Soluzioni del Tutorato di Statistica 1 del 13/12/2010
Docente: Prof.ssa Enza Orlandi
Tutore: Dott.ssa Barbara De Cicco

Esercizio 1.
Siano $X_1, ..., X_n$ v.a. distribuite come $N(\mu, \sigma^2)$ con μ nota.
Si vuole testare il test:

$$\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 = \sigma_1^2 \end{cases}$$

Trovare il test più potente di livello $\alpha = 0,05$ sapendo che $\sigma_1^2 > \sigma_0^2$.
Poichè le ipotesi sono semplici, il test è quello di Neyman-Pearson:
$\frac{L(\sigma_0^2)}{L(\sigma_1^2)} = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i-\mu)^2/2\sigma_0^2}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x_i-\mu)^2/2\sigma_1^2}} = \frac{\sigma_1}{\sigma_0} \frac{e^{-\sum_{i=1}^n (x_i-\mu)^2/2\sigma_0^2}}{e^{-\sum_{i=1}^n (x_i-\mu)^2/2\sigma_1^2}} \leq k$ Passando al logaritmo si ottiene:

$-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu)^2 + \frac{1}{2\sigma_1^2}\sum_{i=1}^n (x_i-\mu)^2 + log(\frac{\sigma_1}{\sigma_0}) \leq logk$
$\sum_{i=1}^n (x_i-\mu)^2 (\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}) \leq k_1$

Dove $k_1 = logk - log(\sigma_1/\sigma_0)$. Quindi:

$\sum_{i=1}^n (x_i-\mu)^2 \geq k^*$

Avendo osservato che $\sigma_1^2 > \sigma_0^2$ e che quindi la quantità: $(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}) < 0$.
Dove $k^* = \frac{k_1}{\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}}$.

Il test più potente di livello α è:
$0,05 = P(\sum_{i=1}^n (X_i-\mu)^2) \geq k^* | \sigma_0^2) = P(\sum_{i=1}^n (\frac{(X_i-\mu)}{\sigma_0})^2 \geq \frac{k^*}{\sigma_0^2} | \sigma_0^2) = P(U \geq \frac{k^*}{\sigma_0^2})$
Dove $U = \sum_{i=1}^n (\frac{(X_i-\mu)}{\sigma_0})^2 \sim \chi_n^2$ e quindi $\frac{k^*}{\sigma_0^2}$ è il quantile di una $\chi^2_{n,0.95}$.

Infatti si noti che sotto le ipotesi H_0, le X_i sono tutte v.a. $N(\mu, \sigma_0^2)$ e quindi $\frac{X_i-\mu}{\sigma_0} \sim N(0,1)$ da cui $\sum_{i=1}^n (\frac{(X_i-\mu)}{\sigma_0})^2 \sim \chi_n^2$.

Quindi il test più potente di livello $\alpha = 0,05$ è dato dal test con regione critica: $C^* = \{(x_1, ..., x_n) t.c. \sum_{i=1}^n (x_i-\mu)^2 \geq \sigma_0^2 \chi^2_{n,0.95}\}$.

Esercizio 2.
Sia $X_1, ..., X_n$ un c.c. dalla distribuzione $N(\mu, 5)$ e sia $n = 20$. Vogliamo testare $H_0: \mu_0 = 7$ contro $H_1: \mu_1 > 7$
Trovare il test uniformemente più potente di livello 0.05.
Per il test ricavato calcolare la funzione di potenza in $\mu = 7.5$; $\mu = 8$; $\mu = 8.5$; $\mu = 9$.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$f_X(x) = \frac{1}{\sqrt{10\pi}} e^{-\frac{1}{10}(x-\mu)^2} = \frac{1}{\sqrt{10\pi}} e^{-\frac{\mu^2}{10}} e^{-\frac{x^2}{10}} e^{\frac{\mu x}{5}}$

La distribuzione appartiene alla famiglia esponenziale allora applichiamo il teorema 9.5 del libro, da cui osserviamo che:
$c(\mu) = \frac{\mu}{5}$ e $d(x) = x$
La statistica da usare è $T = \sum_{i=1}^{n} x_i$ e poichè la funzione $c(\mu)$ è monotona crescente la regione critica sarà:
$C = \{(x_1, ..., x_n)|T > k\}$ con k t.c. l'errore di I specie sia pari ad α
$0.05 = \alpha = P(\sum_{i=1}^{n} x_i > k|\mu_0) = P(\frac{\sum_{i=1}^{n} x_i - \mu_0}{\sqrt{5n}} > \frac{k-n\mu_0}{\sqrt{5n}}|\mu_0) = \Phi\left(\frac{k-n\mu_0}{\sqrt{5n}}\right)$ da cui
$\frac{k-n\mu_0}{\sqrt{5n}} = z_{1-\alpha} = 1.65$ quindi $k = n\mu_0 + 1.65\sqrt{5n} = 156.5$
Quindi il test uniformemente più potente di livello α è dato dalla regione critica:
$C = \{(x_1, ..., x_n)|\bar{X} > 7.825\}$
La funzione di potenza è:
$\Pi_Y(\mu) = P(rifiutareH_0|\mu) = P(\bar{X} > 7.825|\mu) = P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{7.825-\mu}{\sigma/\sqrt{n}}) = \Phi\left(\frac{7.825-\mu}{\sqrt{5}/\sqrt{20}}\right)$
allora
$\Pi(7.5) = \Phi\left(\frac{7.825-7.5}{\sqrt{5}/\sqrt{20}}\right) = 0.2578$
$\Pi(8) = \Phi\left(\frac{7.825-8}{\sqrt{5}/\sqrt{20}}\right) = 0.6368$
$\Pi(8.5) = \Phi\left(\frac{7.825-8.5}{\sqrt{5}/\sqrt{20}}\right) = 0.9115$
$\Pi(9) = \Phi\left(\frac{7.825-9}{\sqrt{5}/\sqrt{20}}\right) = 0.9906$

Esercizio 3. (Esame 7 giugno 2010)
Siano $X_1, ..., X_n$ v.a. iid estratte da una popolazione $N(\mu, \sigma^2)$ con μ e σ^2 oncogniti.
Si vuole verificare $H_0 : \sigma^2 = 1$ contro $H_1 : \sigma^2 > 1$.
Si calcoli la zona di rifiuto R come la funzione $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}^2)$ per un test di ampiezza α usando il rapporto di verosomiglianza generalizzato.
Si dica inoltre cosa si intende per test uniformemente più potente di ampiezza α.
Il test trovato è un test uniformemente più potente? Motivare.

Vedere soluzioni A.A. 2009/2010 $http : //www.mat.uniroma3.it/didattica_interattiva/aa_09_10/st1/esgiugn$

Esercizio 4.
Sia $X_1, ..., X_n$ v.a. i.i..d. estratti da una popoloazione con funzione di densità:
$f(x, \theta) = \frac{1}{\theta^2} x e^{-x/\theta}$ con $x, \theta > 0$.

1. Valutare se la distribuzione appartiene alla famiglia esponenziale.
   La distribuzione appartiene alla famiglia esponenziale infatti possiamo scrivere:
   $a(\theta) = \frac{1}{\theta^2}$; $b(x) = x$; $c(\theta) = -\frac{1}{\theta}$; $d(x) = x$.
2. Trovare una statistica sufficiente.
   Poichè la distribuzione appartiene alla famiglia esponenziale allora
   $T(x) = \sum_{i=1}^{n} X_i$ è una statistica sufficiente.
3. Valutare se la media campionaria $\bar{X}$ è uno stimatore corretto per θ. Altrimenti correggerlo.
   $E[\bar{X}] = E[X] = \sum_0^{+\infty} \frac{1}{\theta^2} x^2 e^{-x/\theta} dx = \int_0^{+\infty} \theta y^2 e^{-y} dy = 2\theta$. Avendo eseguito il cambio di variabili: $x/\theta = y$.
   Quindi $\bar{X}$ è uno stimatore non corretto, per correggerlo sciegliamo $\hat{\theta^*} = \frac{\bar{X}}{2}$.

4. Calcolare l'MSE dello stimatore corretto calcolato al putno precedente.
   Poichè lo stimatore è corretto, l'errore quadratico medio è la varianza dello stimatore:
   $MSE_{\theta^*} = Var(\hat{\theta^*})$
   Calcoliamo quindi la varianza:
   $E[X^2] = \int_0^{+\infty} \frac{1}{\theta^2} x^3 e^{-x/\theta} dx = 6\theta^2$
   $Var(X) = \sigma^2 = 6\theta^2 - 4\theta^2 = 2\theta^2$
   Quindi:
   $Var\bar{X}/2 = \frac{1}{4} Var\bar{X} = \frac{1}{4}\frac{\sigma^2}{n} = \frac{\theta^2}{2n}$.

5. Calcolare lo stimatore di massima verosomiglianza di θ, vedere se è corretto. Trovare l'UMVUE.
   $L(\theta) = \prod_{i=1}^n \frac{1}{\theta^2} x_i e^{-x_i/\theta} 1_{(0,+\infty)}(x_i) = \frac{1}{\theta^{2n}} e^{-\frac{1}{\theta}\sum_i x_i}$
   Passando al logaritmo si ottiene:
   $logL(\theta) = -2nlog\theta - \frac{1}{\theta}\sum_{i=1}^n x_i$
   Derivando in θ si ha:
   $-\frac{2n}{\theta} + \frac{\sum_i x_i}{\theta^2} = 0$ da cui otteniamo che lo stimatore è: $\hat{\theta} = \bar{X}/2$ che per quanto visto al punto precedente è uno stimatore corretto.
   Per uno stimatore corretto, se la sua varianza coincide con il limite inferiore di Cramer-Rao, allora tale stimatore è anche un UMVUE.
   $Var(T) \geq \frac{(\tau')^2}{nE[(\frac{\partial}{\partial\theta} logf(x,\theta))^2]} = \frac{1}{nE[(-\frac{2}{\theta} + \frac{X}{\theta^2})^2]} = \frac{\theta^2}{2n}$
   Quindi $\frac{\bar{X}}{2}$ è l'UMVUE.

Esercizio 5.

Supponiamo che un segnale avente valore μ sia trasmesso da una locazione A e il valore ricevuto dalla locazione B è distribuito come una $N(\mu, 4)$. Quindi quando μ è inviato il valore ricevuto è $\mu + U$ dove $U \sim N(0, 4)$ e rappresenta il rumore nella trasmissione. Per ridurre l'errore supponiamo che il valore è stato inviato per 9 volte. Se i rispettivi valori ricevuti sono:
$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5$
cercare un intervallo di confidenza per μ al 95%.
Calcoliamo la media campionaria:
$\bar{X} = \frac{1}{9}\sum_{i=1}^9 x_i = 81/9 = 9$
Poichè la varianza è nota, scegliamo come quantità pivotale $Q = \frac{\bar{X}-\mu}{\sigma/9}$
$0,95 = P(-q_1 \leq Q \leq q_1) = P(-q_1 \leq \frac{\bar{X}-\mu}{\sigma/9} \leq q_1) = P(-q_1\frac{\sigma}{3} \leq \bar{X} - \mu \leq q_1\frac{\sigma}{3})$
Dalle tavole si ricava che $q_1 = 1,96$. Infatti q_1 è il quantile di una Normale di livello $0,975$.
Quindi l'intervallo di confidenza per μ al 95% è:
$(\bar{X} - q_1\frac{\sigma}{3}, \bar{X} + q_1\frac{\sigma}{3}) = (10,306; 7,694)$.