UNIVERSITÀ DEGLI STUDI ROMA TRE
Corso di Laurea in Matematica
ST410 - Statistica 1 - A.A. 2014/2015

II Esonero - 15 Gennaio 2015

1	2	3	4	5	6	Tot.

Avvertenza: Svolgere ogni esercizio nello spazio assegnato, senza consegnare altri fogli e accompagnando le risposte con spiegazioni complete, chiare ed essenziali. Scrivere il proprio nome su ogni foglio nello spazio predisposto. Non è consentito l'utilizzo di libri o appunti; non è consentito l'uso di calcolatrici.
Tempo: 2h.

COGNOME: .... **NOME:** ....

MATRICOLA: ....

Nome e Cognome:

Esercizio 1. Sia $X_1, \ldots, X_n$ un campione casuale di ampiezza n estratto da una popolazione con densità $f(x;\theta)$, $\theta \in \Theta$. Sia $T = T(X_1, \ldots, X_n)$ stimatore non distorto di θ.

1. (2pt) Definire cosa si intende col dire che "T è una statistica completa".
2. (2pt) Sia $S = S(X_1, \ldots, X_n)$ un altro stimatore non distorto di θ. Supponendo che T sia una statistica sufficiente e completa, dimostrare che $E(S|T) = T$ (con probabilità 1).
3. (3pt) Sia $n = 2$ e $f(x;\theta)$ la densità di una variabile aleatoria discreta, con $\theta \in \mathbb{N} \setminus \{0\}$ e $f(x;\theta) = 1/\theta$ se $x \in \{1, \ldots, \theta\}$, $f(x;\theta) = 0$ altrimenti. Sia $M = \max\{X_1, X_2\}$. Dimostrare che M è una statistica completa. (sugg.: calcolare prima di tutto la densità discreta di M, i.e., $h(m;\theta) = P_\theta(M = m) = \ldots$ e poi verificare la definizione).

Soluzione:

1. Cfr. lezione del 20 novembre 2014 o Mood–Graybill–Boes par. 7.5.2
2. Per il teorema di Lehmann–Scheffé T è stimatore UMVUE per θ. Siccome T è sufficiente, allora per il teorema di Rao–Blackwell anche $S' = E(S|T)$ è stimatore non distorto di θ e $Var_\theta(S') \leq Var_\theta(T)$ per ogni θ, dove la disuguaglianza deve essere stretta per almeno un valore di θ a meno che $S' = T$ con probabilità 1. Ma T è UMVUE, quindi $Var_\theta(S') = Var_\theta(T)$ per ogni θ e perciò $E(S|T) = T$ con probabilità 1.

   Oppure: $S' := E(S|T)$ è funzione di T; siccome T è una statistica sufficiente allora, per il teorema di Rao–Blackwell, anche S' è uno stimatore non distorto di θ. quindi $S' - T$ è funzione di T e $E(S' - T) = 0$ indipendentemente da θ. Siccome T è una statistica completa allora $P(S' = T) = 1$ per ogni $\theta \in \Theta$, i.e., $E(S|T) = T$.
3. $P_\theta(M = m) = P_\theta(X_1 = m, X_2 \leq m - 1) + P_\theta(X_1 \leq m, X_2 = m)$. Quindi $P_\theta(M = m) = 0$ se $m > \theta$, mentre è pari a $\frac{1}{\theta}\frac{m-1}{\theta} + \frac{m}{\theta}\frac{1}{\theta} = \frac{2m-1}{\theta^2}$ se $m \leq \theta$.
   Sia g funzione t.c. $g(M)$ sia una statistica. Assumiamo per ipotesi che $E_\theta(g(M)) = 0$ per ogni θ.
   $$E_\theta(g(M)) = \sum_{m=1}^{\theta} g(m) P_\theta(M = m) = \sum_{m=1}^{\theta} g(m) \frac{2m-1}{\theta^2}.$$
   Quindi $E_{\theta=1}(g(M)) = 0 \Rightarrow g(1) = 0$, $E_{\theta=2}(g(M)) = 0 \Rightarrow g(2) = 0$ etc. Per induzione $g(m) = 0$ per ogni m. Perciò $g(M) = 0$ e quindi $P_\theta(g(M) = 0) = 1$ per ogni θ. Quindi M è una statistica completa.

Nome e Cognome:

Esercizio 2. Sia X campione casuale di ampiezza 1 estratto da una densità esponenziale di parametro $\lambda > 0$, i.e. $f(x;\lambda) = \lambda e^{-\lambda x}$ con $x > 0$.

1. (2pt) Dimostare che $Q = \lambda X$ è una quantità pivotale.
2. (2pt) Utilizzare la Q del punto precedente per costruire un intervallo di confidenza al 100γ percento per λ e con estremo inferiore pari a 0.
3. (2pt) Calcolare esplicitamente l'estremo superiore dell'intervallo nel caso in cui la realizzazione campionaria sia $x = 10$ e $\gamma = 1 - \frac{1}{e}$.

Soluzione:

1. Per definizione una quantità pivotale è una variabile aleatoria funzione di X e λ ma la cui distribuzione non dipende dal parametro incognito λ. La funzione di distribuzione $F_Q(x) = P_\lambda(\lambda X \leq x) = P_\lambda(X \leq x/\lambda) = F_X(x/\lambda)$, quindi $f_Q(x) = f(x/\lambda;\lambda)\frac{1}{\lambda} = e^{-x}$, perciò la distribuzione di Q non dipende da λ, i.e., Q è una quantità pivotale.
2. Q è un'esponenziale di parametro 1. Siano $a \geq 0$ e $b > 0$ tali che $P(a < Q < b) = \gamma$. Allora $P\left(\frac{a}{X} < \lambda < \frac{b}{X}\right) = \gamma$, quindi $\left(\frac{a}{X}, \frac{b}{X}\right)$ è un intervallo di confidenza al 100γ percento per λ . Poniamo $a = 0$. Per ipotesi vogliamo che $P(0 < Q < b) = \gamma$, cioè che $\int_0^b e^{-x}dx = \gamma$, da cui $b = -\log(1-\gamma)$. Quindi $\left(0, \frac{-\log(1-\gamma)}{X}\right)$ è l'intervallo (aleatorio) di confidenza cercato.
3. Sostituendo nella formula precedente: $-\log(1-\gamma) = -\log(1/e) = 1$, quindi l'intervallo di confidenza (osservato) è $\left(0, \frac{1}{10}\right)$.

Nome e Cognome:

Esercizio 3. Il tempo totale necessario a fabbricare una banconota ha distribuzione normale di media μ incognita e varianza σ^2. In un esperimento in cui sono state prodotte 9 banconote, il tempo medio per la loro fabbricazione è risultato di 60 minuti, e la deviazione standard (campionaria) dei tempi è risultata pari a 3 minuti.

1. (2pt) Nel caso in cui non si hanno informazioni su σ^2, si determini un intervallo di confidenza al 95% per la media μ.

2. (2pt) Si supponga di sapere che $\sigma^2 = 9$. Si determini anche in questo caso un intervallo di confidenza al 95% per la media μ.

Per lo svolgimento dell'esercizio può essere utile conoscere alcuni di questi valori:
$t_{9,0.975} \cong -2.3; t_{8,0.975} \cong -2.3; t_{9,0.95} \cong -1.8; t_{8,0.95} \cong -1.9;$
$z_{0.975} \cong -2.0; z_{0.95} \cong -1.6;$
$\chi^2_{9,0.975} \cong 2.7; \chi^2_{8,0.975} \cong 2.2; \chi^2_{9,0.95} \cong 3.3; \chi^2_{8,0.95} \cong 2.7; \chi^2_{9,0.05} \cong 16.9; \chi^2_{8,0.05} \cong 15.5;$
$\chi^2_{9,0.025} \cong 19.0; \chi^2_{8,0.025} \cong 17.5$
dove, come al solito, $t_{n,\alpha}$ è il quantile di ordine $1-\alpha$ di una t di Student con n gradi di libertà; z_α è il quantile di ordine $1-\alpha$ di una normale standardizzata e $\chi^2_{n,\alpha}$ è il quantile di ordine $1-\alpha$ di una chi-quadro con n gradi di libertà.

Soluzione:

1. Per quanto visto nella lezione del 2 dicembre 2014, un intervallo di confidenza al 100γ percento per la media di una popolazione normale, se la varianza è incognita, è

$$\left(\overline{X} - t_{n-1,\frac{1-\gamma}{2}} \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{n-1,\frac{1-\gamma}{2}} \cdot \frac{S}{\sqrt{n}}\right),$$

dove n è l'ampiezza del campione, $\overline{X}$ la media campionaria, $S = \sqrt{S^2}$ la deviazione standard campionaria (i.e., la radice della varianza campionaria). Nel nostro caso $n = 9$ e, sulla realizzazione campionaria, $\overline{X}$ vale 60 e S vale 3. Inoltre $\gamma = 95\%$, perciò ci interessa conoscere $t_{8,0.025}$. Per la simmetria della distribuzione t di Student, $t_{8,0.025} = -t_{8,0.975}$, quindi l'intervallo di confidenza cercato è, svolgendo i calcoli, $(57.7, 62.3)$.

2. Per quanto visto nella lezione del 27 novembre 2014, un intervallo di confidenza al 100γ percento per la media di una popolazione normale, con varianza nota σ^2, è

$$\left(\overline{X} - z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right),$$

dove n è l'ampiezza del campione e $\overline{X}$ la media campionaria. Nel nostro caso $n = 9$ e, sulla realizzazione campionaria, $\overline{X}$ vale 60. Inoltre $\gamma = 95\%$, perciò ci interessa conoscere $z_{0.025}$. Per la simmetria della distribuzione normale, $z_{0.025} = -z_{0.975}$, quindi l'intervallo di confidenza cercato è, svolgendo i calcoli, $(58, 62)$.

Nome e Cognome:

Esercizio 4. Sia

$$f_0(x) = \begin{cases} 4x & \text{se } 0 < x < \frac{1}{2} \\ 4 - 4x & \text{se } \frac{1}{2} \le x < 1 \\ 0 & \text{se } x \le 0 \text{ o } x \ge 1 \end{cases}$$

e

$$f_1(x) = 1 \text{ se } 0 < x < 1 \text{ (0 altrimenti).}$$

Sia $f(x;\theta)$ (con $\theta \in \{0,1\}$) pari a $f_0(x)$ se $\theta = 0$ e pari a $f_1(x)$ se $\theta = 1$. Sia X un campione casuale di ampiezza 1 estratto da una popolazione di densità $f(x;\theta)$. Attraverso il campione si vuole verificare l'ipotesi $H_0 : \theta = 0$ in alternativa all'ipotesi $H_1 : \theta = 1$.

1. (2pt) Determinare il rapporto di verosimiglianza semplice $\lambda(x)$.

2. (2pt) Sia $k > 0$. Sia Y il test del rapporto di verosimiglianza semplice, di ampiezza α, così concepito: si rifiuti H_0 se $\lambda(x) < k$, si accetti H_0 altrimenti. Y è un test uniformemente più potente di ampiezza α?

3. (3pt) Determinare k in modo che α sia pari all'81%.

Soluzione:

1. Per definizione il rapporto di verosimiglianza semplice è il rapporto tra la funzioni di verosimiglianza $L_0(x)$,(i.e., $L(\theta;x)$ con $\theta = 0$) e $L_1(x)$ (i.e., $L(\theta;x)$ con $\theta = 1$). Quindi $\lambda(x) = \begin{cases} 4x & \text{se } 0 < x < \frac{1}{2} \\ 4 - 4x & \text{se } \frac{1}{2} \le x < 1 \end{cases}$, non è definito altrimenti.

2. Sia $C = \{x \text{ t.c. } 0 < x < 1 \text{ e } \lambda(x) < k\}$. Per definizione $C \subseteq (0,1)$ è la regione critica del test Y e $x \in C \Leftrightarrow \lambda(x) < k$. Quindi, banalmente, Y è un UMP-test per il lemma di Neyman–Pearson.

3. Per definizione l'ampiezza del test Y è $P_{\theta=0}(\lambda(X) < k) = P_{\theta=0}(4X < k, X < 1/2) + P_{\theta=0}(4 - 4X < k, X \ge 1/2) = P_{\theta=0}(X < \min\{k/4, 1/2\}) + P_{\theta=0}(X > \max\{1 - k/4, 1/2\})$.
   Supponiamo (come ragionevole) che $k \le 2$: abbiamo quindi che l'ampiezza del test è pari a $P_{\theta=0}(X < k/4) + P_{\theta=0}(X > 1 - k/4)$, cioè $\int_0^{k/4} 4x dx + \int_{1-k/4}^{1} (4 - 4x) dx$. Svolgendo i calcoli, l'ampiezza (in funzione di k) risulta $\frac{k^2}{4}$, quindi il k cercato è $k = 1.8$.

Nome e Cognome:

Esercizio 5. Sia $\{g(t;\theta)\}$, al variare di $\theta \in \Theta$, una famiglia di densità di probabilità per una variabile aleatoria univariata T (Θ intervallo $\subseteq \mathbb{R}$).

1. (2pt) Definire cosa si intende col dire che "T ha rapporto di verosimiglianza monotono non decrescente".
2. (3pt) Siano X_1, X_2 variabili aleatorie Bernoulliane, indipendenti e identicamente distribuite, di parametro $p \in [0,1]$. Dimostrare che $T = X_1 + X_2$ ha rapporto di verosimiglianza monotono non decrescente.
3. (3pt) Si enunci il teorema di Karlin–Rubin per la determinazione di test uniformemente più potenti per ipotesi composte unilaterali.

Soluzione:

1. Cfr. lezione del 9 dicembre 2014 o Mood–Graybill–Boes par. 9.3.2.
2. T ha distribuzione binomiale di parametri $n = 2$ e p. Quindi T ha famiglia di densità $g(x;p) = \binom{2}{x} p^x (1-p)^{2-x}$ con $x \in \{0,1,2\}$ e $p \in [0,1]$. Per $p_2 > p_1$ calcoliamo $r(x) = \frac{g(x;p_2)}{g(x;p_1)} = \frac{p_2^x(1-p_2)^{2-x}}{p_1^x(1-p_1)^{2-x}}$. Quindi $r(0) = \frac{(1-p_2)^2}{(1-p_1)^2}$, $r(1) = \frac{p_2(1-p_2)}{p_1(1-p_1)}$, $r(2) = \frac{p_2^2}{p_1^2}$. Dato che $p_2 > p_1$ allora $\frac{1-p_2}{1-p_1} < \frac{p_2}{p_1}$, perciò $r(0) \le r(1) \le r(2)$ e quindi T ha rapporto di verosimiglianza monotono non decrescente.
3. Cfr. lezione del 9 dicembre 2014 o Mood–Graybill–Boes par. 9.3.2.

Nome e Cognome:

Esercizio 6. La quantità di liquido erogata, ad ogni attivazione, da due distributori A e B di una stessa bevanda, si assume avere distribuzione normale di media μ_A e varianza σ_A^2 (per il distributore A) e di media μ_B e varianza σ_B^2 (per il distributore B). I parametri sono tutti incogniti. Si mettono in funzione i due distributori: su 10 bevande versate dal distributore A si ha un volume medio di 203 ml e una deviazione standard campionaria di 3 ml. Su 15 bevande versate dal distributore B si ha un volume medio di 206 ml e una deviazione standard campionaria di 5 ml.
Si vuole testare l'ipotesi $H_0 : \sigma_A^2 = \sigma_B^2$ in alternativa all'ipotesi $H_1 : \sigma_A^2 \neq \sigma_B^2$.

1. (2pt) Utilizzando i dati raccolti, si può rigettare l'ipotesi nulla ad un livello di significatività del 5%? (Descrivere il test di ipotesi utilizzato).

2. (2pt) Calcolare il p-value dei dati per il test del punto precedente.

Per lo svolgimento dell'esercizio può essere utile conoscere alcuni di questi valori:
$F_{10,15,0.975} \cong 0.28; F_{10,15,0.95} \cong 0.35; F_{10,15,0.05} \cong 2.54; ; F_{10,15,0.025} \cong 3.06; F_{9,14,0.975} \cong 0.26;$
$F_{9,14,0.95} \cong 0.33; F_{9,14,0.05} \cong 2.65; F_{9,14,0.025} \cong 3.21;$
$P(F_{10,15} \leq 0.36) \cong 0.05; P(F_{9,14} \leq 0.36) \cong 0.06; P(F_{10,15} \leq 0.6) \cong 0.21; P(F_{9,14} \leq 0.6) \cong 0.22;$
$t_{25,0.05} \cong 1.71; t_{25,0.025} \cong 2.06; t_{23,0.05} \cong 1.71; t_{23,0.025} \cong 2.07;$
$P(t_{25} \leq 1.7) \cong 0.95; P(t_{23} \leq 1.7) \cong 0.95;$
dove, come al solito, $F_{n,m}$ ha distribuzione F con n e m gradi di libertà e $F_{n,m,\alpha}$ è il quantile di $F_{n,m}$ di ordine $1-\alpha$; t_n ha distribuzione t di Student con n gradi di libertà e $t_{n,\alpha}$ è il quantile di t_n di ordine $1-\alpha$.

Soluzione:

1. Per quanto visto a lezione (16 dicembre 2014, oppure Mood-Graybill-Boes, par. 9.4.4) un possibile test di ampiezza α per verificare l'uguaglianza delle varianze delle due popolazioni è il seguente: si rifiuti H_0 se la statistica test $\frac{S_X^2}{S_Y^2}$, calcolata sui campioni, è maggiore di $F_{n-1,m-1,\alpha/2}$ o è minore di $F_{n-1,m-1,1-\alpha/2}$, dove S_X^2 e S_Y^2 sono le varianze (campionarie) del primo e del secondo campione, n è l'ampiezza del primo campione, m è l'ampiezza del secondo campione. Nel caso specifico: $S_X^2 = 9$, $S_Y^2 = 25$, $n = 10$, $m = 15$, perciò la stastistica test vale 0.36, e tale valore va confrontato con $F_{9,14,0.025} \cong 3.21$ e $F_{9,14,0.975} \cong 0.26$, quindi l'ipotesi nulla non si può rifiutare (ad un livello di significatività del 5%).

2. Il valore p dei dati è il livello di significatività α per cui, in base ai dati raccolti, si passa dall'accettazione al rifiuto dell'ipotesi nulla. È perciò quel valore α per cui $F_{9,14,\alpha/2} = 0.36$ o $F_{9,14,1-\alpha/2} = 0.36$. Sia $P := P(F_{9,14} \leq 0.36) \cong 0.06$, allora il p-value è uguale a $\min\{2P, 2 - 2P\} \cong 0.12$.