

**Corso di Laurea in Informatica Applicata**  
**Esame di Calcolo delle Probabilità e Statistica**  
**Secondo test di verifica a. a. 2006/2007**

Risolvere esattamente due tra gli esercizi seguenti. Le risposte non motivate non saranno prese in considerazione.

**Primo esercizio.** Sia  $X$  una v. a. avente legge uniforme sull'intervallo  $(0, 1)$ , e sia  $Y$  una v. a. di densità  $B(1, p)$ , indipendente da  $X$ .

- (a) Determinare la funzione di ripartizione della v. a.  $Z = X + Y$ .
- (b) Calcolare  $\mathbf{Cov}(Z, X)$ .
- (c) Dire se esistono valori di  $p$  per i quali  $Z$  ha legge uniforme su un opportuno intervallo  $[a, b]$ ; in caso affermativo, calcolare  $p$ ,  $a$  e  $b$ .

**Secondo esercizio.** Siano  $X$  e  $Y$  due v. a. indipendenti, aventi rispettivamente densità  $f$  e  $g$ , dove

$$f(x) = \begin{cases} 2xe^{-x^2} & \text{per } x > 0 \\ 0 & \text{altrimenti,} \end{cases} \quad g(x) = \begin{cases} 4xe^{-2x^2} & \text{per } x > 0 \\ 0 & \text{altrimenti.} \end{cases}$$

- (a) Determinare la legge della v. a.  $Z = \min\{X, Y\}$ . Si tratta di una legge nota?
- (b) Determinare  $P(Z > t | X > 1)$  per ogni numero positivo  $t$ .
- (c) Determinare  $P(X > 1 | Z > t)$  per ogni numero positivo  $t$ .

**Terzo esercizio.** Un'industria dolciaria vuole stimare il potere calorico del burro utilizzato nella fabbricazione dei suoi panettoni; si suppone che tale potere calorico sia una variabile aleatoria normale di media  $\mu$  e varianza  $\sigma^2$  non note. L'industria acquista il burro da due diversi produttori caseari ed ha disposizione due campioni di osservazioni  $(X_1, X_2, \dots, X_m)$  e  $(Y_1, Y_2, \dots, Y_n)$ , con  $m < n$ , relativi rispettivamente al primo e al secondo produttore. Indicate con  $\bar{X}$  e  $\bar{Y}$  le rispettive medie campionarie, si vuole effettuare la stima di  $\mu$  utilizzando uno dei seguenti stimatori:

$$T_1 = \bar{X}; \quad T_2 = \bar{Y}; \quad T_3 = \frac{\bar{X} + \bar{Y}}{2}; \quad T_4 = \frac{m\bar{X} + n\bar{Y}}{m + n}.$$

- (a) Dire quali degli stimatori precedenti sono corretti.
- (b) Determinare per ciascuno di essi il rischio quadratico (in funzione di  $\sigma^2$ ).
- (c) Mostrare che, nella famiglia dei quattro stimatori considerati,  $T_4$  è quello di rischio quadratico minimo.

### SOLUZIONI

**Primo esercizio.** (a) Cerchiamo la funzione di ripartizione  $G$  di  $X + Y$ . Per ogni  $t \in \mathbb{R}$  si ha

$$\{X + Y \leq t\} = \{X + Y \leq t, Y = 0\} \cup \{X + Y \leq t, Y = 1\}$$

e quindi, dato che gli eventi al secondo membro sono tra loro disgiunti e poiché  $X$  e  $Y$  sono indipendenti,

$$\begin{aligned} G(t) &= P(X + Y \leq t) = P(X + Y \leq t, Y = 0) + P(X + Y \leq t, Y = 1) \\ &= P(X \leq t, Y = 0) + P(X + 1 \leq t, Y = 1) \\ &= P(X \leq t)P(Y = 0) + P(X + 1 \leq t)P(Y = 1) \\ &= (1 - p)P(X \leq t) + pP(X + 1 \leq t) \\ &= (1 - p)P(X \leq t) + pP(X \leq t - 1) = (1 - p)F(t) + pF(t - 1), \end{aligned}$$

dove con  $F$  si è indicata la f. d. r. di una v. a. di legge  $\mathcal{U}(0, 1)$ , e cioè la funzione

$$F(t) = \begin{cases} 0 & \text{per } t < 0 \\ t & \text{per } 0 \leq t \leq 1 \\ 1 & \text{per } t > 1. \end{cases}$$

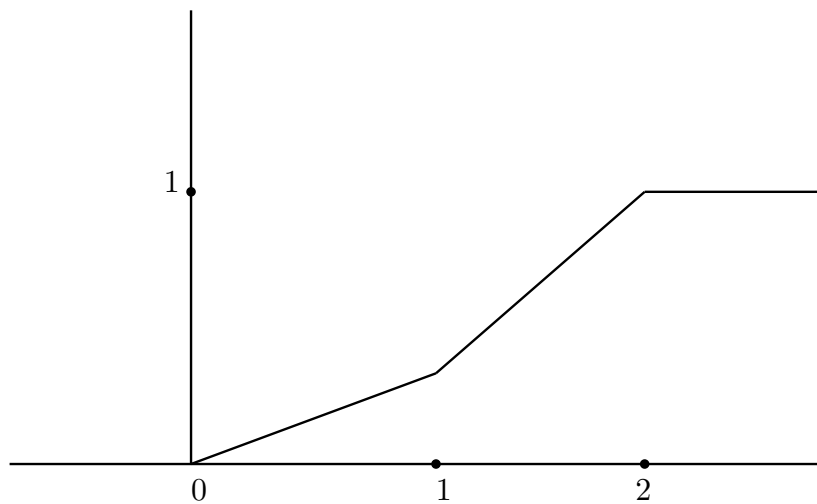
Per scrivere in forma più esplicita la funzione  $G$  ottenuta, distinguiamo i quattro casi (i)  $t < 0$ , (ii)  $0 \leq t < 1$ , (iii)  $1 \leq t < 2$ , (iv)  $t \geq 2$ .

- (i)  $F(t) = F(t - 1) = 0$ , quindi  $G(t) = 0$ .
- (ii)  $F(t) = t$ ;  $F(t - 1) = 0$ , quindi  $G(t) = (1 - p)t$ .
- (iii)  $F(t) = 1$ ,  $F(t - 1) = t - 1$ , quindi  $G(t) = (1 - p) + p(t - 1) = pt + 1 - 2p$
- (iv)  $F(t) = 1$ ,  $F(t - 1) = t - 1$ , quindi  $G(t) = 1$ .

Ricapitolando

$$G(t) = \begin{cases} 0 & \text{per } t < 0 \\ (1-p)t & \text{per } 0 \leq t < 1 \\ pt + 1 - 2p & \text{per } 1 \leq t < 2 \\ 1 & \text{per } t \geq 2; \end{cases}$$

$G$  ha il grafico seguente



(b) Per le proprietà della speranza si ha  $\mathbf{E}[Z] = \mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$   
e

$$\begin{aligned} \mathbf{E}[XZ] &= \mathbf{E}[X(X + Y)] = \mathbf{E}[X^2 + XY] = \mathbf{E}[X^2] + \mathbf{E}[XY] \\ &= \mathbf{E}[X^2] + \mathbf{E}[X]\mathbf{E}[Y], \end{aligned}$$

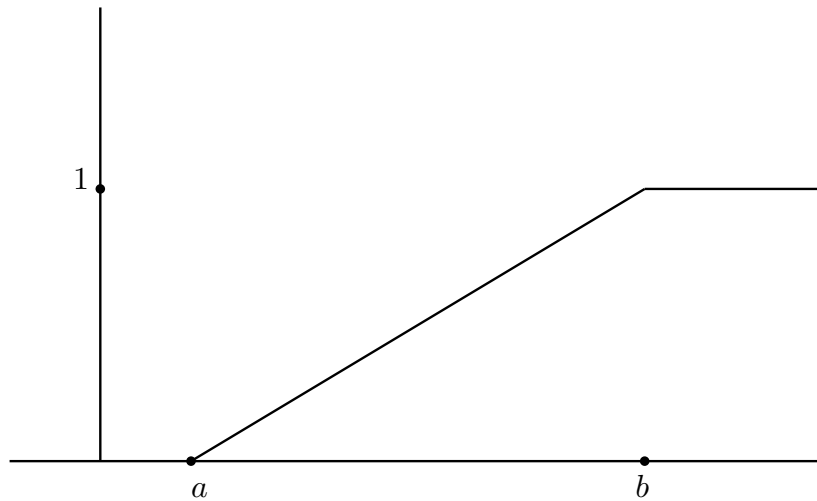
perché  $X$  e  $Y$  sono tra loro indipendenti. Pertanto

$$\begin{aligned} \mathbf{Cov}(X, Z) &= \\ &= \mathbf{E}[XZ] - \mathbf{E}[X]\mathbf{E}[Z] = \mathbf{E}[X^2] + \mathbf{E}[X]\mathbf{E}[Y] - \mathbf{E}[X](\mathbf{E}[X] + \mathbf{E}[Y]) \\ &= \mathbf{E}[X^2] + \mathbf{E}[X]\mathbf{E}[Y] - \mathbf{E}^2[X] - \mathbf{E}[X]\mathbf{E}[Y] = \mathbf{E}[X^2] - \mathbf{E}^2[X] \\ &= \mathbf{Var}X = \frac{1}{12}. \end{aligned}$$

(c) La f. d. r.  $H$  di una v.a. uniforme sull'intervallo  $[a, b]$  vale

$$H(t) = \begin{cases} 0 & \text{per } t < a \\ \frac{t-a}{b-a} & \text{per } a \leq t \leq b \\ 1 & \text{per } t > b. \end{cases}$$

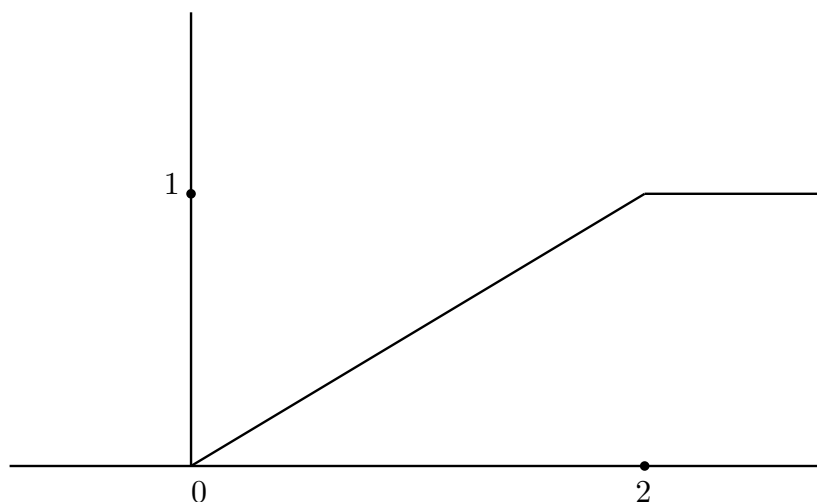
ed ha il grafico seguente



Confrontando con la formula che fornisce  $G$  si vede subito che gli unici valori candidati per essere gli estremi dell'intervallo  $[a, b]$  sono 0 e 2. D'altra parte, con tale scelta di  $a$  e  $b$   $H$  assume la forma

$$H(t) = \begin{cases} 0 & \text{per } t < 0 \\ \frac{t}{2} & \text{per } 0 \leq t \leq 2 \\ 1 & \text{per } t > 2, \end{cases}$$

e il grafico è



Di nuovo confrontando con la formula che dà  $G$ , si vede che si tratta di trovare  $p$  in modo che, per ogni  $t$  appartenente all'intervallo  $[0, 1]$ , si abbia

$$(1 - p)t = \frac{t}{2}$$

e, per ogni  $t$  appartenente all'intervallo  $[1, 2]$ , risulti

$$pt + 1 - 2p = \frac{t}{2}.$$

È immediato verificare che entrambe queste identità sono soddisfatte se e solo se  $p = 1/2$ . Dunque, per  $p = 1/2$ , la v. a.  $Z$  ha legge uniforme sull'intervallo  $[0, 2]$ .

**Secondo esercizio.** Dato che  $X$  e  $Y$  sono due v. a. non negative (in quanto la loro densità è nulla sulla semiretta dei numeri  $x < 0$ ), anche  $Z$  è una v. a. non negativa e dunque, per  $t < 0$  si ha  $P(Z \leq t) = 0$ . Per  $t \geq 0$ , per l'indipendenza di  $X$  e  $Y$  si ha

$$P(Z \leq t) = 1 - P(Z > t) = 1 - P(X > t, Y > t) = 1 - P(X > t)P(Y > t).$$

Sia  $U$  una v. a. avente densità

$$h(x) = \begin{cases} \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} & \text{per } x > 0 \\ 0 & \text{altrimenti,} \end{cases}$$

dove  $\lambda$  e  $\alpha$  sono due costanti positive assegnate (Weibull di parametri  $\lambda$  e  $\alpha$ ). Si ha allora, per  $t \geq 0$ ,

$$P(U > t) = \int_t^{+\infty} \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} dx = \int_{\lambda t^\alpha}^{+\infty} e^{-u} du = e^{-\lambda t^\alpha}$$

(si è utilizzato il cambiamento di variabile  $\lambda x^\alpha = u$ , che dà  $\alpha \lambda x^{\alpha-1} dx = du$ ).

Utilizzando questa formula con  $\lambda = 1$  e  $\alpha = 2$  (resp.  $\lambda = 2$  e  $\alpha = 2$ ) si trova

$$P(X > t) = e^{-t^2} \quad (\text{resp. } P(Y > t) = e^{-2t^2}),$$

e quindi, infine

$$P(Z \leq t) = 1 - e^{-t^2} e^{-2t^2} = 1 - e^{-3t^2}.$$

Ricapitolando si ha

$$P(Z \leq t) = \begin{cases} 0 & \text{per } t < 0 \\ 1 - e^{-3t^2} & \text{per } t \geq 0. \end{cases}$$

Si riconosce allora (per esempio derivando, ma non sarebbe necessario) che  $Z$  ha densità di Weibull di parametri  $\lambda = 3 = 1 + 2$  e  $\alpha = 2$ .

(b) Si ha

$$\begin{aligned} P(Z > t | X > 1) &= \frac{P(Z > t, X > 1)}{P(X > 1)} = \frac{P(X > t, Y > t, X > 1)}{P(X > 1)} \\ &= \frac{P(X > \max(t, 1), Y > t)}{P(X > 1)} = \frac{P(X > \max(t, 1)) P(Y > t)}{P(X > 1)}, \end{aligned}$$

dove l'ultima eguaglianza segue dall'indipendenza di  $X$  e  $Y$ . Se  $t \geq 1$  si ha  $\max(t, 1) = t$ , e dunque la frazione precedente, per le formule ricavate nel punto (a), risulta uguale a

$$\frac{P(X > t)P(Y > t)}{P(X > 1)} = \frac{e^{-t^2} e^{-2t^2}}{e^{-1}} = e^{-3t^2+1}.$$

Invece, per  $t < 1$  si ha  $\max(t, 1) = 1$ , e dunque

$$\frac{P(X > \max(t, 1))P(Y > t)}{P(X > 1)} = \frac{P(X > 1)P(Y > t)}{P(X > 1)} = P(Y > t) = e^{-2t^2}.$$

(c) Ragionando analogamente al punto (b) e ricordando (ved. punto (a)) che  $P(Z > t) = P(X > t)P(Y > t)$ , si trova

$$\begin{aligned} P(X > 1|Z > t) &= \frac{P(Z > t, X > 1)}{P(Z > t)} = \frac{P(X > \max(t, 1))P(Y > t)}{P(X > t)P(Y > t)} \\ &= \frac{P(X > \max(t, 1))}{P(X > t)} = \begin{cases} 1 & \text{per } t \geq 1 \\ \frac{P(X > 1)}{P(X > t)} & \text{per } t < 1 \end{cases} \\ &= \begin{cases} 1 & \text{per } t \geq 1 \\ e^{t^2-1} & \text{per } t < 1. \end{cases} \end{aligned}$$

**Terzo esercizio.** (a) È noto che la media campionaria è uno stimatore corretto della media. Dunque intanto  $T_1$  e  $T_2$  sono stimatori corretti. Gli altri due stimatori sono del tipo  $aT_1 + bT_2$ , dove  $a$  e  $b$  sono due costanti. Per le note proprietà della media e per la correttezza di  $T_1$  e  $T_2$  si ha

$$\mathbf{E}[aT_1 + bT_2] = a\mathbf{E}[T_1] + b\mathbf{E}[T_2] = a\mu + b\mu = (a + b)\mu.$$

Per lo stimatore  $T_3$  si ha  $a = b = 1/2$ , mentre per lo stimatore  $T_4$  si ha  $a = m/(m+n)$ ,  $b = n/(m+n)$ . Poiché in entrambi i casi risulta  $a + b = 1$ , anche  $T_3$  e  $T_4$  sono stimatori corretti.

(b) Poiché tutti gli stimatori considerati sono corretti, per ciascuno di essi il rischio quadratico coincide con la varianza. D'altra parte, per note proprietà della varianza e ricordando che le osservazioni  $(X_1, X_2, \dots, X_m)$  sono indipendenti si ha

$$\mathbf{Var}(T_1) = \mathbf{Var}(\bar{X}) = \mathbf{Var}\left(\frac{\sum_{i=1}^m X_i}{m}\right) = \frac{1}{m^2} \sum_{i=1}^m \mathbf{Var}(X_i) = \frac{m\sigma^2}{m^2} = \frac{\sigma^2}{m},$$

e analogamente

$$\mathbf{Var}(T_2) = \frac{\sigma^2}{n}.$$

Per gli altri due stimatori, osserviamo che  $T_1$  e  $T_2$  sono indipendenti tra loro (si tratta di funzioni di gruppi separati di variabili indipendenti) e quindi, di nuovo per le proprietà della varianza,

$$\mathbf{Var}(aT_1 + bT_2) = a^2\mathbf{Var}(T_1) + b^2\mathbf{Var}(T_2) = \left(\frac{a^2}{m} + \frac{b^2}{n}\right)\sigma^2.$$

Per lo stimatore  $T_3$  ( $a = b = 1/2$ ) si trova

$$\mathbf{Var}(T_3) = \left(\frac{1}{m} + \frac{1}{n}\right)\frac{\sigma^2}{4},$$

mentre per lo stimatore  $T_4$  ( $a = m/(m+n)$ ,  $b = n/(m+n)$ ) si trova

$$\mathbf{Var}(T_4) = \left(\frac{m}{(m+n)^2} + \frac{n}{(m+n)^2}\right)\sigma^2 = \frac{\sigma^2}{m+n}.$$

(c) Si vede subito che  $T_4$  ha rischio minore sia di  $T_1$  che di  $T_2$ . Dunque basta confrontare il suo rischio con quello di  $T_3$ . Si tratta di far vedere che

$$\frac{1}{4}\left(\frac{1}{m} + \frac{1}{n}\right) \geq \frac{1}{m+n}.$$

Con semplici calcoli la disuguaglianza precedente si trasforma nella disuguaglianza  $(m+n)^2 \geq 4mn$ , e cioè  $m^2 - 2mn + n^2 \geq 0$ . Quest'ultima disuguaglianza è vera in quanto si può scrivere nella forma  $(m-n)^2 \geq 0$ .