Corso di Laurea in Ingegneria Informatica
Anno Accademico 2016/2017
Calcolo delle Probabilità e Statistica Matematica

Nome ....

N. Matricola .... Ancona, 11 gennaio 2017

1. (8 punti) Si vuole stimare il parametro p di una legge di Bernoulli. Si considerino due stimatori

$$\widehat{p}_1 = \frac{X_1 + X_2 + ... + X_n}{n}$$
$$\widehat{p}_2 = \frac{X_1 + 2\,X_2 + X_3 + 2\,X_4 + ... + 2\,X_{n-3} + X_{n-2} + 2\,X_{n-1} + X_n}{n + (n-1)/2}$$

dove il campione ha rango n dispari e maggiore di 1. Si chiede di:

- (2 punti) determinare se sono stimatori corretti;
- (3 punti) determinare quale dei due è più efficace;
- (3 punti) verificare che entrambi soddisfano la disuguagnianza di Cramer-Rao.

Soluzione.

$$E[\widehat{p}_1] = \frac{n\,p}{n}$$
$$E[\widehat{p}_2] = \frac{(n + (n-1)/2)\,p}{n + (n-1)/2}$$

Sono entrambi stimatori corretti. Sia ora $\sigma^2 = p\,(1-p)$ la varianza di una delle variabili del campione.

$$Var(\widehat{p}_1) = \frac{1}{n^2}\,n\,Var(X_1) = \frac{1}{n}\,\sigma^2$$
$$Var(\widehat{p}_2) = \frac{1}{(n + (n-1)/2)^2}\left[\frac{n+1}{2}\,\sigma^2 + \frac{n-1}{2}\,4\,\sigma^2\right] = \frac{p\,(1-p)}{n + (n-1)/2} =$$
$$= 2\,\sigma^2\,\frac{5\,n-3}{(3\,n-1)^2}$$

Si vede facilmente che

$$\frac{1}{n} < 2\,\frac{5\,n-3}{(3\,n-1)^2}$$

per $n > 1$ e quindi è più efficace $\widehat{p}_1$.

Il limite inferiore della diguguaglianza di Cramer-Rao per la legge di Bernoulli è

$$C(n,p) = \frac{p\,(1-p)}{n}$$

soddisfatto con il segno di uguaglianza da $\widehat{p}_1$ e quindi da $\widehat{p}_2$ con la disuguaglianza stretta.

2. (6 punti) Dante e Beatrice, non avendo altro da fare, decidono di giocare "testa o croce" con una moneta truccata. La probabilità di ottenere testa è 3/4. Quando esce testa, Dante dà a Beatrice 5 fiorini, se esce croce Beatrice ne dà 10 a Dante. Quant'è il guadagno medio di Beatrice? E quello di Dante?

Soluzione. Indichiamo con Y la v.a. che rappresenta il guadagno di Dante in una singola giocata, e con Z quello di Beatrice. Abbiamo:

$$E[Y] = (-5)\,\frac{3}{4} + (10)\,\frac{1}{4} = -\frac{5}{4}$$
$$E[Z] = (5)\,\frac{3}{4} + (-10)\,\frac{1}{4} = \frac{5}{4}$$

Dante perde in media 1.25 fiorini, Beatrice ne guadagna altrettanti.

3. (6 punti) Un'organizzazione di volontariato vuole stimare l'era media dei suoi aderenti. Un campione di rango $n = 20$ fornisce $\overline{X}_n = 31$ per la media campionaria e $S^2 = 25$ per la varianza campionaria. Fornire l'intervallo di confidenza al 90%, 95% e 99% per l'età media.

Soluzione. I valori dei quantili di Student a 19 gradi di libertà per $\alpha/2$ sono

$$t_{0.05} = 1.729 \quad \text{al } 90\%$$
$$t_{0.025} = 2.093 \quad \text{al } 95\%$$
$$t_{0.005} = 2.861 \quad \text{al } 99\%$$

Gli intervalli di confidenza sono dati da

$$\left[\overline{X}_n - \frac{S}{\sqrt{n}}\,t_{\alpha/2}, \overline{X}_n + \frac{S}{\sqrt{n}}\,t_{\alpha/2}\right]$$

che diventano

$$[29.07, 32.93] \quad \text{al } 90\%$$
$$[28.66, 33, 34] \quad \text{al } 95\%$$
$$[27.80, 34.20] \quad \text{al } 99\%$$

4. (6 punti) Un ricco turista decide di fare una vacanza senza badare a spese. All'inizio della vacanza possiede un somma di 100.000 (centomila) euro; la quantità di denaro che spende in un giorno può essere rappresentata da una variabile normale di media 5.000 euro e deviazione standard $\sigma = 2000$ euro. Qualè la probabilità che debba tornare a casa dopo 18 giorni?

Soluzione. Indichiamo con X_1, X_2 ... X_n le variabili che rappresentano la spesa sostenuta nel primo, nel secondo, ..., nell'$n-$esimo giorno (con $n = 18$). La probabilità richiesta è allora

$$
\begin{aligned}
&P\left(X_1 + X_2 + ... + X_n \geq 100000\right) = \\
&\quad = P\left(\frac{X_1 + X_2 + ... + X_n - 5000\,n}{\sigma\sqrt{n}} \geq \frac{100000 - 18 \times 5000}{2000\sqrt{18}}\right) \approx \\
&\quad \approx P\left(S_n \geq 1.18\right) = 1 - 0.8810 = 0.119
\end{aligned}
$$

5. (28 punti) All'ambulatorio di un medico arrivano pazienti alla media di 10 all'ora. Il tempo che ciascun paziente trascorre nello studio del medico per la visita può essere rappresentato da una variabile uniforme Y definita al modo seguente:

$$\begin{array}{ll} Y \in [0, 15] & \text{se nella prima mezz'ora sono giunte non più di 3 persone} \\ Y \in [0, 10] & \text{se nella prima mezz'ora sono giunte più di 3 persone} \end{array}$$

(significa che se il medico vede poca gente dedica più tempo a ciascun paziente; se vede tanta gente si sbriga prima su ciascuno di essi). Indicata con X la v.a. che rappresenta gli arrivi nella prima mezz'ora, si chiede di:

- (4 punti) determinare la densità congiunta di X e Y;
- (4 punti) determinare le densità marginali;
- (2 punti) verificare l'indipendenza o meno delle variabili;
- (9 punti) nel caso non siano indipendenti, calcolare il coefficiente di correlazione;
- (9 punti) calcolare la probabilità che un paziente che arriva nella seconda mezz'ora trascorra meno di 5 minuti nello studio del medico.

Soluzione. La variabile X segue una legge di Poisson di parametro $\lambda = 5$, che indichiamo con

$$P(X = k) = p(k) = e^{-5}\,\frac{5^k}{k!},$$

e può assumere tutti i numeri naturali. La variabile Y segue una legge uniforme in un intervallo che dipende dal valore assunto dalla X. La densità congiunta é quindi una funzione che dipende da una variabile discret, X, e una variabile continua, Y. Possiamo quindi indicare la densità congiunta simbolicamente con una successione di funzioni $\{f_k(y)\}_{k=0}^{\infty}$. Ricaviamo l'espressione esplicita delle $f_k(y)$ con il ragionamento seguente. Il dato del problema, cioè che Y è uniforme su $[0, 15]$ se $X \leq 3$ e su $[0, 10]$ se $X > 3$ si traduce in

$$\begin{aligned} P(Y \in I|X = k) &= \frac{|I|}{15},\ \forall\, I \subset [0, 15],\ k \leq 3 \\ P(Y \in J|X = k) &= \frac{|J|}{10},\ \forall\, J \subset [0, 10],\ k > 3 \end{aligned}$$

dove $|I|$ e $|J|$ sono le misure degli intervalli I e J. Ricordando la definizione di probabilità condizionata abbiamo:

$$\begin{aligned} P(\{Y \in I\} \cap \{X = k\}) &= \frac{|I|}{15}\,p(k),\ \forall\, I \subset [0, 15],\ k \leq 3 \\ P(\{Y \in J\} \cap \{X = k\}) &= \frac{|J|}{10}\,p(k),\ \forall\, J \subset [0, 10],\ k > 3. \end{aligned}$$

Le funzioni $f_k(y)$ sono pertanto date da

$$\begin{aligned} f_k(y) &= \frac{1}{15}\,\chi([0, 15])\,p(k),\ k \leq 3 \\ f_k(y) &= \frac{1}{10}\,\chi([0, 10])\,p(k),\ k > 3, \end{aligned}$$

dove $\chi([a, b])$ è la funzione caratteristica dell'intervallo $[a, b]$.

Veniamo alle densità marginali. La densità marginale della X è data da

$$f_X(k) = \int_0^{15} \frac{1}{15}\, p(k)\, dy = p(k),\;\; k \leq 3$$
$$f_X(k) = \int_0^{10} \frac{1}{10}\, p(k)\, dy = p(k),\;\; k > 3.$$

La marginale f_X è pertanto una variabile di Poisson con parametro $\lambda = 5$, com'era da aspettarsi. Per la Y abbiamo:

$$f_Y(y) = \sum_{k=0}^{3} p(k)\, \frac{1}{15} + \sum_{k=4}^{\infty} p(k)\, \frac{1}{10}\;\; y \in [0, 10]$$
$$f_Y(y) = \sum_{k=0}^{3} p(k)\, \frac{1}{15}\;\; y \in [10, 15]$$

Nel prosieguo, indichiamo con

$$\alpha = \sum_{k=0}^{3} p(k) \approx 0.265$$
$$\beta = \sum_{k=0}^{3} k\, p(k) \approx 0.623.$$

La marginale di Y diventa pertanto

$$f_Y(y) = \frac{\alpha}{15} + \frac{1-\alpha}{10} = 0.0912, \quad y \in [0, 10]$$
$$f_Y(y) = \frac{\alpha}{15} = 0.0177, \quad y \in [10, 15].$$

Si vede subito che le due variabili non sono indipendenti.

Per il calcolo del coefficiente di correlazione abbiamo:

$$\begin{aligned}
&E[X] = Var(X) = \lambda = 5;\\
&E(Y) = 0.0912 \int_0^{10} y\, dy + 0.0177 \int_{10}^{15} \frac{1}{10}\, y\, dy = 5.663;\\
&E(Y^2) = 0.0912 \int_0^{10} y^2\, dy + 0.0177 \int_{10}^{15} \frac{1}{10}\, y^2\, dy = 44.376;\\
&Var(Y) = 44.375 - 32.064 = 12.311;\\
&E[X\,Y] = \sum_{k=0}^{\infty} k\, p(k) \int_0^{15} f_k(y)\, y\, dy =\\
&\quad = \sum_{k=0}^{3} k\, p(k) \int_0^{15} \frac{1}{15}\, y\, dy + \sum_{k=4}^{\infty} k\, p(k) \int_0^{10} \frac{1}{10}\, y\, dy =\\
&\beta\, \frac{15^2}{30} + (5-\beta)\, \frac{10^2}{20} = 26.558,
\end{aligned}$$

dove abbiamo sfruttato la proprietà della distribuzione di Poisson che

$$\sum_{k=0}^{\infty} k\, p(k) = \lambda = 5.$$

Abbiamo pertanto per la covarianza

$$Cov(X,Y) = E[X\,Y] - E[X]\,E[Y] = 26.558 - 5 \times 5.662 = -1.755$$

e per il coefficiente di correlazione

$$\rho = \frac{-1.755}{\sqrt{5 \times 12.311}} = -0.224$$

Infine, la probabilità che un paziente che arriva nella seconda mezz'ora trascorra meno di 5 minuti nello studio del medico, usando il teorema delle probabilità totali, è data da

$$\begin{aligned} & P(Y < 5 | k \leq 3)\,P(k \leq 3) + P(Y < 5 | k > 3)\,P(k > 3) = \\ & = \alpha \int_0^5 \frac{1}{15}\,dy + (1-\alpha) \int_0^5 \frac{1}{10}\,dy = \frac{\alpha}{3} + \frac{1-\alpha}{2} = 0.456 \end{aligned}$$