Teoria e Laboratorio di Inferenza Statistica
Corso di Laurea in Statistica, Popolazione e Ricerca Sociale

Statistica 2
Corso di Laurea in Statistica, Finanza e Assicurazioni

Anno Accademico 2005-2006

Prima Prova Scritta - 12 giugno 2006

COMPITO A

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

Esercizio 1A. Sia $X_1, \ldots X_n$ un campione casuale proveniente da una popolazione con distribuzione uniforme nell'intervallo $(\theta, 2\theta)$, con parametro incognito θ e funzione di densità

$$f_X(x;\theta) = \frac{1}{\theta} I_{(\theta,2\theta)}(x), \qquad \theta > 0.$$

a) Verificare che
$$E_\theta[X] = \frac{3}{2}\theta, \qquad V_\theta[X] = \frac{\theta^2}{12}.$$

b) Verificare (giustificando tutte le affermazioni) che l'approssimazione normale della distribuzione campionaria della v.a. media campionaria, $\bar{X}_n$, risulta essere $N(\frac{3}{2}\theta, \frac{\theta^2}{12n})$.

c) Sulla base del precedente punto, determinare la probabilità che la v.a. $\bar{X}_n$ assuma valori superiori a $\frac{19}{6}$, ponendo $\theta = 2$ e $n = 48$.

d) Verificare che, per il modello considerato, una statistica sufficiente è rappresentata dal vettore $\big(x_{(1)}, x_{(n)}\big)$, ovvero dalla coppia costituita dal minimo e massimo dei valori campionari.

Svolgimento.

a)

$E_\theta[X] = \int_\theta^{2\theta} \frac{x}{\theta} d\theta = \ldots = \frac{3}{2}\theta.$ $\qquad E_\theta[X^2] = \int_\theta^{2\theta} \frac{x^2}{\theta} d\theta = \ldots = \frac{7}{3}\theta^2.$
$V_\theta[X] = E_\theta[X^2] - (E_\theta[X])^2 = \ldots = \frac{1}{12}\theta^2.$

b)

Segue dal teorema del limite centrale (le cui ipotesi sono soddisfatte) osservando che:

$$E_\theta[\bar{X}_n] = E_\theta[X] = \frac{3}{2}\theta, \qquad V_\theta[\bar{X}_n] = \frac{V_\theta[X]}{n} = \frac{1}{12n}\theta^2.$$

c)

$$\Pr\left(\bar{X}_n > \frac{19}{6}\right) \simeq \Pr\left(Z > \frac{19/6 - 3}{1/12}\right) = \Pr(Z > 2) = 1 - \Phi(2) = 0.023.$$

d)

Poichè

$$\theta < x_1 < \ldots < x_n < 2\theta \qquad \Leftrightarrow \theta < x_{(1)} < \ldots < x_{(n)} < 2\theta,$$

si ha che $\prod_{i=1}^n I_{(\theta,2\theta)}(x_i) = I_{(\theta,2\theta)}(x_{(1)})\, I_{(\theta,2\theta)}(x_{(n)})$. Pertanto

$$L(\theta; \mathbf{x}_n) = \frac{1}{\theta^n} \prod_{i=1}^n I_{(\theta,2\theta)}(x_i) = \frac{1}{\theta^n} I_{(\theta,2\theta)}(x_{(1)})\, I_{(\theta,2\theta)}(x_{(n)}),$$

e risultato discende dal criterio di fattorizzazione.

Esercizio 2A. Sia $X_1, \dots, X_n$ un campione casuale proveniente da una popolazione geometrica [1] di parametro incognito θ, la cui funzione di massa di probabilità è:

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

a) Determinare la funzione di verosimiglianza di θ associata a un generico campione osservato e una statistica sufficiente per il modello.

b) Determinare la stima di massima verosimiglianza di θ.

c) Verificare che la famiglia delle distribuzioni geometriche costituisce una famiglia esponenziale.

d) In un campione di $n = 5$ osservazioni, si è rilevato che $x_1 = 3,\ x_2 = 5,\ x_3 = 1,\ x_4 = 2,\ x_5 = 4$. Utilizzando la stima di massima verosimiglianza di θ, determinare la probabilità che la v.a. X assuma valori maggiori di 3.

Svolgimento.
a)
La funzione di verosimiglianza è:

$$L(\theta; \mathbf{x}_n) = \theta^n (1-\theta)^{\sum_{i=1}^n x_i - n}.$$

Per il criterio di fattorizzazione segue che $\sum_{i=1}^n x_i$ è una stat. sufficiente per il modello.

b)
Risolvendo l'equazione di log-verosimiglianza si trova che la soluzione

$$\hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^n x_i},$$

che risulta essere effettivamente un punto di massimo per la funzione.

c)
Si ha che:

$$f_X(x;\theta) = \exp(\log[\theta(1-\theta)^{x-1}]) = (\text{propr. dei logaritmi}) = \exp(x\log(1-\theta) + \log\frac{\theta}{1-\theta}).$$

Il modello considerato è quindi famiglia esponenziale con

$$h(x) = 1, \qquad T(x) = x, \qquad \eta(\theta) = \log(1-\theta), \qquad c(\theta) = \log\frac{\theta}{1-\theta}.$$

d)
Per un θ generico si ha:

$$\Pr(X > 3; \theta) = 1 - \Pr(X \leq 3; \theta) = 1 - \sum_{x=1}^{3} f_X(x;\theta) = 1 - [\theta + \theta(1-\theta) + \theta(1-\theta)^2]|.$$

Il risultato si ottiene sostituendo nella precedente espressione a θ il valore della stima di massima verosimiglianza che, con $n = 5$ e $\sum_{i=1}^n x_i = 15$, risulta essere in quest'esempio $\hat{\theta}_{MV} = 5/15 = 1/3$.

[1] La v.a. aleatoria geometrica viene utilizzata per descrivere il numero (aleatorio) minimo di prove, ciascuna con possibile esito di tipo *successo* o *insuccesso*, da effettuare per osservare un successo. Si noti infatti che la v.a. geometrica assume valori nei numeri naturali e che il valore minimo che può assumere è pari a 1.

Esercizio 3A. In un studio di è interessati a determinare il tasso di mutazioni geniche in una specifica popolazione batterica sottoposta a un fattore di rischio. Si considerano n colonie del batterio e si indica con X_i il numero di batteri mutanti nell'i-esima colonia. Si ipotizza inoltre che le n v.a. X_i costituiscano un campione casuale da una popolazione di Poisson di parametro incognito θ. Siamo interessati a effettuare inferenza sul parametro incognito θ (numero medio di mutazioni in una singola colonia) e sulla sua funzione $\lambda = g(\theta) = \Pr\{X = 1; \theta\}$, (probabilità che in una singola colonia vi sia una sola mutazione).

a) Determinare gli stimatori di massima verosimiglianza di θ e di λ.

b) Determinare la varianza di $\hat{\theta}$ e verificare che la varianza asintotica di $\hat{\lambda}$ risulta pari a $\theta(1-\theta)^2 e^{-2\theta}/n$.

c) Determinare le stime per le varianze degli stimatori ottenute al punto precedente.

d) Determinare le approssimazioni asintotiche delle distribuzioni campionarie degli stimatori $\hat{\theta}$ e $\hat{\lambda}$ e l'intervallo di confidenza approssimato di livello 0.95 per λ.

e) Supponendo che, per un campione di dimensione $n = 30$, il valore osservato della media campionaria è pari a 3.7, determinare i valori delle stime puntuali di θ e λ e l'intervallo di confidenza per λ.

Svolgimento.

Esercizio 4A. Si assume che il ricavo delle vendite (in migliaia di euro) di prodotto in un periodo di riferimento sia una v.a. normale di parametri (θ, σ^2), entrambi incogniti. Si considera un campione casuali di ricavi relativi a $n = 9$ punti vendita. I valori osservati dei ricavi sono riportati nella seguente tabella.

P.to Vendita	Ricavi (x_i)
1	1.72
2	0.50
3	1.01
4	1.14
5	1.13
6	1.55
7	2.32
8	0.71
9	0.94

a) Determinare le stime di massima verosimiglinza dei due parametri e l'intervallo di confidenza al 95% per il parametro incognito σ^2.

b) Si sottoponga a verifica il seguente sistema di ipotesi, assumendo un livello di significatività pari a $\alpha = 0.05$:

$$H_0 : \ \theta < 1.1$$
$$H_1 : \ \theta \geq 1.1$$

e commentare l'esito del test.

c) Calcolare il valore p e rappresentarlo graficamente, assieme alla regione di rifiuto del test del punto precedente.

Svolgimento.