Corsi di Inferenza Statistica – 1° prova scritta – 9 giugno 2017

ATTENZIONE: SCRIVERE LE RISPOSTE SOLO SU QUESTI FOGLI.
EVENTUALI ALTRI FOGLI A QUADRETTI CONSEGNATI
NON VERRANNO CONSIDERATI

Cognome e Nome: ________________________ **Matricola e Corso Studi:** ____________

Esercizio A. Siano $X_1, ..., X_n|\theta \sim \text{Geom}(\theta)$ i.i.d., con valore atteso $\mathbb{E}_\theta[X] = \frac{1}{\theta}$, $\mathbb{V}_\theta[X] = \frac{1-\theta}{\theta^2}$ e funzione di massa di probabilità

$$f_X(x;\theta) = \theta(1-\theta)^{x-1}, \qquad x = 1,2,3,\dots \qquad \theta \in (0,1).$$

A1 Verificare che il modello appartiene alla famiglia esponenziale e determinare una statistica sufficiente.

A2 Determinare $L(\theta;\mathbf{x}_n)$ e $\widehat{\theta}_{mv}(\mathbf{x}_n)$, rispettivamente funzione di verosimiglianza e stima di massima verosimiglianza del parametro θ.

A3 Determinare lo stimatore del momenti di θ e verificare che coincide con $\widehat{\theta}_{mv}(\mathbf{x}_n)$.

A4 Verificare che l'informazione attesa di Fisher risulta uguale a $I_n(\theta) = \frac{n}{\theta^2(1-\theta)}$.

A5 Determinare $\widehat{\lambda}(\mathbf{X}_n)$, stimatore di massima verosimiglianza di $\lambda = g(\theta) = \frac{1}{\theta}$ e verificare che è non distorto per λ.

A6 Determinare $\mathbb{E}_\theta[(\widehat{\lambda}(\mathbf{X}_n) - \lambda)^2]$, errore quadratico medio di $\widehat{\lambda}(\mathbf{X}_n)$, e stabilire se lo stimatore è consistente per $\lambda = 1/\theta$.

A7 Determinare la distribuzione asintotica di $\widehat{\lambda}(\mathbf{X}_n)$ (in funzione del parametro θ).

A8 Verificare che $\widehat{\lambda}(\mathbf{X}_n)$ è stimatore UMVUE di λ utilizzando il limite inferiore di Cramer-Rao (per stimatori di funzioni $\tau(\theta)$ del parametro di base di un modello statistico).

A9 Giustificare il fatto che $\widehat{\lambda}(\mathbf{X}_n)$ sia UMVUE di λ utilizzando i teoremi di Rao-Balckwell e Lehmann-Scheffé.

A10 Supponendo che il campione abbia dimensione $n = 5$ e che $\sum_{i=1}^n x_i = 15$, calcolare il valore numerico di $\widehat{\theta}_{mv}(\mathbf{x}_n)$, di $\widehat{\lambda}(\mathbf{x}_n)$.

Esercizio B. Si consideri ancora un campione casuale $X_1, ..., X_n|\theta \sim \text{Geom}(\theta)$ i.i.d., con valore atteso $\mathbb{E}_\theta[X] = \frac{1}{\theta}$, $\mathbb{V}_\theta[X] = \frac{1-\theta}{\theta^2}$ e funzione di massa di probabilità

$$f_X(x;\theta) = \theta(1-\theta)^{x-1}, \qquad x = 1, 2, 3, \ldots \qquad \theta \in (0,1).$$

B1 Verificare che il modello ha effettivamente il rapporto delle verosimiglianze monotono rispetto alla statistica sufficiente minimale $T(\mathbf{X}_n)$ e stabilire se tale rapporto è crescente o decrescente in $T(\mathbf{X}_n)$.

B2 Considerare il sistema di ipotesi $H_0 : \theta = \theta_0 = 1/2$ vs. $H_1 : \theta = \theta_1 = 3/4$. Verificare che la regione di rifiuto di H_0 del test ottenuto utilizzando il Lemma di Neyman-Pearson è:

$$R = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \sum_{i=1}^{n} x_i > k \right\}, \qquad k > 0. \tag{1}$$

B3 Calcolare le probabilità di errore di I e II tipo e la potenza del test che si ottengono ponendo $k = 2$ e $n = 1$.

B4 Determinare l'intervallo di confidenza asintotico di livello $1 - \alpha = 0.95$ per il parametro $\lambda = g(\theta) = \frac{1}{\theta}$.

B5 Stabilire se è possibile accettare l'ipotesi $H_0 : \lambda = 3.3$ vs. $H_1 : \theta \neq 3.3$ in corrispondenza di un campione osservato in cui $n = 20$, $\sum_{i=1}^{n} x_i = 60$ e assumendo $\alpha = 0.05$.

Esercizio C. Sia $X_1, ..., X_n$ un campione casuale a componenti indipendenti e identicamente distribuite tali che ciascuna $X_i \sim \mathrm{N}(\theta, \sigma_0^2)$ con valore atteso $\mathbb{E}_\theta[X] = \theta$ incognito e varianza $\mathbb{V}_\theta[X] = \sigma_0^2 = 9$ nota. Si consideri il seguente sistema di ipotesi

$$\begin{cases} H_0: & \theta = \theta_0 = 2 \\ H_1: & \theta = \theta_1 = 4 \end{cases}$$

per il quale sono stati costruiti due test basati su statistiche campionarie diverse. Il primo test è basato sulla statistica *media campionaria* $S(X_1, ..., X_n) = \overline{X}_n$ ed ha la seguente forma

$$R_S(s) = \{(x_1, ..., x_n) \in \mathcal{X}_n : S(x_1, ..., x_n) > s\}$$

associata ad un generico valore di soglia s. Il secondo test è basato sulla statistica *mediana campionaria* $T(X_1, ..., X_n) = Med(X_1, ..., X_n)$ ed ha la seguente forma

$$R_T(t) = \{(x_1, ..., x_n) \in \mathcal{X}_n : T(x_1, ..., x_n) > t\}$$

associata ad un generico valore di soglia t. Per le due statistiche test sono rappresentate nella Figura 1 le distribuzioni campionarie esatte sotto l'ipotesi che $\theta = \theta_0 = 2$.

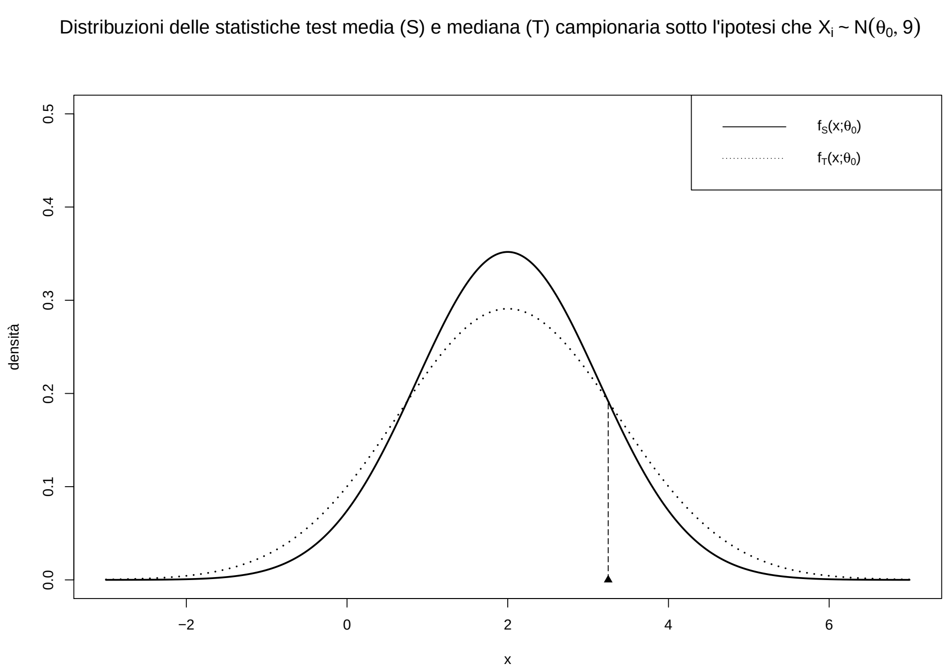

Figure 1: *Distribuzioni delle statistiche test S (media) e T (mediana) sotto l'ipotesi che $X_i \sim N(\theta_0, 9)$*

C1 Fornisci un'argomentazione grafica appropriata per giustificare l'affermazione che, utilizzando come soglia per le due diverse statistiche test lo stesso valore $s = t = 3.25$ (indicato nella Figura 1 da un triangolino), il test $R_S(3.25)$ basato sulla media campionaria assicura un miglior controllo della probabilità dell'errore di prima specie rispetto al test $R_T(3.25)$ basato sulla mediana campionaria.

C2 Si consideri il seguente sistema di ipotesi

$$\begin{cases} H_0: & \theta \leq \theta_0 \\ H_1: & \theta > \theta_0 \end{cases}$$

Per un valore fissato s determina l'espressione della funzione di potenza

$$\eta_{R_S(s)}(\theta) = \mathbb{P}_\theta \left\{ (X_1, ..., X_n) \in R_S(s) \right\}$$

corrispondente alla regione di rifiuto

$$R_S(s) = \{(x_1, ..., x_n) \in \mathcal{X}_n : S(x_1, ..., x_n) > s\}$$

per un generico $\theta \in \mathbb{R}$.

C3 Supponiamo ora di utilizzare due soglie distinte per le due diverse statistiche test ($s = 3.865$ per la media e $t = 4.263$ per la mediana) e di considerare quindi le regioni di rifiuto

$$R_S(3.865) = \{(x_1, ..., x_n) \in \mathcal{X}_n : S(x_1, ..., x_n) > 3.865\}$$

e

$$R_T(4.263) = \{(x_1, ..., x_n) \in \mathcal{X}_n : T(x_1, ..., x_n) > 4.263\}$$

Nella seguente Figura 2 abbiamo rappresentato le due funzioni di potenza corrispondenti alle regioni $R_S(3.865)$ e $R_T(4.263)$. Sapresti argomentare in base al grafico che le due regioni assicurano la stessa

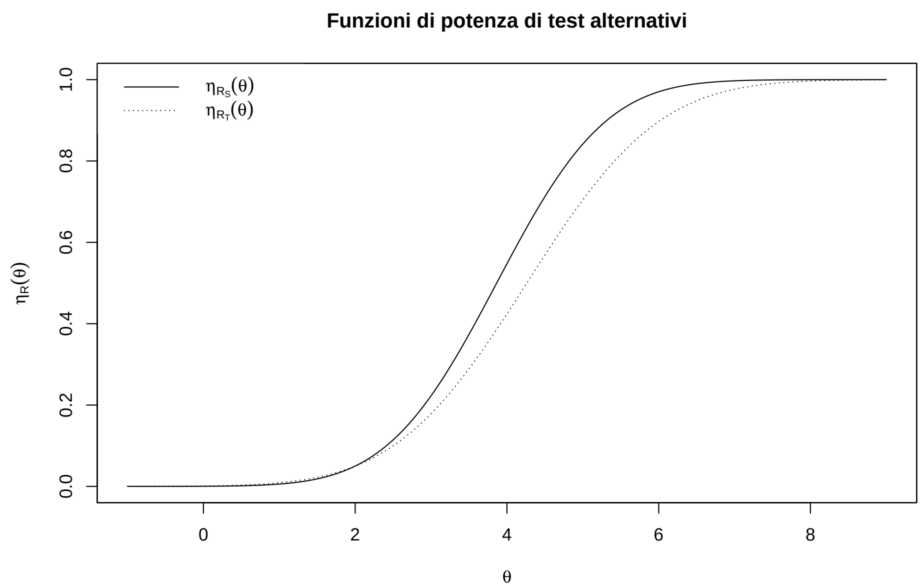

Figure 2: *Funzioni di potenza di test alternativi*

dimensione del test?

C4 Sapresti argomentare in base al grafico che la regione basata sulla media è preferibile a quella basata sulla mediana?

C5 Sapresti giustificare con un'appropriata argomentazione teorica l'ultima affermazione senza usare il grafico?

C6 Supponiamo che sia stato osservato il seguente campione:

$$\boldsymbol{X}_n = \boldsymbol{x}_n^{oss} = (-0.1, 10.33, 3.38, -0.37, -1.86, 20.06, 1.14)$$

Fornire le conclusioni per i test associati alle regioni $R_S(3.865)$ e $R_T(4.263)$

C7 Evidenziare sull'opportuna figura il p-value corrispondente al test $R_S(3.865)$.