Teoria e Laboratorio di Inferenza Statistica
Corso di Laurea in Statistica, Popolazione e Ricerca Sociale
Anno Accademico 2004-2005

Verifica n. 1 - 19 aprile 2005

COMPITO A

COGNOME E NOME ——————————————————————-

Esercizio 1. Si consideri un campione casuale $X_1, \ldots, X_n$, dove X_i indica il tempo che un impiegato di banca dedica a ciascun cliente, e si supponga che X_i abbia distribuzione normale di parametri μ e σ^2, entrambi incogniti. Si consideri quindi un campione osservato di dimensione $n = 16$, per il quale si ha $\sum_{i=1}^n x_i = 49.6$ e $\sum_{i=1}^n (x_i - \bar{x}_n)^2 = 2.56$.

a) Verificare che le stime di massima verosimiglianza dei parametri incogniti sono pari a $\hat{\mu} = 3.10$ minuti e $\hat{\sigma} = 0.40$ minuti.

b) Utilizzando le stime riportate al punto a), determinare la probabilità che il tempo dedicato dall'impiegato a un singolo cliente sia superiore a 3 minuti.

c) Utilizzando le stime riportate al punto a), determinare la probabilità che il tempo complessivo dedicato a 10 clienti sia inferiore a 35 minuti.

Svolgimento.

a) Sappiamo che le stime di MV sono, in questo modello,

$$\hat{\mu} = \bar{x}_n = \frac{49.6}{16} = 3.10, \qquad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{2.56}{16} \quad \Rightarrow \quad \hat{\sigma} = 0.40.$$

b) Indicando con Z la v.a. $\mathrm{N}(0,1)$ e con $\Phi(\cdot)$ la sua f. di ripartizione, si ha:

$$\Pr(X > 3) = \Pr\left(Z > \frac{3 - 3.1}{0.4}\right) = \Pr(Z > -0.25) = 1 - \Pr(Z < -0.25) = 1 - \Phi(-0.25) \simeq 0.6$$

c) $X_1, \ldots, X_n$ i.i.d. $\Rightarrow Y_n = \sum_{i=1}^n X_i \sim \mathrm{N}(n\mu; n\sigma^2) =$ (in questo caso) $= \mathrm{N}(31, 10(0.4)^2)$. Quindi:

$$\Pr(Y_n < 35) = \Pr\left(Z < \frac{35 - 31}{0.4\sqrt{10}}\right) = \Pr(Z < 3.17) = \Phi(3.17) \simeq 0.99$$

Esercizio 2. Si consideri un campione casuale di n osservazioni da una popolazione X con distribuzione di probabilità

$$f_X(x;\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x} \qquad x = \{0,1,2\}, \qquad 0<\theta<1.$$

a) Verificare che
$$E_\theta[X] = 2\theta \qquad V_\theta[X] = 2\theta(1-\theta).$$

b) Determinare $E_\theta[\bar{X}_n]$ e $V_\theta[\bar{X}_n]$.

c) Determinare l'espressione della funzione di verosimiglianza del parametro θ, $L(\theta;\mathbf{x}_n)$, associata a un generico campione osservato, $\mathbf{x}_n$, individuare il nucleo della funzione di verosimiglianza e, se esiste, una statistica sufficiente unidimensionale.

Svolgimento.

a) Si osservi che $X_i|\theta \sim \text{Bin}(2,\theta) \quad \Rightarrow \quad E_\theta[X] = 2\theta, \qquad V_\theta[X] = 2\theta(1-\theta).$

Per il calcolo esplicito si ha:

$$E_\theta[X] = \sum_{x=0}^{2} x f_X(x;\theta) = 0\times\binom{2}{0}\theta^0(1-\theta)^2 + 1\times\binom{2}{1}\theta^1(1-\theta)^1 + 2\times\binom{2}{2}\theta^2(1-\theta)^0 = ... = 2\theta,$$

e che
$$E_\theta[X^2] = \sum_{x=0}^{2} x^2 f_X(x;\theta) = ... = 2\theta^2 + 2\theta,$$

da cui
$$V_\theta[X] = E_\theta[X^2] - (E_\theta[X])^2 = 2\theta^2 + 2\theta - (2\theta)^2 = 2\theta(1-\theta).$$

b) Per le note proprietà di media e varianza campionarie per campioni casuali, si ha che

$$E_\theta[\bar{X}_n] = E_\theta[X] = 2\theta$$

e che
$$V_\theta[\bar{X}_n] = \frac{1}{n}V_\theta[X] = \frac{2}{n}\theta(1-\theta).$$

c) Si ha $L(\theta;\mathbf{x}_n) = \left[\prod_{i=1}^n \binom{2}{x_i}\right]\theta^{\sum_{i=1}^n x_i}(1-\theta)^{2n-\sum_{i=1}^n x_i}$. Pertanto:

$$L(\theta;\mathbf{x}_n) = h(\mathbf{x}_n)g(T(\mathbf{x}_n);\theta), \quad \text{dove} \quad h(\mathbf{x}_n) = \prod_{i=1}^{n}\binom{2}{x_i}, \quad g(T(\mathbf{x}_n);\theta) = \theta^{T(\mathbf{x}_n)}\theta^{2n-T(\mathbf{x}_n)}.$$

In base al criterio di fattorizzazione si ha quindi che $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$ è statistica sufficiente unidimensionale.

Esercizio 3. Con riferimento al precedente esercizio, si consideri un campione osservato di $n = 20$ osservazioni, tale che $x_{(1)} = x_{(2)} = \ldots = x_{(14)} = 0,\ x_{(15)} = x_{(16)} = 1,\ x_{(17)} = \ldots = x_{(20)} = 2$.

a) Determinare la stima di massima verosimiglianza di θ.

b) Determinare il valore dell'informazione osservata, I_n^{oss}, e l'espressione dell'approssimazione normale, $\bar{L}_N(\theta; \mathbf{x}_n)$, per la funzione di verosimiglianza relativa.

c) Determinare l'insieme di verosimiglianza di livello $q = 0.85$, utilizzando l'approssimazione normale ottenuta al punto precedente.

d) Sulla base dei precedenti risultati, come è possibile stimare la quantità $E_\theta[\bar{X}_n]$?

Svolgimento.

a)

$$\frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n) = \frac{d}{d\theta}[\sum_{i=1}^{n} x_i \ln\theta + (2n - \sum_{i=1}^{n} x_i)\ln(1-\theta)] = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{2n - \sum_{i=1}^n x_i}{1-\theta}.$$

Uguagliando a zero e risolvendo rispetto a θ si trova

$$\hat{\theta}_{smv}(\mathbf{x}_n) = \frac{\sum_{i=1}^n x_i}{2n} = 10/40 = 1/4,$$

che è punto di massimo in quanto la derivata seconda di $\ln L(\theta; \mathbf{x}_n)$ è negativa $\forall\theta$.

b) Poichè

$$= -\frac{d^2}{d\theta^2}[\frac{\sum_{i=1}^n x_i}{\theta} - \frac{2n - \sum_{i=1}^n x_i}{1-\theta}] = \frac{\sum_{i=1}^n x_i}{\theta^2} + \frac{2n - \sum_{i=1}^n x_i}{(1-\theta)^2},$$

Si ha

$$I_n^{oss} = -\frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n) \mid_{\theta=\hat{\theta}_{smv}} = \frac{\sum_{i=1}^n x_i}{\hat{\theta}_{smv}^2} + \frac{2n - \sum_{i=1}^n x_i}{(1-\hat{\theta}_{smv})^2} = \frac{10}{1/16} + \frac{40-10}{9/16} = \frac{640}{3}.$$

Quindi

$$\bar{L}_N(\theta; \mathbf{x}_n) = \exp\{-\frac{1}{2}(\theta - 1/4)^2(640/3)\}.$$

c) Poichè $\sqrt{-2\ln(0.85)} = 0.57$, si ha

$$\tilde{L}_q(\mathbf{x}_n) = [1/4 - 0.57\sqrt{3/640}, 1/4 + 0.57\sqrt{3/640}] = [0.211, 0.289].$$

d) $\widehat{E_\theta[X]} = \widehat{[2\theta]} = 2\hat{\theta}_{smv}(\mathbf{x}_n) = 1/2.$