallo di confidenza per la varianza σ^2 è

$$\frac{(n-1) S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

sicché

$$IC_{\sigma^2}(1-\alpha): \left[\frac{(n-1) S_n^2}{\chi^2_{n-1;\alpha/2}} , \frac{(n-1) S_n^2}{\chi^2_{n-1;1-\alpha/2}} \right]$$

Ora $1-\alpha = 0,99$ sicché

$$\chi^2_{15-1;0,995} = 4,07$$

$$\chi^2_{15-1;0,005} = 31,32$$

e perciò

$$IC_{\sigma^2}(0,99): \left[\frac{(15-1)\cdot 5,168}{31,32} , \frac{(15-1)\, 5,168}{4,07} \right]$$

$$= [\, 2,310 \, , \, 17,776 \,]$$

Ex. 5)

la regione critica è definita in funzione di una STATISTICA test, nel nostro caso data da $T_n(X_1, X_2, X_3) = 2X_1 - 2X_2 + X_3$ quindi il primo passo da effettuare è determinare la DISTRIBUZIONE di T_n da cui dipendono, sotto H_0 e sotto H_1, le probabilità α e β di errore di 1° e 2° tipo, rispettivamente.
Poichè $X \sim N(\mu, \sigma^2 = 3)$, abbiamo che

$$T_n(X_1, X_2, X_3) = 2X_1 - 2X_2 + X_3 \sim N(\mu^*, \sigma^{2*} = 27)$$

poichè

$$\mu^* = E_\mu(T_n) = E(2X_1 - 2X_2 + X_3) = 2\mu - 2\mu + \mu = \mu$$

$$\sigma^{2*} = Var_\mu(T_n) = Var_\mu(2X_1 \; 2X_2 + X_3) = 4\sigma^2 + 4\sigma^2 + \sigma^2$$
$$= 9\sigma^2 = 9 \cdot 3 = 27$$

e tenuto conto del fatto che T_n è una COMBINAZIONE LINEARE di v.c. NORMALI i.i.d. (propr. di RIPRODUCIBILITÀ della distr NORMALE).

Ora

a) La probabilità α di commettere un errore di 1° tipo ossia di RIFIUTARE H_0 quando essa è VERA è

$$\alpha = P(T_n < 1.2 \mid \mu = 2)$$

$$= P\left(\frac{T_n - 2}{\sqrt{27}} < \frac{1.2 - 2}{\sqrt{27}}\right) = P(Z < -0{,}15) = 0.4404$$

b) Poiché la REGIONE CRITICA è

$$C_\alpha = \{ (x_1, x_2, x_3) \in \mathcal{X} : 2x_1 - 2x_2 + x_3 < 1.2 \}$$

La REGIONE di NON RIFIUTO (o ACCETTAZIONE) di H_0 sarà

$$\bar{C}_\alpha = \{ (x_1, x_2, x_3) \in \mathcal{X} : 2x_1 - 2x_2 + x_3 \geq 1.2 \}$$

Allora, per la definizione che ne abbiamo dato,

$$\beta = P(T_n \geq 1.2 \mid \mu = 1)$$

dove $T_n = T_n(X_1, X_2, X_3) = 2X_1 - 2X_2 + X_3$; sicché

$$\beta = P\left(\frac{T_n - \mu}{\sqrt{27}} \geq \frac{1.2 - \mu}{\sqrt{27}} \,\Big|\, \mu = 1 \right) \qquad \text{poiché } T_n \underset{H_1}{\sim} N(1, 27)$$

$$= P\left(Z \geq \frac{1.2 - 1}{\sqrt{27}} \right) = P(Z \geq 0.04) = 0.484$$

c) In questo caso la POTENZA del test è data da

$$\eta_{C_\alpha} \stackrel{d.}{=} 1 - \beta = 1 - 0.484 = 0.516$$

Da notare che il test corrispondente alla data regione critica NON è il test PIÙ POTENTE per il problema in questione in quanto NON COINCIDE con quello restituito dal LEMMA di NEYMAN-PEARSON per un dato livello α fissato (peraltro non è neanche costruito a partire dalla STATISTICA sufficiente per μ per il problema in questione che è data da $\sum_{i=1}^{n} X_i$)

EX. 6)

Qui si tratta di verificare il seguente sistema di IPOTESI costruito su β_1, coefficiente angolare della retta di regressione

$$\begin{cases} H_0: \beta_1 = 0 \\ \text{vs} \\ H_1: \beta_1 \neq 0 \end{cases}$$

fissato in α il LIVELLO di SIGNIFICATIVITÀ del test.

Come abbiamo avuto modo di vedere trattando del modello di regressione lineare, indicato con b_1 lo stimatore a minimi quadrati di β_1 e ipotizzato che la componente stocastica ε del modello segua una distribuzione $N(0, \sigma^2)$, si ha che

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}\right)$$

e quindi $E(b_1) = \beta_1$ mentre $Var(b_1) = \sigma_{b_1}^2 = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$.

Ma allora

$$\frac{b_1 - \beta_1}{\sigma_{b_1}} \sim N(0,1)$$

mentre indicato con $s_{b_1}^2 = \left(\frac{1}{n-2} SSE\right) / \sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ - con $\frac{1}{n-2} SSE$ stimatore della VARIANZA σ^2 della componente stocastica ε del modello - lo STIMATORE della varianza di b_1

si ha che

$$\frac{s_{b_1}^2}{\sigma_{b_1}^2} = \frac{\frac{SSE}{n-2} \cdot \frac{1}{\sum (x_i - \bar{x}_n)^2}}{\sigma_{b_1}^2} = \frac{1}{(n-2)\sum_{i=1}^{n}(x_i-\bar{x}_n)^2} \cdot \frac{SSE}{\sigma_{b_1}^2} \sim \chi^2_{n-2}$$

dove $SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ è la somma dei QUADRATI dei RESIDUI di REGRESSIONE $(Y_i - \hat{Y}_i)$.

Infine

$$\frac{b_1 - \beta_1}{\sqrt{\frac{SSE}{(n-2)\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}}} \sim t_{n-2}$$

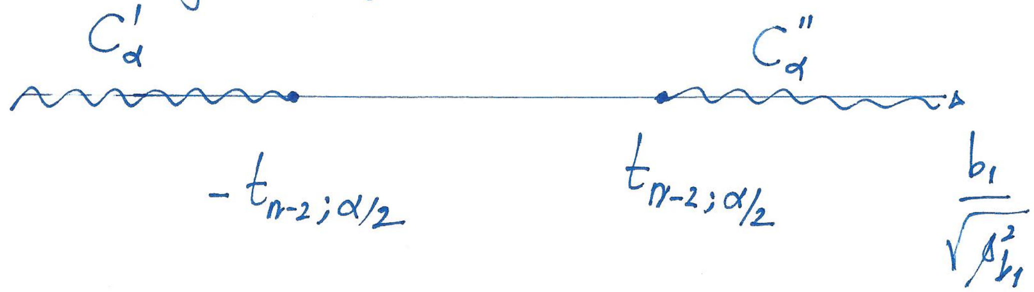
essendo il rapporto di una v.c. NORMALE STANDARD e della RADICE QUADRATA di una v.c. CHI-QUADRATO divisa per i suoi gradi di libertà, INDIPENDENTI.

Sulla base del risultato appena raggiunto, disponiamo della STATISTICA TEST per la VERIFICA dell'IPOTESI di SIGNIFICATIVITÀ del parametro β_1.

Sotto H_0 dunque

$$\frac{b_1 - 0}{\sqrt{s_{b_1}^2}} \sim t_{n-2}$$

sicché tenuto conto della forma dell'ipotesi alternativa avremo la seguente regione CRITICA BILATERALE



Avendo libertà di fissare α (e ricordando che di PROBABILITÀ di un ERRORE si tratta) poniamo $\alpha = 0{,}05$; di conseguenza $t_{9-2;0.025} = 2.365$ (dalle TAVOLE della distr t-STUDENT)

mentre

$$\frac{b_1}{\sqrt{s^2_{b_1}}} = \frac{1.3925}{\sqrt{0.025}} = 8{,}807$$

che cade a destra di 2.365 e quindi in C''_α, comportando il RIFIUTO di $H_0: \beta_1 = 0$ e portandoci a concludere a FAVORE della SIGNIFICATIVITÀ STATISTICA del parametro β_1.