Dipartimento di Matematica, Roma Tre
Pietro Caputo

2019-20, II semestre
31 agosto, 2020

CP210 Introduzione alla Probabilità: Esame 3

Cognome	
Nome	
Matricola	
Firma	

Codice etico: Durante la prova d'esame la studentessa/lo studente:

1. non si avvale di alcun ausilio o supporto esterno, cartaceo o elettronico (es.: manuali, dispense, fogli propri, libri, pubblicazioni, telefoni cellulari, computer o altri dispositivi elettronici) che non siano penna e fogli per scrivere;
2. non copia né osserva le prove di altri candidati;
3. non contatta o tenta di contattare in alcun modo altre persone.

Firma

....

Modalità di esame:

1. Risposte implicite sotto forma di coefficienti binomiali, potenze, esponenziali ecc. sono benvenute. Mostrate in dettaglio il vostro lavoro.
2. Scrivete il vostro nome su ogni pagina. In caso di utilizzo di piu' pagine per un singolo esercizio indicare chiaramente l'ordine.
3. Il punteggio massimo per ogni esercizio è indicato nel testo. Notare che con cinque esercizi risolti correttamente si arriva 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 4 palline di cui alcune rosse e alcune nere ma non sappiamo esattamente in quale proporzione. Supponiamo di sapere che l'urna contiene 2 palline nere e 2 rosse con probabilità $\frac{1}{3}$, mentre contiene 3 palline nere e 1 rossa con probabilità $\frac{2}{3}$. Estraiamo una pallina a caso e la reinseriamo nell'urna insieme a una nuova pallina dello stesso colore di quella estratta. A questo punto l'urna contiene 5 palline. Calcolare:

   (a) la probabilità che l'urna contenga 4 palline nere e 1 pallina rossa;

   (b) la probabilità che l'urna contenga 3 palline nere e 2 palline rosse;

   (c) la probabilità condizionata che l'urna iniziale avesse 3 palline nere e 1 pallina rossa sapendo che l'urna attuale contiene 3 palline nere e 2 palline rosse.

Soluzione:

a). Sia A l'evento che l'urna iniziale contiene 3 palline nere e 1 rossa. Allora A^c indica l'evento che l'urna iniziale contiene 2 rosse e 2 nere. Inoltre $P(A) = 1 - P(A^c) = 2/3$. Sia B l'evento che dopo l'estrazione si hanno 4 nere e 1 rossa. Allora

$$P(B) = P(B \cap A) + P(B \cap A^c) = P(B \cap A),$$

dove usiamo il fatto che $B \cap A^c = \emptyset$ (non possiamo ottenere 4 nere e 1 rossa dopo l'estrazione se inizialmente si avevano 2 nere e 2 rosse). Inoltre $P(B|A) = 3/4$ poiché per ottenere 4 nere e 1 rossa si deve estrarre una nera dall'urna contenente 3 nere e 1 rossa. In conclusione

$$P(B) = P(B \cap A) = P(B|A)P(A) = \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}.$$

b). Sia C l'evento che dopo l'estrazione si hanno 3 nere e 2 rosse. Allora

$$P(C) = P(C \cap A) + P(C \cap A^c) = P(C|A)P(A) + P(C|A^c)P(A^c).$$

Ma $P(C|A) = 1/4$ e $P(C|A^c) = 1/2$. Dunque

$$P(C) = \frac{1}{4} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}.$$

c). Qui vogliamo calcolare

$$P(A|C) = P(C|A)\frac{P(A)}{P(C)} = \frac{1}{4}\frac{\frac{2}{3}}{\frac{1}{3}} = \frac{1}{2}.$$

Nome:____

2. **(6 punti)** Consideriamo due giochi. Nel primo ogni giorno indipendentemente, vinciamo 10 euro con probabilità 0.7 e perdiamo 10 euro con probabilità 0.3. Nel secondo, ogni giorno indipendentemente, vinciamo 5 euro con probabilità 0.9 e perdiamo 5 euro con probabilità 0.1. Supponendo di giocare per 5 giorni:

   (a) Calcolare il valore atteso del guadagno netto nei due giochi.

   (b) Calcolare la varianza del guadagno netto nei due giochi.

   (c) A quale gioco preferireste giocare ?

Soluzione: (a). Indichiamo con X il guadagno netto in un giorno nel primo gioco e con Y il guadagno netto in un giorno nel secondo gioco. Il valore atteso di X vale

$$E[X] = 10 \times 0.7 - 10 \times 0.3 = 4$$

Nel secondo caso si ha:

$$E[Y] = 5 \times 0.9 - 5 \times 0.1 = 4$$

Dunque per 5 giorni avremo in media un guadagno $5 \times 4 = 20$ euro in entrambi i giochi.

(b). La varianza del guadagno netto in un giorno nel primo caso vale

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = 100 \times 0.7 + 100 \times 0.3 - 16 = 84.$$

Nel secondo caso si ha:

$$\mathrm{Var}[Y] = E[Y^2] - E[Y]^2 = 25 \times 0.9 + 25 \times 0.1 - 16 = 9.$$

Poiché i guadagni in giorni differenti sono indipendenti la varianza per 5 giorni è 5 volte la varianza in un giorno e nel primo caso avremo una varianza pari a $5 \times 84 = 420$, mentre nel secondo caso vale $5 \times 9 = 54$. Dunque la varianza è molto maggiore nel primo caso.

(c). I calcoli precedenti mostrano che i guadagni attesi nei due giochi coincidono ma il primo gioco è molto piú rischioso nel senso che permette perdite e vincite molto maggiori rispetto al secondo gioco. La risposta dunque dipende dai gusti personali: chi è attratto dal rischio sarà tentato di scegliere il primo, chi è piú prudente sceglierà il secondo. Per essere piú quantitativi possiamo osservare che, se giochiamo per un numero n di giorni, per il teorema del limite centrale si avrà che il guadagno netto, per n grande, si comporta come una normale $Z_1 = N(4n, 84n)$ nel primo caso e come una normale $Z_2 = N(4n, 9n)$ nel secondo. Se $n = 5$, vediamo che nel primo caso l'evento $Z_1 < 0$ (perdita netta di denaro) ha probabilità almeno 0.16 (una deviazione standard sotto alla media) mentre nel secondo caso l'evento $Z_2 < 0$ ha una probab. quasi trascurabile (servono quasi tre deviazioni standard sotto alla media).

Nome:______________________________________

3. **(6 punti)** Sia S_n la posizione dopo n passi della passeggiata aleatoria con incrementi indipendenti ± 1 con probabilità $1/2, 1/2$, con punto di partenza $S_0 = 0$, e sia

$$Y_n = S_n^2 - n.$$

Calcolare i valori attesi:

(a) $E[Y_n]$

(b) $E[Y_{n+1} - Y_n]$

(c) $E[(Y_{n+1} - Y_n)^2]$

Soluzione:

a). Scriviamo

$$S_n = \sum_{i=1}^{n} Z_i$$

dove Z_i sono indipendenti tali che $Z_i = \pm 1$ con probabilità $1/2, 1/2$. La media di S_n è zero poiché gli incrementi Z_i hanno media $E[Z_i] = 0$. La varianza, per l'indipendenza, vale:

$$\mathrm{Var}(S_n) = n\mathrm{Var}(Z_1) = n,$$

dove usiamo il fatto che $\mathrm{Var}(Z_1) = \mathbb{E}[Z_1^2] = 1$. Dunque $E[S_n^2] = \mathrm{Var}[S_n] = n$, e

$$E[Y_n] = 0.$$

b). Si ha

$$E[Y_{n+1} - Y_n] = E[Y_{n+1}] - E[Y_n] = 0.$$

b). Notiamo che

$$Y_{n+1} - Y_n = (S_n + Z_{n+1})^2 - (n+1) - S_n^2 + n = 2S_n Z_{n+1} + Z_{n+1}^2 - 1 = 2S_n Z_{n+1}.$$

Inoltre per l'indipendenza si ha $E[S_n^2 Z_{n+1}^2] = E[S_n^2]E[Z_{n+1}^2] = E[S_n^2] = n$. Dunque

$$E[(Y_{n+1} - Y_n)^2] = 4E[S_n^2 Z_{n+1}^2] = 4n.$$

Nome:____

4. **(6 punti)** Siano X e Y due variabili aleatorie normali $N(\mu, \sigma^2)$ indipendenti, e sia

$$Z = X + Y.$$

Calcolare, al variare di $\mu \in \mathbb{R}$ e $\sigma^2 > 0$:

(a) Il valore atteso di Z^2;

(b) il valore atteso di e^Z;

(c) La covarianza tra Z e Y.

Soluzione:

a). Si ha
$$E[Z^2] = E[X^2 + 2XY + Y^2] = E[X^2] + 2E[XY] + E[Y^2].$$
Per l'indipendenza $E[XY] = E[X]E[Y] = \mu^2$ poiché $E[X] = E[Y] = \mu$. Inoltre

$$E[X^2] = E[Y^2] = \mu^2 + \sigma^2,$$

e dunque $E[Z^2] = 4\mu^2 + 2\sigma^2$. Alla stessa conclusione si può arrivare osservando che Z è una variabile $N(2\mu, 2\sigma^2)$.

b). Si ha
$$E[e^Z] = E[e^X e^Y] = E[e^X]E[e^Y] = e^{2\mu+\sigma^2},$$
dove abbiamo usato l'indipendenza e il fatto che

$$E[e^X] = E[e^Y] = \int_{-\infty}^{\infty} e^x f_{\mu,\sigma^2}(x)dx = e^{\mu+\sigma^2/2},$$

dove f_{μ,σ^2} è la densità della normale standard.
c). La covarianza tra Z e Y vale

$$\mathrm{Cov}(Z, Y) = E[ZY] - E[Z]E[Y] = 2\mu^2 + \sigma^2 - 2\mu^2 = \sigma^2,$$

dove usiamo il fatto che $E[ZY] = E[(X + Y)Y] = E[XY] + E[Y^2] = \mu^2 + \mu^2 + \sigma^2$.

Nome:______________________________________

5. **(6 punti)** Supponiamo che il numero di nascite in Italia ogni anno sia una variabile di Poisson di parametro 5×10^5. E supponiamo che le nascite corrispondenti a anni differenti siano indipendenti. Sia E l"evento che il numero di nascite in 2 anni è compreso tra 998000 e 1002000.

   (a) Usando la disuguaglianza di Chebyshev, fornire una stima dal basso della probabilità dell'evento E.

   (b) Calcolare un valore approssimato della probabilità dell'evento E.

Soluzione: (a). Per l'indipendenza si ha che il numero di nascite in 2 anni è una variabile di Poisson di parametro $\lambda = 10^6$. Allora il suo valor medio e la sua varianza sono pari a $\lambda = 10^6$. Se X indica il numero di nascite in 2 anni, l'evento in questione è

$$E = \{998000 \leqslant X \leqslant 1002000\} = \{|X - \lambda| > 2000\}^c.$$

Allora per Chebyshev si ha

$$P(E) \geqslant 1 - \frac{\mathrm{Var}(X)}{(2000)^2} = 1 - \frac{10^6}{4 \times 10^6} = \frac{3}{4}.$$

(b). X si può rappresentare come somma di λ variabili aleatorie di Poisson di parametro 1 indipendenti. Per il teorema del limite centrale $Z = (X - \lambda)/\sqrt{\lambda}$ è approssimata da una normale standard, e dunque

$$P(E) = P(|X - \lambda| \leqslant 2000) = P(|Z| \leqslant 2) \approx 2\Phi(2) - 1 \approx 0.95.$$

Nome:____

6. **(6 punti)** Sia $P = (X, Y)$ un punto scelto uniformemente a caso nel disco di raggio 1 centrato nell'origine del piano:

   (a) dire se X e Y sono indipendenti;

   (b) scrivere la densità di probabilità della variabile $X^2 + Y^2$;

   (c) calcolare il valore atteso della variabile $1/(X^2 + Y^2)$.

Soluzione:
(a) Le variabili non sono indiependenti poiché la densità congiunta di (X, Y) è positiva se e solo se il punto appartiene al disco di raggio 1 centrato nell'origine e dunque non è supportata da un rettangolo.

(b). Sia $D = \sqrt{X^2 + Y^2}$ la distanza dal centro del disco. Allora D^2 soddisfa $P(D^2 \leqslant t) = \pi t^2/\pi = t^2$ per ogni $t \in [0, 1]$, dove abbiamo usato il fatto che per l'uniformità la probabilità che P appartenga al disco di raggio t è il rapporto fra le aree tra il disco di raggio t e il disco di raggio 1, se $t \in [0, 1]$. Dunque D^2 ha densità di probabilità $f(t) = \frac{d}{dt} P(D^2 \leqslant t) = 2t$ per ogni $t \in [0, 1]$ e $f(t) = 0$ per $t \notin [0, 1]$.

(c). Dal punto precedente abbiamo

$$E[1/(X^2 + Y^2)] = E[1/D^2] = \int_0^1 t^{-1} f(t)dt = \int_0^1 2dt = 2.$$

Nome:____