Dipartimento di Matematica, Roma Tre
Pietro Caputo

2018-19, II semestre
9 luglio, 2019

CP210 Introduzione alla Probabilità: Esame 2

Cognome	
Nome	
Matricola	
Firma	

Nota:

1. L'unica cosa che si può usare durante l'esame è una penna o una matita. Tutto il resto (calcolatrice, libri, appunti, altri fogli di carta, ...) deve essere messo da parte.
2. Risposte implicite sotto forma di coefficienti binomiali, potenze, esponenziali ecc. sono benvenute. Mostrate in dettaglio il vostro lavoro.
3. Non parlate durante l'esame. Copiare o far copiare non è tollerabile.
4. Scrivete il vostro nome su ogni pagina. In caso di utilizzo di piu' pagine per un singolo esercizio indicare chiaramente l'ordine.
5. Il punteggio massimo per ogni esercizio è indicato nel testo. Notare che già con quattro esercizi risolti correttamente si arriva vicino al trenta.

Buon lavoro!

esercizio	1	2	3	4	5	6	totale
punti							
su	6	6	6	6	6	6	36

Nome:______________________________________

1. **(6 Punti)** Un'urna contiene n palline nere e b palline bianche. In funzione di $n \in \mathbb{N}$ e $b \in \mathbb{N}$, calcolare la probabilità di estrarre due palline dello stesso colore nei due seguenti casi:

   (a) l'estrazione avviene senza reinserimento;

   (b) l'estrazione avviene con reinserimento.

   (c) In quale dei due casi la probabilità è maggiore ?

Soluzione:

a). Supponiamo che le palline siano tutte distinte, quindi abbiamo n palline nere numerate da 1 a n e b palline bianche numerate da $n+1$ a $n+b$. Se non c'e' reinserimento le prime due estrazioni possono risultare in $(n+b)(n+b-1)$ possibili esiti distinti. Di questi, $n(n-1)+b(b-1)$ hanno lo stesso colore. Allora la probabilità vale

$$\frac{n(n-1)+b(b-1)}{(n+b)(n+b-1)}.$$

b). Se c'e' reinserimento allora le prime due estrazioni possono risultare in $(n+b)^2$ possibili esiti distinti. Di questi n^2+b^2 hanno lo stesso colore. Allora la probabilità vale

$$\frac{n^2+b^2}{(n+b)^2}.$$

c). Scrivendo $x = n^2+b^2$ e $y = (n+b)^2$, $a = n+b$, si ha che la prima probabilità vale $(x^2-a)/(y^2-a)$, mentre la seconda vale x^2/y^2. Poiché $a > 0$ e $x < y$ si vede facilmente che nel secondo caso la probabilità è sempre maggiore, per qualsiasi valore di $n, b \in \mathbb{N}$.

Nome:______________________________________

2. **(6 Punti)** Siano $X_1, \ldots, X_{25}$ variabili di Poisson indipendenti con media $\mathbb{E}[X_i] = 1$. Consideriamo la probabilità

$$p = \mathbb{P}\left(\sum_{k=1}^{25} X_i > 30\right).$$

   (a) Dare una stima dall'alto per p usando la disuguaglianza di Markov;

   (b) Fornire un valore approssimato per p usando il teorema del limite centrale;

   (c) Cosa cambia nei punti a) e b) se le X_i sono variabili esponenziali indipendenti con media $\mathbb{E}[X_i] = 1$?

Soluzione:

a). Dalla disuguaglianza di Markov abbiamo

$$p = \mathbb{P}\left(\sum_{k=1}^{25} X_i > 30\right) \leqslant \frac{1}{30}\sum_{k=1}^{25} \mathbb{E}[X_i] = \frac{5}{6}.$$

b). Sia $S = \sum_{k=1}^{25} X_i$. La varianza, per l'indipendenza, vale:

$$\mathrm{Var}(S) = 25\mathrm{Var}(X_1) = 25,$$

dove usiamo il fatto che $\mathrm{Var}(X_1) = 1$ essendo X_1 una Poisson di parametro 1. Dunque per il teorema del limite centrale $(S - 25)/\sqrt{25}$ è approssimativamente normale. Segue che

$$p = \mathbb{P}((S-25)/\sqrt{25} > (30-25)/\sqrt{25}) \approx \mathbb{P}(Z > 1) = 1 - \Phi(1) \approx 0.16$$

Nota: se usiamo l'approssimazione di continuità otteniamo

$$p = \mathbb{P}((S-25)/\sqrt{25} \geqslant (30.5-25)/\sqrt{25}) \approx \mathbb{P}(Z \geqslant 1.1) = 1 - \Phi(1.1) \approx 0.14.$$

c). Notiamo che se le variabili fossero esponenziali indipendenti con media $\mathbb{E}[X_i] = 1$ si avrebbe lo stesso $\mathbb{E}[S] = 25$ e $\mathrm{Var}(S) = 25$, e dunque non cambia nulla nei due punti precedenti.

Nome:____

3. **(6 Punti)** Siano Z_1, Z_2 due variabili normali standard indipendenti. Calcolare la densità di probabilità delle variabili

   (a) $Z_1 + Z_2$;

   (b) $Z_1 - Z_2$;

   (c) $Z_1^2 + Z_2^2$.

Soluzione:

a). La somma di normali $N(\mu_i, \sigma_i^2)$ indipendenti è normale con parametri $\mu = \sum_i \mu_i$ e $\sigma^2 = \sum_i \sigma_i^2$. Allora $Z_1 + Z_2$ è normale di media $\mu = 0$ e varianza $\sigma^2 = 2$. Dunque ha densità di probabilità

$$f_{Z_1+Z_2}(x) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{x^2}{4}\right), \qquad x \in \mathbb{R}.$$

b). La variabile $-Z_2$ è normale standard indipendente da Z_1. Dunque $Z_1 - Z_2$ ha la stessa distribuzione di $Z_1 + Z_2$, e quindi ha densità di probabilità

$$f_{Z_1-Z_2}(x) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{x^2}{4}\right), \qquad x \in \mathbb{R}.$$

c). La variabile Z_1^2 ha funzione di distribuzione

$$F(y) = \mathbb{P}(Z_1^2 \leqslant y) = \mathbb{P}(|Z_1| \leqslant \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)\,dx$$

dove $f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ e $y > 0$. Derivando rispetto a y si ha

$$F'(y) = f(\sqrt{y})\frac{1}{2\sqrt{y}} + f(-\sqrt{y})\frac{1}{2\sqrt{y}} = f(\sqrt{y})\frac{1}{\sqrt{y}} = \frac{1}{\sqrt{2\pi}}\, y^{-\frac{1}{2}}\, e^{-\frac{y}{2}}.$$

Allora Z_1^2 è una gamma di parametri $\alpha = \frac{1}{2}$ e $\lambda = \frac{1}{2}$. Poiché la somma di gamma con parametri (α, λ) e (β, λ) indipendenti è una gamma con parametri $(\alpha+\beta, \lambda)$ si ha che $Z_1^2 + Z_2^2$ ha la densità

$$f_{Z_1^2+Z_2^2}(x) = \frac{1}{2}\, e^{-\frac{x}{2}}\, \mathbf{1}_{x \geqslant 0}.$$

Dunque $Z_1^2 + Z_2^2$ è una variabile esponenziale di parametro $\frac{1}{2}$.

Nome:_______________________________________

4. **(6 Punti)** Un account di posta elettronica riceve messaggi secondo un processo di Poisson con una media di 4 messaggi ogni ora. Stimiamo che il 25 per cento di questi finisce nella cartella spam. Supponendo che i messaggi spam e i messaggi non spam seguano due processi di Poisson indipendenti, calcolare:
   (a) Il numero medio di messaggi spam ricevuti in 24 ore;
   (b) La probabilità di ricevere due messaggi spam in una data ora;
   (c) Sapendo che in una data ora sono stati ricevuti 4 messaggi in totale, dire qual è la probabilità che 2 di questi siano spam.

Soluzione: (a). In 24 ore si ha una media di $4 * 24 = 96$ messaggi totali, dunque il numero medio di messaggi spam è $96/4 = 24$.

(b). I messaggi spam costituiscono un processo di Poisson con media 1 all'ora. Ne segue che il numero di tali messaggi in un'ora è una variabile di Poisson di parametro $\lambda = 1$. Dunque la probabilità richiesta vale

$$\frac{\lambda^2}{2} e^{-\lambda} = 0.5e^{-1} \approx 0.18$$

(c). In una data ora, sia N_1 il numero di messaggi non spam e sia N_2 il numero di messaggi spam. Allora N_1, N_2 sono due v.a. di Poisson indipendenti, con parametri $\lambda_1 = 3$ e $\lambda_2 = 1$ rispettivamente. Scriviamo

$$\begin{aligned}\mathbb{P}(N_2 = 2|N_1 + N_2 = 4) &= \frac{\mathbb{P}(N_2 = 2, N_1 = 2)}{\mathbb{P}(N_1 + N_2 = 4)} \\ &= \frac{\mathbb{P}(N_2 = 2)\mathbb{P}(N_1 = 2)}{(4^4/4!)e^{-4}} = \frac{(1^2/2!)e^{-1}(3^2/2!)e^{-3}}{(4^4/4!)e^{-4}} = \frac{54}{256}.\end{aligned}$$

Nome:____

5. **(6 Punti)** Due giocatori lanciano 3 dadi A,B,C. Il giocatore 1 ha il punteggio X dato dalla somma dei dadi A,B, mentre il giocatore 2 ha il punteggio Y dato dalla somma dei dadi B,C. Calcolare:
   (a) la probabilità condizionata $\mathbb{P}(X = 8|Y = 6)$.
   (b) il valore atteso $\mathbb{E}\big[|X - Y|\big]$

Soluzione: (a). Notiamo che $\mathbb{P}(Y = 6)$ è la probabilità che la somma di due dadi valga 6 e dunque

$$\mathbb{P}(Y = 6) = \frac{5}{36}.$$

Calcoliamo ora $\mathbb{P}(X = 8, Y = 6)$. L'evento $\{X = 8, Y = 6\}$ si realizza in 4 modi differenti: $(A = 6, B = 2, C = 4)$, $(A = 5, B = 3, C = 3)$, $(A = 4, B = 4, C = 2)$, $(A = 3, B = 5, C = 1)$. Ognuno di questi eventi ha probabilità 6^{-3}. Allora

$$\mathbb{P}(X = 8, Y = 6) = \frac{4}{6^3}.$$

Concludiamo che

$$\mathbb{P}(X = 8|Y = 6) = \frac{\mathbb{P}(X = 8, Y = 6)}{\mathbb{P}(Y = 6)} = \frac{2}{15}.$$

(b). La differenza $X - Y$ equivale a $A - C$ dove A, C indicano la faccia del dado A e del dado C rispettivamente, e dunque assume i valori $\pm 1, \pm 2, \pm 3, \pm 4, \pm 5$. Notiamo che

$$\begin{aligned}\mathbb{P}(|X - Y| = 1) &= \mathbb{P}(A - C = 1) + \mathbb{P}(A - C = -1) = 2\mathbb{P}(A - C = 1)\\ &= 2(P(A = 2, C = 1) + \cdots + \mathbb{P}(A = 6, C = 5)) = \frac{10}{36}.\end{aligned}$$

Allo stesso modo si ha

$$\begin{aligned}\mathbb{P}(|X - Y| = 2) &= 2\mathbb{P}(A - C = 2) = 2(P(A = 3, C = 1) + \cdots + \mathbb{P}(A = 6, C = 4)) = \frac{8}{36}\\ \mathbb{P}(|X - Y| = 3) &= 2\mathbb{P}(A - C = 3) = 2(P(A = 4, C = 1) + \cdots + \mathbb{P}(A = 6, C = 3)) = \frac{6}{36}\\ \mathbb{P}(|X - Y| = 4) &= 2\mathbb{P}(A - C = 4) = 2(P(A = 5, C = 1) + \mathbb{P}(A = 6, C = 2)) = \frac{4}{36}\\ \mathbb{P}(|X - Y| = 5) &= 2\mathbb{P}(A - C = 5) = 2(P(A = 6, C = 1) = \frac{2}{36}\end{aligned}$$

Allora

$$\mathbb{E}\big[|X - Y|\big] = \sum_{k=1}^{5} k\mathbb{P}(|X - Y| = k) = \frac{1}{36}(10 + 16 + 18 + 16 + 10) = \frac{35}{18}.$$

Nome:______________________________________

6. **(6 Punti)** Lanciamo un dado equo n volte. Sia T_n la variabile aleatoria definita come segue: se c'e' almeno un 6 negli n lanci poniamo T_n uguale al numero del lancio in cui appare il 6 per la prima volta; se non ci sono 6 negli n lanci poniamo $T_n = n$. Calcolare:
   (a) la densità di probabilità di T_n, in funzione di n;
   (b) il limite per $n \to \infty$ del valore atteso di T_n;
   (c) la probabilità condizionata dell'evento $\{T_n < n\}$ sapendo che c'e' esattamente un 6 negli n lanci, in funzione di n.

Soluzione: (a). Se $k = 1, \ldots, n-1$, l'evento $T_n = k$ si ottiene se e solo se non ci sono 6 nei primi $k-1$ lanci e si ha 6 nel k-esimo lancio. Inoltre $T_n = n$ se e solo se tutti i primi $n-1$ lanci sono diversi da 6. Dunque

$$p(k) = \mathbb{P}(T_n = k) = \tfrac{1}{6}(\tfrac{5}{6})^{k-1}, \quad k = 1, \ldots, n-1; \qquad p(n) = \mathbb{P}(T_n = n) = (\tfrac{5}{6})^{n-1}.$$

In altre parole, T è una geometrica di parametro $1/6$ troncata al livello n.

(b). Per quanto visto sopra ci aspettiamo che il limite per $n \to \infty$ del valore atteso sia il valore atteso della geometrica di parametro $1/6$, ossia $\lim_{n\to\infty} \mathbb{E}[T_n] = 6$. Possiamo verificare questa affermazione come segue:

$$\mathbb{E}[T_n] = \sum_{k=1}^{n} kp(k) = \tfrac{1}{6}\sum_{k=1}^{n-1} k(\tfrac{5}{6})^{k-1} + n(\tfrac{5}{6})^{n-1}.$$

Per $n \to \infty$ si ha $n(\frac{5}{6})^{n-1} \to 0$ e $\frac{1}{6}\sum_{k=1}^{n-1} k(\frac{5}{6})^{k-1} \to \frac{1}{6}\sum_{k=1}^{\infty} k(\frac{5}{6})^{k-1}$. Allora

$$\lim_{n\to\infty} \mathbb{E}[T_n] = \tfrac{1}{6}\sum_{k=1}^{\infty} k(\tfrac{5}{6})^{k-1} = 6.$$

(c). Sia N il numero di 6 negli n lanci. Vogliamo calcolare

$$\mathbb{P}(T_n < n | N = 1) = \frac{\mathbb{P}(T_n < n, N = 1)}{\mathbb{P}(N = 1)}.$$

Notiamo che

$$\mathbb{P}(T_n < n, N = 1) = \sum_{k=1}^{n-1} \mathbb{P}(T_n = k, N = 1) = (n-1)\tfrac{1}{6}(\tfrac{5}{6})^{n-1},$$

dove usiamo il fatto che per ogni $k \in \{1, \ldots, n-1\}$ l'evento $\{T_n = k, N = 1\}$ equivale a avere un 6 al k-esimo lancio e tutti diversi da 6 negli $n-1$ lanci rimanenti, e dunque $\mathbb{P}(T_n = k, N = 1) = \frac{1}{6}(\frac{5}{6})^{n-1}$. D'altra parte, N è una binomiale di parametri n e $\frac{1}{6}$. Allora $\mathbb{P}(N = 1) = n\frac{1}{6}(\frac{5}{6})^{n-1}$. In conclusione,

$$\mathbb{P}(T_n < n | N = 1) = \frac{n-1}{n}.$$

Questo risultato si può indovinare facilmente senza calcoli osservando che condizionatamente all'evento $N = 1$, l'unico 6 negli n lanci deve essere uniformemente distribuito su $\{1, \ldots, n\}$.

Nome:_______________________________________