

Corso di Laurea in Informatica
Esame di Calcolo delle Probabilità e Statistica
Prova scritta del 9 gennaio 2007

Primo esercizio. Si ha a disposizione un cestino contenente 5 uova; si sa che 2 di queste sono marce, mentre le altre 3 sono fresche. Le uova vengono controllate rompendole una a una. Siano X il numero di uova che è necessario rompere per trovare il primo uovo marcio, Y il numero di uova che è necessario rompere, dopo aver trovato il primo uovo marcio, per trovare il secondo, Z il numero di uova (fresche) che restano nel cestino (dopo che è stato trovato anche il secondo uovo marcio).

- (a) Calcolare la densità congiunta di X, Y, Z .
- (b) Calcolare le densità di X , di Y e di Z . Si tratta di v. a. indipendenti?
- (c) In media, quante uova fresche restano una volta trovate entrambe le uova marce?
- (d) Un cuoco ha bisogno di 2 uova per cucinare una frittata. Quanto vale la probabilità che egli riesca a cucinarla, una volta terminato il controllo delle uova?

Secondo esercizio. Si consideri la funzione

$$f(x) = \begin{cases} c|x| & \text{per } x \in (-1, 1) \\ 0 & \text{altrimenti.} \end{cases}$$

dove c è una costante reale.

- (a) Determinare c in modo che f sia una densità di probabilità.
- (b) Sia X una v. a. assolutamente continua di densità f (con la costante determinata in (a)), e sia Y una v. a. avente densità $B(1, p)$. Determinare la funzione di ripartizione della v. a.

$$Z = \max\{XY, X - XY\}.$$

Si tratta di una v. a. continua?

Terzo esercizio. Si ha a disposizione un campione (X_1, X_2) di osservazioni di una v. a. X avente media μ e varianza σ^2 . Per stimare μ , si vuole utilizzare uno stimatore del tipo

$$T = aX_1 + bX_2,$$

dove a e b sono due costanti non negative.

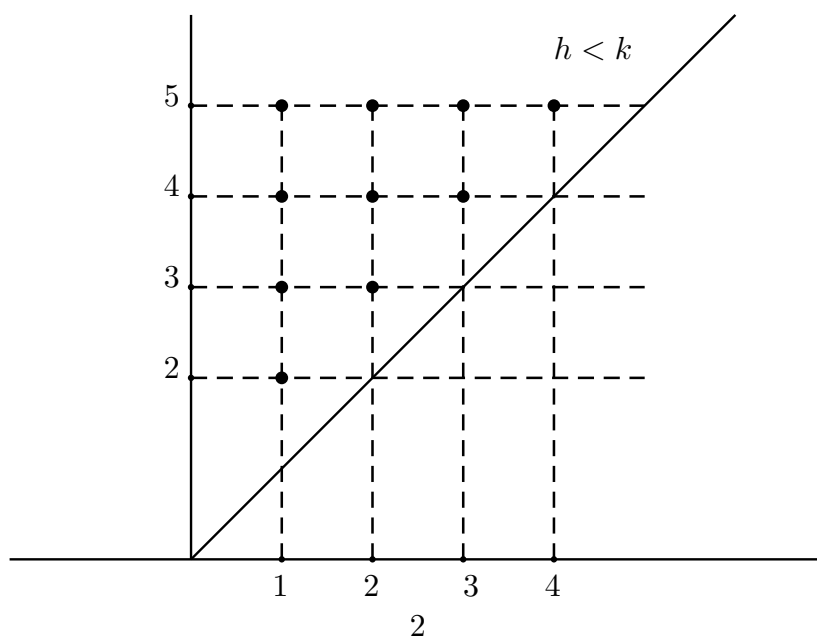
- (a) Determinare per quali valori di a e b T è uno stimatore corretto di μ .
- (b) Determinare il rischio quadratico di T .
- (c) Tra gli stimatori corretti trovati nel punto (a), determinare quelli ammissibili.

SOLUZIONI

Primo esercizio. (a) Ovviamente si ha $Y + Z = 5$, e quindi, intanto, possiamo dire che

$$P(X = h, Y = k, Z = r) = \begin{cases} 0 & \text{per } r \neq 5 - k \\ P(X = h, Y = k) & \text{per } r = 5 - k. \end{cases}$$

Basta allora calcolare la densità congiunta di X e Y . Come è facile capire, il vettore aleatorio (X, Y) assume come valori le coppie (h, k) di numeri interi con $h < k$ appartenenti all'insieme $\{1, 2, 3, 4\} \times \{2, 3, 4, 5\}$; ved. la figura sottostante



Per il modo con cui viene effettuato il controllo, è chiaro che ogni coppia di valori è assunta con la stessa probabilità delle altre, e quindi, per ogni (h, k) appartenente all'insieme A sopra descritto, e cioè

$$A = \{(h, k) \in \mathbb{N}^2 \cap \{1, 2, 3, 4\} \times \{2, 3, 4, 5\}, h < k\}$$

(che ha 10 elementi, come si verifica subito), si avrà

$$P(X = h, Y = k) = \begin{cases} \frac{1}{10} & \text{per } (h, k) \in A \\ 0 & \text{altrimenti.} \end{cases}.$$

Infine

$$P(X = h, Y = k, Z = r) = \begin{cases} \frac{1}{10} & \text{per } (h, k) \in A, r = 5 - k \\ 0 & \text{altrimenti.} \end{cases}$$

(b) Per calcolare le densità marginali di X e di Y , conviene sfruttare la loro densità congiunta. Sommando rispetto a k (cioè sulle rette verticali della figura, si trova la densità di X , e cioè (ved. figura)

$$P(X = 1) = \frac{4}{10}, P(X = 2) = \frac{3}{10}, P(X = 3) = \frac{2}{10}, P(X = 4) = \frac{1}{10};$$

(in un'unica formula $P(X = h) = \frac{5-h}{10}$, $h = 1, 2, 3, 4$); per gli altri valori di h la densità è nulla. In modo analogo si procede per Y (sommando sulle rette orizzontali) e si trova

$$P(Y = 2) = \frac{1}{10}, P(Y = 3) = \frac{2}{10}, P(Y = 4) = \frac{3}{10}, P(Y = 5) = \frac{4}{10};$$

(in un'unica formula $P(Y = k) = \frac{k-1}{10}$, $k = 2, 3, 4, 5$); per gli altri valori di k la densità è nulla.

Per la densità di Z , si può procedere così: dato che $Y + Z = 5$, per ogni r si ha (ved. la formula qui sopra)

$$P(Z = r) = P(5 - Y = r) = P(Y = 5 - r) = \frac{4 - r}{10}, \text{ per } 5 - r = 2, 3, 4, 5,$$

cioè per $r = 0, 1, 2, 3$. Per gli altri valori di r la densità è nulla.

Le variabili X , Y e Z non sono indipendenti: infatti, ad esempio, si ha $P(X = 2, Y = 2) = 0$, mentre $P(X = 2)P(Y = 2) \neq 0$.

(c) Qui si chiede semplicemente la media di Z , che vale

$$\mathbf{E}[Z] = 1 \times \frac{3}{10} + 2 \times \frac{2}{10} + 3 \times \frac{1}{10} = 1.$$

(d) Il cuoco può cucinare la frittata se nel cestino restano almeno 2 uova fresche, dunque se $Z \geq 2$, e la probabilità che questo accada vale

$$P(Z \geq 2) = P(Z = 2) + P(Z = 3) = \frac{2}{10} + \frac{1}{10} = \frac{3}{10}.$$

Secondo esercizio. (a) Si tratta di trovare c in modo che $\int_{\mathbb{R}} f(x) dx = 1$, cioè

$$c \int_{-1}^1 |x| dx = 1.$$

D'altra parte, per simmetria,

$$\int_{-1}^1 |x| dx = 2 \int_0^1 x dx = 2 \left[\frac{x^2}{2} \right]_0^1 = 1,$$

e quindi si conclude che $c = 1$.

(b) Per ogni $t \in \mathbb{R}$ si ha

$$\begin{aligned} P(Z \leq t) &= P(XY \leq t, X - XY \leq t) \\ &= P(XY \leq t, X - XY \leq t, Y = 0) + P(XY \leq t, X - XY \leq t, Y = 1) \\ &= P(0 \leq t, X \leq t, Y = 0) + P(0 \leq t, X \leq t, Y = 1) = P(0 \leq t, X \leq t) \\ &= \begin{cases} 0 & \text{per } t < 0 \\ P(X \leq t) & \text{per } t \geq 0. \end{cases} \end{aligned}$$

D'altra parte

$$\begin{aligned}
P(X \leq t) &= \int_{-\infty}^t f(x) dx = \begin{cases} 0 & \text{per } t < -1 \\ \int_{-1}^t (-x) dx & \text{per } -1 \leq t < 0 \\ \int_{-1}^0 (-x) dx + \int_0^t x dx & \text{per } 0 \leq t < 1 \\ 1 & \text{per } t \geq 1 \end{cases} \\
&= \begin{cases} 0 & \text{per } t < -1 \\ \frac{1-t^2}{2} & \text{per } -1 \leq t < 0 \\ \frac{1+t^2}{2} & \text{per } 0 \leq t < 1 \\ 1 & \text{per } t \geq 1. \end{cases}
\end{aligned}$$

Si conclude che

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

La funzione di ripartizione G non è continua per $t = 0$. Infatti

$$\lim_{t \rightarrow 0^-} G(t) = 0 \neq G(0) = \frac{1}{2}.$$

Dunque Z non è una v. a. continua.

Terzo esercizio. La media di T vale

$$\mathbf{E}[T] = \mathbf{E}[aX_1 + bX_2] = a\mathbf{E}[X_1] + b\mathbf{E}[X_2] = a\mu + b\mu = (a+b)\mu.$$

Dunque T è uno stimatore corretto di μ se e solo se $a+b=1$.

(b) *Primo metodo* (un po' lungo).

$$\begin{aligned}
R_T &= \mathbf{E}[(T - \mu)^2] = \mathbf{E}[(aX_1 + bX_2 - \mu)^2] \\
&= \mathbf{E}[a^2X_1^2 + b^2X_2^2 + \mu^2 - 2a\mu X_1 - 2b\mu X_2 + 2abX_1X_2] \\
&= a^2\mathbf{E}[X_1^2] + b^2\mathbf{E}[X_2^2] + \mu^2 - 2a\mu^2 - 2b\mu^2 + 2ab\mathbf{E}[X_1X_2] \\
&= a^2\mathbf{E}[X_1^2] + b^2\mathbf{E}[X_2^2] + \mu^2 - 2a\mu^2 - 2b\mu^2 + 2ab\mathbf{E}[X_1]\mathbf{E}[X_2]
\end{aligned}$$

dato che X_1 e X_2 sono indipendenti. D'altra parte (dato che X_1 e X_2 hanno la stessa legge)

$$\mathbf{E}[X_2^2] = \mathbf{E}[X_1^2] = \mathbf{Var}[X_1] + \mathbf{E}^2[X_1] = \sigma^2 + \mu^2.$$

Si ottiene dunque, sostituendo

$$\begin{aligned} R_T &= a^2(\sigma^2 + \mu^2) + b^2(\sigma^2 + \mu^2) + \mu^2 - 2a\mu^2 - 2b\mu^2 + 2ab\mu^2 \\ &= (a^2 + b^2)\sigma^2 + \mu^2(a + b - 1)^2. \end{aligned}$$

Secondo metodo.

$$\begin{aligned} R_T &= \mathbf{E}[(T - \mu)^2] = \mathbf{E}[(T - \mathbf{E}[T] + \mathbf{E}[T] - \mu)^2] \\ &= \mathbf{E}[(T - \mathbf{E}[T])^2 + 2(T - \mathbf{E}[T])(\mathbf{E}[T] - \mu) + (\mathbf{E}[T] - \mu)^2] \\ &= \mathbf{E}[(T - \mathbf{E}[T])^2] + 2\mathbf{E}[(T - \mathbf{E}[T])(\mathbf{E}[T] - \mu)] + \mathbf{E}[(\mathbf{E}[T] - \mu)^2], \end{aligned}$$

sviluppando il quadrato con la regola del binomio e usando le proprietà della media. D'altra parte, per le note proprietà della varianza (ricordare che X_1 e X_2 sono indipendenti)

$$\begin{aligned} (i) \quad \mathbf{E}[(T - \mathbf{E}[T])^2] &= \mathbf{Var}T = \mathbf{Var}(aX_1 + bX_2) = a^2\mathbf{Var}X_1 + b^2\mathbf{Var}X_2 \\ &= (a^2 + b^2)\sigma^2; \end{aligned}$$

inoltre

$$\begin{aligned} (ii) \quad 2\mathbf{E}[(T - \mathbf{E}[T])(\mathbf{E}[T] - \mu)] &= 2(\mathbf{E}[T] - \mu)\mathbf{E}[(T - \mathbf{E}[T])] \\ &= 2(\mathbf{E}[T] - \mu)(\mathbf{E}[T] - \mathbf{E}[T]) = 0; \end{aligned}$$

e infine

$$(iii) \quad \mathbf{E}[(\mathbf{E}[T] - \mu)^2] = (\mathbf{E}[T] - \mu)^2 = ((a + b)\mu - \mu)^2 = (a + b - 1)^2\mu^2.$$

Sostituendo nella formula che dà R_T , si trova di nuovo

$$R_T = (a^2 + b^2)\sigma^2 + \mu^2(a + b - 1)^2.$$

(c) Dal punto (a) sappiamo che T è corretto se $a + b - 1 = 0$, e dunque, in questo caso, si ha

$$R_T = (a^2 + b^2)\sigma^2.$$

Cominciamo a chiederci intanto se nella famiglia considerata esiste uno stimatore preferibile a tutti gli altri, cioè tale che il suo rischio sia minore o uguale a quello degli altri. In altri termini, si tratta di trovare il minimo della funzione $f(a, b) = (a^2 + b^2)\sigma^2$ al variare di (a, b) nell'insieme

$$\{a \geq 0, \quad b \geq 0, \quad a + b = 1\}$$

ovvero, come si dice, di risolvere il *problema di minimo vincolato*

$$\begin{cases} \min (a^2 + b^2)\sigma^2 \\ a \geq 0, \quad b \geq 0 \\ a + b = 1. \end{cases}$$

Dato che $b = 1 - a$, il problema si riduce facilmente ad un problema di minimo con una sola variabile, cioè

$$\begin{cases} \min (a^2 + (1 - a)^2)\sigma^2 \\ 0 \leq a \leq 1, \end{cases}$$

Con i noti metodi di analisi, si vede che la funzione $h(a) = a^2 + (1 - a)^2$, $0 \leq a \leq 1$ assume il minimo (assoluto) per $a = 1/2$ (e quindi si avrà $b = 1 - (1/2) = 1/2$), e tale minimo vale $h(1/2) = 1/2$. Pertanto lo stimatore (media campionaria)

$$T = \frac{1}{2}X_1 + \frac{1}{2}X_2 = \frac{X_1 + X_2}{2}$$

è preferibile a tutti gli altri, e il suo rischio vale $\sigma^2/2$. Poiché ogni altro stimatore ha un rischio strettamente maggiore, la media campionaria è uno stimatore ammissibile.