

Prova Scritta di Probabilità e Statistica Laurea in Matematica 17 luglio 2012	Cognome: _____ Nome: _____ Matricola: _____
--	---

ESERCIZIO 1. Siano (Ω, P) uno spazio di probabilità discreto, e X, Y variabili casuali scalari indipendenti.

- (a) Dimostrare che se $X, Y \in L^1$, allora $XY \in L^1$ e

$$E(XY) = E(X)E(Y).$$

(si può usare senza dimostrare il fatto che la densità congiunta è il prodotto delle densità marginali).

- (b) Dimostrare che se $X, Y \in L^2$, allora $XY \in L^2$ e

$$\text{Var}(XY) = E(X^2)\text{Var}(Y) + E(Y^2)\text{Var}(X).$$

SOLUZIONE.

- (a) Cominciamo col mostrare che XY ammette valor medio. Usando il fatto che $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ si ha:

$$\begin{aligned} E(|XY|) &= \sum_{x,y} |x||y|p_{X,Y}(x, y) = \sum_x |x|p_X(x) \sum_y |y|p_Y(y) \\ &= E(|X|)E(|Y|) < +\infty. \end{aligned}$$

La dimostrazione che $E(XY) = E(X)E(Y)$ è del tutto analoga.

- (b) Se $X, Y \in L^2$ allora $X^2, Y^2 \in L^1$. Per quanto visto sopra, segue che $X^2Y^2 \in L^1$, cioè $XY \in L^2$. Inoltre dalle identità

$$E(X^2Y^2) = E(X^2)E(Y^2) \quad E(XY) = E(X)E(Y),$$

l'identità

$$\text{Var}(XY) = E(X^2)\text{Var}(Y) + E(Y^2)\text{Var}(X).$$

segue facilmente.

ESERCIZIO 2. Usando solo la definizione di spazio di probabilità, dimostrare che per ogni n -pla di eventi $A_1, A_2, \dots, A_n$, si ha

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

SOLUZIONE. Dimostriamo per induzione che la disuguaglianza vale per ogni $n \geq 2$. Per $n = 2$,

$$A_1 \cup A_2 = [A_1 \setminus (A_1 \cap A_2)] \cup (A_1 \cap A_2) \cup [A_2 \setminus (A_1 \cap A_2)],$$

e quest'ultima unione è disgiunta. Pertanto, per l'additività della probabilità,

$$P(A_1 \cup A_2) = P[A_1 \setminus (A_1 \cap A_2)] + P(A_1 \cap A_2) + P[A_2 \setminus (A_1 \cap A_2)].$$

Inoltre

$$P[A_1 \setminus (A_1 \cap A_2)] = P(A_1) - P(A_1 \cap A_2),$$

e similmente con A_2 al posto di A_1 , essendo A_1 unione disgiunta di $A_1 \setminus (A_1 \cap A_2)$ e $A_1 \cap A_2$. Mettendo tutto assieme si giunge a

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2).$$

Pertanto la tesi è vera per $n = 2$ dimostriamo ora il passo induttivo. Usando quanto l'ipotesi induttiva e quanto dimostrato per $n = 2$,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\left[\bigcup_{i=1}^n A_i\right] \cup A_{n+1}\right) \leq P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) \\ &\leq \sum_{i=1}^n P(A_i) + P(A_{n+1}) = \sum_{i=1}^{n+1} P(A_i) \end{aligned}$$

ESERCIZIO 3. Si ricordi il modello della passeggiata aleatoria semplice: $\Omega = \{-1, 1\}^N$, $P =$ probabilità uniforme su Ω , se $x \in \Omega$, $s_n := x_1 + \cdots + x_n$. Calcolare la probabilità condizionata

$$P(s_{2n} = 0 | s_{2k} = 0)$$

sia per $k < n$ che per $k > n$.

SOLUZIONE.

Sia $k < n$. Ricordiamo che non è restrittivo assumere $N = 2n$. Inoltre $P(s_{2k} = 0) = \binom{2k}{k}/2^k$. Si tratta di calcolare la probabilità dell'evento $\{s_{2n} = 0\} \cap \{s_{2k} = 0\}$. Gli elementi di questo evento sono tutti e soli i cammini che sono in 0 dopo $2k$ passi ($\binom{2k}{k}$ scelte possibili) e che poi sono ancora in 0 dopo ulteriori $2n - 2k$ passi ($\binom{2n-2k}{n-k}$ scelte possibili). In altre parole

$$P(\{s_{2n} = 0\} \cap \{s_{2k} = 0\}) = \frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^n},$$

cioè

$$P(s_{2n} = 0 | s_{2k} = 0) = \frac{\frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^n}}{\frac{1}{2^k} \binom{2k}{k}} = \frac{1}{2^{2n-2k}} \binom{2n-2k}{n-k}.$$

Se invece $k > n$, possiamo assumere $N = 2k$. Lo stesso ragionamento di prima porta a

$$P(\{s_{2n} = 0\} \cap \{s_{2k} = 0\}) = \frac{\binom{2n}{n} \binom{2k-2n}{k-n}}{2^{2k}},$$

$$P(s_{2n} = 0 | s_{2k} = 0) = \frac{\frac{\binom{2n}{n} \binom{2k-2n}{k-n}}{2^{2k}}}{\frac{1}{2^k} \binom{2k}{k}} = \frac{\binom{2n}{n} \binom{2k-2n}{k-n}}{\binom{2k}{k}}.$$

ESERCIZIO 4. Siano X e Y due variabili casuali indipendenti, assolutamente continue, con la stessa densità

$$f(x) := \frac{1}{2}e^{-|x|}.$$

- (a) Calcolare la funzione generatrice dei momenti della variabile casuale X .
- (b) Calcolare la densità e la funzione generatrice dei momenti della variabile casuale $X + Y$.

SOLUZIONE.

(a)

$$\gamma_X(t) = E(e^{tX}) = \frac{1}{2} \int_{-\infty}^0 e^{tx} e^x dx + \frac{1}{2} \int_0^{+\infty} e^{tx} e^{-x} dx = \frac{1}{1-t^2}$$

per $|t| < 1$ mentre $\gamma_X(t) = +\infty$ se $|t| \geq 1$.

(b) Sia $z > 0$.

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{+\infty} f(x)f(z-x)dx = \frac{1}{4} \int_{-\infty}^0 e^x e^{x-z} dx \\ &\quad + \frac{1}{4} \int_0^z e^{-x} e^{x-z} dx + \frac{1}{4} \int_z^{+\infty} e^{-x} e^{z-x} dx = \frac{1}{4}(1+z)e^{-z}. \end{aligned}$$

Facendo il calcolo analogo per $z < 0$, troviamo

$$f_{X+Y}(z) = (1+|z|)e^{-|z|}.$$

Infine, per l'indipendenza di X e Y ,

$$\gamma_{X+Y}(t) = \gamma_X(t)\gamma_Y(t) = \gamma_X^2(t).$$

ESERCIZIO 5. Sia $X \in L^1$. Sappiamo che se $f(x) = ax + b$, allora $E(f(X)) = f(E(X))$.

(a) Consideriamo ora una funzione del tipo

$$f(x) := \max(a_1x + b, a_2x + b).$$

Si mostri che $f(X) \in L^1$,

$$E(f(X)) \geq f(E(X)).$$

(Sugg.: usare la monotonia del valor medio)

(b) Si generalizzi l'argomento del punto precedente per mostrare che $E(f(X)) \geq f(E(X))$ per ogni funzione del tipo

$$f(x) := \sup_{i \in I} (a_i x + b_i),$$

dove I è un arbitrario insieme di indici, e *assumendo* (non è vero in generale) che $f(X) \in L^1$.

SOLUZIONE. Se I è un arbitrario insieme di indici, sia $f_i(x) = a_i x + b_i$. Allora $E(f_i(X)) = f_i(E(X))$. Poiché $f \geq f_i$, dalla monotonià del valor medio segue che

$$E(f(X)) \geq f_i(E(X)).$$

Prendendo il $\sup_i$ nella precedente disuguaglianza si ottiene $E(f(X)) \geq f(E(X))$. Nel caso in cui $I = \{1, 2\}$ non è necessario supporre che $f(X) \in L^1$, in quanto

$$E(|f(X)|) \leq E[|a_1 X| + |b_1| + |a_2 X| + |b_2|] = (|a_1| + |a_2|)E(|X|) + |b_1| + |b_2| < +\infty.$$

ESERCIZIO 6. Un produttore di lampade da proiettore ha due centri di produzione. Nel centro A si lavora in modo molto accurato, e le lampade prodotte hanno una durata media di 400 ore di utilizzo. Nel centro B la produzione non raggiunge gli stessi standard, e le lampade hanno una durata media di 300 ore. In entrambi i casi la distribuzione della durata di una lampadina è esponenziale. Supponiamo di acquistare una lampadina, e di non avere, al momento dell'acquisto, alcuna informazione sulla sua provenienza: riteniamo pertanto che essa possa provenire con ugual probabilità dal centro A o dal centro B. Dopo aver usato la lampadina per 500 ore, essa è ancora funzionante. Sulla base di ciò, qual è la probabilità che essa provenga dal centro A?

SOLUZIONE. Consideriamo gli eventi A = “la lampadina proviene dal centro A”, B = “la lampadina proviene dal centro B”, C = “la lampadina funziona dopo 500 ore di utilizzo”. Siano inoltre $X \sim \text{Exp}(1/400)$ e $Y \sim \text{Exp}(1/300)$. Sappiamo che $P(A) = P(B) = 1/2$, e

$$P(C|A) = P(X \geq 500) = e^{-5/4} \quad P(C|B) = P(Y \geq 500) = e^{-5/3}.$$

Quindi

$$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)} = \frac{e^{-5/4}}{e^{-5/4} + e^{-5/3}}.$$

ESERCIZIO 7. Siano $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n$ variabili casuali indipendenti tali che $X_k \sim \text{Exp}(\lambda)$, $Y_k \sim \text{Exp}(\mu)$, per $1 \leq k \leq n$. Si determini la funzione di ripartizione delle seguenti variabili casuali:

$$Z := \min_{1 \leq k \leq n} \{\max(X_k, Y_k)\},$$

$$W := \max\left(\min_{1 \leq k \leq n} X_k, \min_{1 \leq k \leq n} Y_k\right).$$

SOLUZIONE. Sia $z > 0$.

$$\begin{aligned} F_Z(z) &= 1 - P(Z > z) = 1 - \prod_{k=1}^n P(\max(X_k, Y_k) > z) = 1 - \prod_{k=1}^n [1 - P(\max(X_k, Y_k) \leq z)] \\ &= 1 - \prod_{k=1}^n [1 - P(X_k \leq z) P(Y_k \leq z)] = 1 - \prod_{k=1}^n \left[1 - (1 - e^{-\lambda z})(1 - e^{-\mu z})\right] \\ &= 1 - \left[e^{-\lambda z} + e^{-\mu z} - e^{-(\lambda+\mu)z}\right]^n. \end{aligned}$$

$$\begin{aligned} F_W(z) &= P\left(\min_{1 \leq k \leq n} X_k \leq z\right) P\left(\min_{1 \leq k \leq n} Y_k \leq z\right) \\ &= \left[1 - P\left(\min_{1 \leq k \leq n} X_k > z\right)\right] \left[1 - P\left(\min_{1 \leq k \leq n} Y_k > z\right)\right] \\ &= \left[1 - \prod_{k=1}^n P(X_k > z)\right] \left[1 - \prod_{k=1}^n P(Y_k > z)\right] = \left[1 - e^{-n\lambda z}\right] \left[1 - e^{-n\mu z}\right] \end{aligned}$$

ESERCIZIO 8. Un'urna contiene N palline numerate da 1 a N . Si eseguono due estrazioni *con reimmissione*, e sia D la differenza tra il numero estratto più grande e quello più piccolo. Si determini media e momento secondo di D . (Sugg: ricordare che $\sum_1^N i = \frac{N(N+1)}{2}$ e $\sum_1^N i^2 = \frac{N(N+1)(2N+1)}{6}$; inoltre per il calcolo di $E(D^2)$ non è consigliabile usare la densità di D)

SOLUZIONE. Siano X e Y rispettivamente il primo e il secondo numero estratto. X e Y sono due variabili casuali indipendenti, con distribuzione $P(X = k) = P(Y = k) = \frac{1}{N}$, per $k = 1, 2, \dots, N$. Facilmente:

$$E(X) = E(Y) = \frac{1}{N} \sum_{k=1}^N k = \frac{N+1}{2},$$

$$E(X^2) = E(Y^2) = \frac{1}{N} \sum_{k=1}^N k^2 = \frac{(N+1)(2N+1)}{6}.$$

Inoltre

$$E(D^2) = E[(X - Y)^2] = 2E(X^2) - 2E^2(X) = \frac{(N+1)(2N+1)}{3} - \frac{(N+1)^2}{2} = \frac{N^2 - 1}{6}.$$

Per il calcolo di $E(D)$ ricorriamo alla densità di D . Per $k = 1, \dots, N-1$:

$$P(D = k) = \sum_{i=1}^{N-k} [P(X = i)P(Y = i+k) + P(Y = i)P(X = i+k)] = \frac{2(N-k)}{N^2}.$$

per cui

$$E(D) = \sum_{k=1}^{N-1} k \frac{2(N-k)}{N^2} = \frac{N^2 - 1}{3N}.$$

ESERCIZIO 9. Sia $(X_n)_{n \geq 1}$ una successione di variabili casuali i.i.d. con $X_n \sim \text{Exp}(\lambda)$. Si dica per quali valori di λ

$$P(X_1 + X_2 + \cdots + X_{100} > 10^3) < 0.05.$$

SOLUZIONE. Ricordando che $E(X_n) = 1/\lambda$, $\text{Var}(X_n) = 1/\lambda^2$, e posto $Z_n := \frac{\bar{X}_n - 1/\lambda}{1/\lambda} \sqrt{n}$, si ha

$$P(X_1 + X_2 + \cdots + X_{100} > 10^3) = P(Z_{100} > 10(10\lambda - 1)) \simeq P(Z > 10(10\lambda - 1)) = 1 - \Phi(10(10\lambda - 1)).$$

quindi dev'essere $1 - \Phi(10(10\lambda - 1)) < 0.05$, cioè $10(10\lambda - 1) > \Phi^{-1}(0.95) = 1.64$, che vale se e solo se

$$\lambda > \frac{1}{10} + \frac{1.64}{100}.$$

ESERCIZIO 10. Un'urna contiene $n \geq 2$ palline, di cui 2 rosse e $n - 2$ verdi. Si effettuano n estrazioni successive, senza reimmissione, fino a svuotare l'urna. Numerando da 1 a n le estrazioni, si X_n il numero dell'estrazione in cui si estrae la seconda pallina rossa.

- (a) Determinare la funzione di ripartizione e la densità di X_n .
 (b) Posto $Y_n := \frac{X_n}{n}$, determinare, per ogni $x \in \mathbb{R}$,

$$F(x) := \lim_{n \rightarrow +\infty} F_{Y_n}(x).$$

Qual è la distribuzione di una variabile casuale che ha F come funzione di ripartizione?

SOLUZIONE.

- (a) Anzitutto si noti che la variabile casuale $2 \leq X_n \leq n$. Se $2 \leq k \leq n$, calcoliamo $P(X_n \leq k)$. L'evento $\{X_n \leq k\}$ equivale all'affermazione che nelle prime k estrazioni sono state estratte entrambe le palline rosse. Per calcolarne la probabilità, consideriamo il seguente modello per le n estrazioni. Supponiamo di numerare le palline da 1 a n , in cui le palline rosse sono la 1 e la 2. Posto $\Omega = S_n$, per $\sigma \in \Omega$, intendiamo che $\sigma(i)$ è il numero dell'estrazione in cui è stata estratta la pallina i . P è la probabilità uniforme su Ω , e $X_n(\sigma) := \max(\sigma(1), \sigma(2))$. Dunque $\sigma \in \{X_n \leq k\}$ se e solo se $\sigma(1), \sigma(2) \leq k$. Il numero di tali permutazioni è $k(k-1)[(n-2)!]$. Perciò

$$P(X_n \leq k) = \frac{k(k-1)[(n-2)!]}{n!} = \frac{k(k-1)}{n(n-1)}.$$

In particolare

$$P(X = k) = P(X \leq k) - P(X \leq k-1) = \frac{k(k-1) - (k-1)(k-2)}{n(n-1)} = \frac{2(k-1)}{n(n-1)}.$$

- (b) Essendo $2 \leq X_n \leq n$, si ha che $F_{Y_n}(x) = 0$ se $x \leq 0$ o $x \leq 1$. Per $x \in (0, 1)$, se $\lfloor \cdot \rfloor$ denota la parte intera di un numero reale,

$$F_{Y_n}(x) = P(X_n \leq nx) = P(X_n \leq \lfloor nx \rfloor) = \frac{\lfloor nx \rfloor(\lfloor nx \rfloor - 1)}{n(n-1)}.$$

Da ciò segue che

$$F(x) := \lim_{n \rightarrow +\infty} F_{Y_n}(x) = x^2.$$

Si vede inoltre che F è la funzione di ripartizione di una variabile casuale assolutamente continua con densità

$$f(x) = F'(x) = 2x \mathbf{1}_{(0,1)}(x).$$