Probabilità e Statistica (LT in Matematica)
Prof. P.Dai Pra, prova scritta 07/09/2004.

Cognome: Nome:

Matricola: Firma:

ESERCIZIO 1. Sia (X, Y) un vettore aleatorio con densità congiunta

$$f_{X,Y}(x, y) = e^{-x} 1_{[0,x]}(y) 1_{[0,+\infty)}(y).$$

a. Determinare la densità marginale e la funzione generatrice dei momenti di X.
b. Determinare le densità congiunte e marginali delle variabili $X - Y$ e Y.
c. Calcolare $P(X - Y > Y)$.

Soluzione. a.

$$f_X(x) = \int f_{X,Y}(x, y)dy = xe^{-x} 1_{[0,+\infty)}(x).$$

Perciò, con una facile integrazione per parti

$$\gamma_X(t) = \int_0^{+\infty} xe^{(t-1)x}dx = \frac{1}{(1-t)^2}.$$

b. Posto $Z = X - Y$, si osservi che

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A^{-1} \begin{pmatrix} Z \\ Y \end{pmatrix},$$

con

$$A^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Perciò

$$f_{Z,Y}(z, y) = e^{-z-y} 1_{[0,+\infty)}(z + y) 1_{[0,z+y)}(y) = e^{-z-y} 1_{[0,+\infty)}(z) 1_{[0,+\infty)}(y).$$

Risulta perciò che $Z, Y \sim Exp(1)$ e che sono indipendenti.

c. Per simmetria $P(Z > Y) = P(Y > Z)$. Inoltre, essendo (Z, Y) assolutamente continua, $P(Z = Y) = 0$, sicché $P(Z > Y) + P(Y > Z) = 1$. Ne segue che $P(Z > Y) = 1/2$.

ESERCIZIO 2 Sia X una variabile scalare non negativa, e γ_X la sua funzione generatrice. Sia, inoltre, $\varphi_X(t) = \log \gamma_X(t)$.

a. Mostrare che per ogni $t, c > 0$ vale la disuguaglianza

$$P(X \geq c) \leq e^{-(tc-\varphi(t))}.$$

(Sugg: $X \geq c \iff e^{tX} \geq e^{tc}$, e usare una nota disuguaglianza).

b. Definiamo $\varphi^*(c) = \sup_{t>0}[tc - \varphi(t)]$. Dedurre da a. che

$$P(X \geq c) \leq e^{-\varphi^*(c)}. \tag{1}$$

Sia ora $X \sim Po(1)$.

c. Calcolare $\varphi^*(c)$ per $c \geq 1$.

d. L'usuale disuguaglianza di Chebischev implica

$$P(X \geq c) \leq \frac{E(X)}{c} = \frac{1}{c}. \tag{2}$$

Quale delle disuguaglianze (1) e (2) risulta più forte per c grande?

Soluzione. a. Per la disuguaglianza di Chebischev

$$P(X \geq c) = P\left(e^{tX} \geq e^{tc}\right) \leq \frac{E\left(e^{tX}\right)}{e^{tc}} = e^{-(tc-\varphi(t))}.$$

b. Essenso la disuguaglianza in a. valida per tutti i $t > 0$,

$$P(X \geq c) \leq \inf_{t>0} e^{-(tc-\varphi(t))} = e^{-\sup_{t>0}[tc-\varphi(t)]} = e^{-\varphi^*(c)}.$$

c. Sappiamo che $\gamma_X(t) = \exp\left[e^t - 1\right]$. Dobbiamo perciò calcolare

$$\varphi^*(c) = sup_{t>0}\left[tc - e^t + 1\right].$$

La funzione $t \mapsto tc - e^t + 1$ assume massimo assoluto per $t = \log(c) > 0$, dove vale

$$\varphi^*(c) = c\log(c) - c + 1.$$

d. La (2) è di gran lunga la più forte, dato che, per c grande

$$e^{-c\log(c)+c-1} \ll \frac{1}{c}.$$

ESERCIZIO 3. Un gruppo di biologi ha individuato una mutazione genetica presente nel 2% degli individui di una popolazione. Per una ricerca statistica essi hanno la necessità di trovare almeno 40 individui “mutanti”. Non vi è, tuttavia, alcun modo per predire la presenza o meno della mutazione senza l’analisi del DNA. Usando opportuni metodi di approssimazione si risponda ai seguenti quesiti.

a. Quanti individui dovranno essere sottoposti all’analisi del DNA affinché, con probabilità maggiore o uguale di 0.75, almeno 40 di essi siano portatori della mutazione?

b. Qual è la probabilità che tra i primi cento individui esaminati i mutanti siano non più di tre?

Soluzione. a. Sia

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-mo individio esaminato è mutante} \\ 0 & \text{altrimenti.} \end{cases}$$

Notare che $X_i \sim Be(0.02)$. Si tratta di determinare n in modo tale che

$$P\left(\overline{X}_n > \frac{40}{n}\right) \geq 0.75.$$

Usando l’approssimazione normale e la correzione di continuità:

$$P\left(\overline{X}_n > \frac{40.5}{n}\right) = P\left(\frac{\overline{X}_n - 0.02}{\sqrt{0.02 \times 0.98}}\sqrt{n} \geq -\frac{0.02 - \frac{40.5}{n}}{\sqrt{0.02 \times 0.98}}\sqrt{n}\right)$$

$$\simeq \Phi\left(\frac{0.02 - \frac{40.5}{n}}{\sqrt{0.02 \times 0.98}}\sqrt{n}\right).$$

Quest’ultima quantità è maggiore o uguale di 0.75 se e solo se

$$\frac{0.02 - \frac{40.5}{n}}{\sqrt{0.02 \times 0.98}}\sqrt{n} \geq \Phi^{-}1(0.75) = 0.773$$

cioè

$$0.02n - 0.108\sqrt{n} - 40.5 \geq 0.$$

Risolvendo la precedente, come equazione di secondo grado in $\sqrt{n}$, si trova $n \geq 2284$.

b. Sia $Y = X_1 + X_2 + \cdots + X_{100} \sim B(100, 0.02) \approx Po(20)$. Pertanto

$$P(Y \leq 3) \simeq e^{-2}\left[1 + 2 + \frac{4}{2} + \frac{8}{6}\right] \simeq 0.857.$$

ESERCIZIO 4. Dimostrare le seguenti affermazioni, relative a due eventi A e B.

a. Se $P(B|A) = P(B|A^c)$, allora A e B sono indipendenti.

b. Se $a = P(A)$ e $b = P(B)$, allora $P(A|B) \geq \frac{a+b-1}{b}$.

Soluzione. a. Se $P(B|A) = P(B|A^c)$ allora

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = P(B|A)[P(A) + P(A^c)] = P(B|A),$$

da cui segue l'indipendenza.

b.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{b}.$$

Ma

$$P(A \cap B) = 1 - P(A^c \cup B^c) \geq 1 - P(A^c) - P(B^c) = a + b - 1.$$

ESERCIZIO 5. Un'urna A contiene tre palline verdi e due rosse, un'altra urna B contiene due palline verdi e tre rosse. Si estrae una pallina dall'urna A e la si inserisce nell'urna B. Da B si estrae poi una pallina, che risulta essere rossa. Qual è la probabilità che la pallina trasferita da A a B fosse rossa?

Soluzione. Si considerino gli eventi $E =$ "la pallina trasferita da A a B è rossa" e $F =$ " la pallina estratta da B è rossa. Allora

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)} = \frac{\frac{2}{3}\frac{2}{5}}{\frac{2}{3}\frac{2}{5} + \frac{1}{2}\frac{3}{5}} = \frac{8}{17}.$$