Metodi Matematici e Statistici

Prova scritta – 17/7/2007

Esercizio 1. Siano T_0 e T_1 due v.a. esponenziali di parametri 5 e 6, W una Bernoulli di parametro 1/2, tutte indipendenti.

i) Calcolare la generatrice di $T_0 + T_1$.

ii) Calcolare $E[T_1^2 T_0]$.

iii) Calcolare $P(T_W > 1)$.

Esercizio 2. Sia X una v.a. gaussiana di media θ incognita e deviazione standard pari a 4.

i) Vogliamo stimare θ con un errore assoluto pari a 1. Se vogliamo il risultato al 95%, quanto numeroso dev'essere il campione?

ii) Supponiamo che da quel campione sia emersa la stima $\widehat{\theta} = 22$. Cosa possiamo dire della media vera θ?

iii) Cosa possiamo dire della probabilità che X superi 24?.

Esercizio 3. Indichiamo con X una v.a. di Bernoulli di parametro p e poniamo $X' = 2X - 1$.

i) Trovare p in modo che X' abbia media nulla. Chiamare p^* tale valore.

ii) Siamo $X_1', ..., X_n'$ delle v.a. indipendenti distribuite come X', con p generico (non necessariamente uguale a p^*). Dare tutte le formule che si conoscono, esatte o approssimate, che si possono usare per il calcolo di

$$P(X_1' + ... + X_n' > n\lambda).$$

iii) Eseguire il calcolo del punto (ii) nel caso in cui le X_k' siano $N(\mu, \sigma^2)$ indipendenti.

Soluzioni

Esercizio 1. i)

$$E[e^{t(T_0+T_1)}] = E[e^{tT_0}]E[e^{tT_1}] = \frac{5}{5-t}\frac{6}{6-t}.$$

ii)

$$\begin{aligned} E[T_1^2 T_0] &= E[T_1^2]E[T_0] = \left(Var[T_1] + E[T_1]^2\right)\frac{1}{5} \\ &= \left(\frac{1}{6^2} + \frac{1}{6^2}\right)\frac{1}{5} = \frac{1}{90} = 1.1111 \times 10^{-2}. \end{aligned}$$

iii)

$$\begin{aligned} P(T_W > 1) &= P(T_W > 1|W = 0)P(W = 0) + \\ &\quad P(T_W > 1|W = 1)P(W = 1) \\ &= \frac{1}{2}(P(T_0 > 1) + P(T_1 > 1)) \\ &= \frac{1}{2}(e^{-5} + e^{-6}) = 4.6083 \times 10^{-3}. \end{aligned}$$

Esercizio 2. Sia X una v.a. gaussiana di media θ incognita e deviazione standard pari a 4.

i)

$$\frac{4q_{0.975}}{\sqrt{n}} \leq 1$$

da cui

$$n \geq (4q_{0.975})^2 = 61.466$$

ovvero $n = 62$.

ii)

$$\theta = 22 \pm 1 \text{ a livello di confidenza } 95\%.$$

In realtà, ricalcolando l'errore relativo a 62 prove, si trova un intervallo anche un po' più piccolo, quindi migliore (questo è dovuto all'approssimazione $n = 62$ rispetto a 61.466). Entrambi comunque sono intervalli di confidenza al 95%.

iii) A livello di confidenza 95%, $21 \leq \theta \leq 23$, quindi

$$P^{21}(X > 24) \leq P^{\theta}(X > 24) \leq P^{23}(X > 24)$$

$$1 - \Phi\left(\frac{24-21}{4}\right) \leq P^{\theta}(X > 24) \leq 1 - \Phi\left(\frac{24-23}{4}\right)$$

$$1 - \Phi\left(\frac{3}{4}\right) \leq P^{\theta}(X > 24) \leq 1 - \Phi\left(\frac{1}{4}\right)$$

$$1 - 0.7734 \leq P^{\theta}(X > 24) \leq 1 - 0.5987$$

$$0.2266 \leq P^{\theta}(X > 24) \leq 0.4013.$$

Esercizio 3. i)

$$E[X'] = -1 \bullet (1-p) + 1 \bullet p = -1 + 2p$$

quindi è zero se e solo se $p = \frac{1}{2}$. Quindi $p^* = \frac{1}{2}$.

ii) Siano $X_1', \ldots, X_n'$ delle v.a. indipendenti distribuite come X', con p generico (non necessariamente uguale a p^*). Dare tutte le formule che si conoscono, esatte o approssimate, che si possono usare per il calcolo di

Osserviamo che $X' = 2X - 1$ con $X \sim B(1,p)$, quindi

$$X_1' + \ldots + X_n' = 2(X_1 + \ldots + X_n) - 1$$

con $X_1, \ldots, X_n \sim B(1,p)$ indipendenti, e $X_1 + \ldots + X_n \sim B(n,p)$. Pertanto

$$\begin{aligned} P(X_1' + \ldots + X_n' > n\lambda) &= P(2(X_1 + \ldots + X_n) - 1 > n\lambda) \\ &= P\left(X_1 + \ldots + X_n > \frac{n\lambda + 1}{2}\right) \\ &= \sum_{k > \frac{n\lambda+1}{2}} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Per il teorema degli eventi rari possiamo anche dire

$$P(X_1' + ... + X_n' > n\lambda) \sim 1 - \sum_{k \leq \frac{n\lambda+1}{2}} e^{-np} \frac{(np)^k}{k!}.$$

Infine, per il teorema limite centrale, osservando che

$$Var[X'] = 4Var[X] = 4p(1-p),$$

$$\begin{aligned} P(X_1' + ... + X_n' > n\lambda) &\sim P\left(Z > \frac{n\lambda - n(2p-1)}{2\sqrt{n}\sqrt{p(1-p)}} \right) \\ &= 1 - \Phi\left(\sqrt{n} \frac{\lambda + 1 - 2p}{2\sqrt{p(1-p)}} \right). \end{aligned}$$

dove $Z \sim N(0,1)$.

iii) In questo caso $X_1' + ... + X_n' \sim N(n\mu, n\sigma^2)$ (le combinazioni lineari di gaussiane indipendenti sono gaussiane) quindi

$$\begin{aligned} P(X_1' + ... + X_n' > n\lambda) &= 1 - \Phi\left(\frac{n\lambda - n\mu}{\sqrt{n}\,\sigma} \right) \\ &= 1 - \Phi\left(\sqrt{n} \frac{\lambda - \mu}{\sigma} \right) \end{aligned}$$

che è la stessa formula trovata sopra col TLC, essendo $2p-1$ la media di X' e $2\sqrt{p(1-p)}$ la deviazione standard.