Una raccolta di esercizi

Statistica A, Informatica. Anno 21-22

1 Esercizi di Probabilità

Esercizio 1.1. Qual è la probabilità che, estraendo di fila tre carte da un mazzo di poker, escano tre cuori?

Esercizio 1.2. In un paese nel quale vi sono 5 alberghi arrivano 3 turisti, questi si sono prenotati in modo del tutto casuale.
a) Qual è la probabilità che finiscano tutti in uno stesso albergo?
b) Qual è la probabilità che finiscano in 3 alberghi differenti?

Esercizio 1.3. Il re (di una monarchia che ammette solo regnanti maschi) non è figlio unico: qual è la probabilità che abbia una sorella?

Esercizio 1.4. Vi sono tre monete, una con doppia testa, una con doppia croce ed una testa e croce. Si sceglie a occhi chiusi una moneta e la si lancia, ed esce testa: qual è la probabilità che anche l'altra faccia sia testa?

Esercizio 1.5. Si lancia tre volte un dado: qual è la probabilità che la somma dei numeri usciti sia 15?

Esercizio 1.6. Si estraggono tre numeri da un'urna che contiene i numeri da 1 a 20: qual è la probabilità di aver estratto tra gli altri i numeri 1 e 2?

Esercizio 1.7. (Problema dei compleanni) Qual è la probabilità che in un insieme di n persone che ne siano (almeno) due che sono nate lo stesso giorno?

Osservazione: questo è un classico problema, e la soluzione (eguale a $1 - \dfrac{365!}{(365-n)!\,365^n}$) non può essere calcolata senza una calcolatrice. È interessante sapere che per $n = 23$ si ha ≈ 0.5 e per $n = 40$ si ha ≈ 0.9.

Esercizio 1.8. Supponiamo di avere due scatole, la prima con 5 palline rosse e 7 blu, la seconda con 8 palline rosse e 3 blu. Si sceglie a caso una scatola e si estraggono due palline che risultano essere entrambe rosse: qual è la probabilità che sia stata scelta la prima scatola?

Esercizio 1.9. Ripetiamo 4 volte un esperimento che ha probabilità di successo eguale a 2/3: calcolare la probabilità che
a) si abbia almeno un successo
b) si abbia esattamente un successo.

2 Esercizi su Variabili Aleatorie

Esercizio 2.1. Sia X una v.a. discreta che prende i valori 1, 2 e 3 con probabilità rispettive p_1, p_2 e p_3: sotto la condizione $\mathbf{E}[X] = 2$, per quali valori delle probabilità p_i la varianza è rispettivamente minima o massima?

Esercizio 2.2. Sia X una v.a. avente densità

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{altrove} \end{cases}$$

Qual è la costante c che rende la funzione sopra scritta una densità? Quanto vale $\mathbf{P}\{X > 1\}$? Quanto vale $\mathbf{E}[X]$?

Esercizio 2.3. In un concorso la prova scritta consiste in un test composto di 5 domande: per ciascuna di esse il test richiede di scegliere una tra 4 possibili risposte, delle quali una sola è corretta. Il test viene valutato attribuendo punteggio 1 ad ogni risposta esatta e $-\frac{1}{4}$ ad ogni risposta errata.

Si consideri un concorrente che risponda a caso ad ogni domanda: quale punteggio ottiene in media?

Esercizio 2.4. Sia X una variabile di Poisson di parametro λ: dire qual è il valore più probabile.

Più in generale dire per quali interi n, $\mathbf{P}\{X = n+1\} \geq \mathbf{P}\{X = n\}$.

Esercizio 2.5. Si lanciano due dadi equilibrati, e consideriamo i due eventi A "*la somma dei due numeri usciti è 7*", e B "*il numero comparso nel primo dado è 4*".

1) Provare che gli eventi A e B sono indipendenti.

2) Più in generale, indicando rispettivamente con le v.a. X ed Y il numero uscito nel primo e nel secondo dado, si considerino gli eventi $\{X + Y = n\}$ e $\{X = k\}$: esaminare per quali valori di n e di k questi due eventi rusultano indipendenti.

Esercizio 2.6. Vi sono due monete, indistinguibili al tatto, delle quali una è perfettamente equilibrata e l'altra è truccata: per quest'ultima è $\frac{2}{3}$ la probabilità che esca testa. Se ne sceglie una a caso e la si lancia 2 volte di seguito, e siano X_1 e X_2 i risultati dei due lanci (X_i vale 1 oppure 0 a seconda che all'i–mo lancio sia uscita testa oppure croce).

1) Le variabili X_1 e X_2 sono indipendenti?

2) Se è uscita testa in entrambi i lanci, qual è la probabilità che sia stata scelta la moneta truccata?

Esercizio 2.7. Sia (X, Y) una variabile doppia avente densità

$$f(x,y) = \begin{cases} x+y & 0<x<1,\, 0<y<1 \\ 0 & \text{altrove} \end{cases}$$

Calcolare le densità marginali delle componenti X e Y , calcolare valore atteso e varianza della variabile X .

Esercizio 2.8. Siano X e Y indipendenti, con densità rispettive f_1 e f_2 , supponiamo inoltre che X e Y prendano solo valori positivi: cercare una formula (analoga a quella della *convoluzione*) per la densità di XY.

Esercizio 2.9. Scrivere una formula esplicita per i quantili di una v.a. con densità uniforme sull'intervallo [a,b] e di una v.a. con densità esponenziale di parametro λ .

Esercizio 2.10. Sia X con densità gaussiana $N(3,4)$: calcolare $\mathbf{E}\big[X^4\big]$.

Esercizio 2.11. Sia X una v.a. con densità uniforme sull'intervallo $[-1,1]$: presa $Y = X^2$, calcolare densità (se esiste), media e varianza della variabile Y .

Esercizio 2.12. Sia T una variabile con densità esponenziale di parametro λ , e sia $S = \log T$.
a) Per quale t vale l'eguaglianza $\mathbf{P}\{T > t\} = 0.1$?
b) Esaminare se la variabile S ha densità e in caso affermativo calcolarla, ed esaminare se S^2 ha densità.

Esercizio 2.13. Ricordiamo questa *formula generale*: data X v.a. con densità f diversa da 0 su un intervallo A , e presa $Y = h(X)$ dove h è biunivoca, derivabile con inversa derivabile da A su un intervallo B , allora Y ha densità g che è nulla fuori di B e su B vale $g(y) = f(h^{-1}(y)).\left|\frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y}\right|$.

Nell'esercizio precedente, si può applicare questa formula al caso di S ed al caso di S^2 ?

3 Esercizi su momenti, funzioni generatrici e teoremi limite

Esercizio 3.1. Siano X, Y indipendenti, con X gaussiana $N(0,2)$ e Y discreta che prende i valori 1 e -1 con probabilità 1/2.
a) Calcolare $\mathbf{P}\{\,|X| > 1\,\}$.
b) Calcolare la varianza di Y.
c) Calcolare $\mathbf{E}\Big[\big(\frac{X}{\sqrt{2}} - Y\big)^2\Big\}$.

Esercizio 3.2. Siano X_1, X_2 indipendenti, con distribuzione di Poisson di parametro λ,
a) Determinare per quale valore di λ si ha $\mathbf{E}\big[(X_1 - X_2)^2\big] = 1$,
b) Utilizzando il valore di λ determinato precedentemente, calcolare $\mathbf{P}\big\{X_1 + X_2 \geq 2\big\}$.

Esercizio 3.3. Siano X e Y indipendenti, con X binomiale B(2,1/5) ed Y con densità uniforme su $[0,2]$: posto $Z = \dfrac{X(Y-1)^2}{X+1}$, calcolare $\mathbf{E}[Z]$.

Esercizio 3.4. Si chiama variabile *lognormale* di parametri m e σ^2 una variabile Y della forma $Y = \mathrm{e}^X$, dove X è gaussiana $N(m, \sigma^2)$.
a) Calcolare la funzione di ripartizione (c.d.f.) e la densità della variabile Y.
b) Calcolare il valore atteso $\mathbf{E}[Y]$ se $m = 0$ e $\sigma^2 = 3$.

Esercizio 3.5. Calcolare i momenti primo e secondo di una variabile Binomiale di parametri n e p utilizzando la *funzione generatrice dei momenti.*

Esercizio 3.6. Gli aerei della compagnia XYZ hanno 180 posti, ma la compagnia sa che in media quelli che si sono prenotati si presentano alla partenza con probabilità 0.9 : per questo la compagnia mette in vendita 195 biglietti (*overbooking*). Qual è la probabilità che qualche cliente rimanga a terra (e di conseguenza la compagnia sia costretta a un risarcimento)?

Esercizio 3.7. Siano $X_1, X_2, \ldots, X_{50}$ indipendenti con densità esponenziale di parametro 2.
a) Stimare la probabilità $\mathbf{P}\{X_1 + \cdots + X_{50} \geq 32\}$
b) Calcolare $\mathbf{E}\big[X_1\,\mathrm{e}^{X_2}\,\big]$.

Esercizio 3.8. Siano $(X_1, X_2, \ldots)$ di v.a. indipendenti equidistribuite con densità uniforme su $[0,1]$: calcolare la densità della variabile $M_n = \max(X_1, \ldots, X_n)$. Esaminare poi se la successione $(M_n)_{n\geq 1}$ converge (in probabilità o in distribuzione) ed eventualmente a quale limite.

Rispondere alle stesse domande considerando questa volra la v.a. $V_n = \min(X_1, \ldots, X_n)$.

Rispondere infine alle domande precedenti supponendo invece che le variabili siano esponenziali di parametro λ.

Esercizio 3.9. Siano $X_1, \ldots, X_{60}$ i.i.d. con densità uniforme su $[0, 2]$ e sia $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$ la media empirica.
a) Stimare la probabilità $\mathbf{P}\{\frac{5}{6} \le \overline{X} \le \frac{7}{6}\}$.
b) Stimare il più piccolo x per il quale si abbia $\mathbf{P}\{\overline{X} > x\} \le 0.05$.

4 Esercizi su campioni statistici e stima parametrica

Esercizio 4.1. Consideriamo, per $\theta \in \mathbb{R}$, la funzione

$$f_\theta(x) = \begin{cases} C_\theta\,(2x + \theta) & 0 < x < 1 \\ 0 & \text{altrove} \end{cases}$$

a) Per quali valori di θ e C_θ la funzione sopra scritta è una densità di probabilità?
b) Scrivere esplicitamente la c.d.f., media e varianza di una v.a. che ha la densità sopra scritta.

Esercizio 4.2. Scrivere esplicitamente la c.d.f. e calcolare il momento primo di una variabile che abbia densità

$$f_\theta(x) = \begin{cases} \mathrm{e}^{-(x-\theta)} & x \ge \theta \\ 0 & x < \theta \end{cases}$$

al variare di θ in $\mathbb{R}$.

Considerato poi un campione $X_1, \ldots, X_n$ avente la densità sopra scritta, indagare sulla stima di θ col metodo della massima verosimiglianza e col metodo dei momenti.

Esercizio 4.3. Preso $\theta > 0$, si consideri la funzione

$$f_\theta(x) = \begin{cases} c_\theta\, x^\theta & 0 < x < \theta \\ 0 & \text{altrimenti} \end{cases}$$

a) Dopo aver calcolato la costante c_θ che rende la funzione sopra scritta una densità, calcolare, per una v.a. X che ha quella densità, $\mathbf{P}\{X > \theta/2\}$.

b) Se $\theta = 2$, e $X_1, \ldots, X_{100}$ sono indipendenti aventi quella densità, stimare $\mathbf{P}\big\{X_1 + \cdots + X_{100} \geq 157\big\}$.

c) Determinare una stima di θ col *metodo dei momenti*.

Esercizio 4.4. Consideriamo, per $\theta > 0$, la funzione

$$f_\theta(x) = C_\theta \begin{cases} \mathrm{e}^{x} & x < -\theta \\ 0 & -\theta \leq x \leq \theta \\ \mathrm{e}^{-x} & x > \theta \end{cases}$$

a) Determinare la costante che rende la funzione sopra scritta una densità di probabilità e tracciare un grafico approssimativo di $f_\theta(.)$.
b) Tracciare un grafico approssimativo della funzione di ripartizione e scriverla esplicitamente.
c) Considerato un campione con la densità sopra scritta, indagare sulla stima del parametro θ.

Esercizio 4.5. Preso $\theta > 0$, sia $f_\theta(.)$ la funzione

$$f_\theta(x) = \begin{cases} \frac{c(\theta)}{x^2} & \theta < x < 2\,\theta \\ 0 & \text{altrimenti} \end{cases}$$

a) Dopo aver determinato il valore di $c(\theta)$ che rende la funzione sopra scritta una densità di probabilità, calcolare media e varianza di una variabile che ha quella densità.

b) Se si osservano n valori $(x_1, \ldots, x_n)$ di un campione estratto da una popolazione con la densità sopra scritta, esaminare la stima di θ col metodo dei momenti.

Esercizio 4.6. Consideriamo, per $\theta > -1$, la funzione

$$F_\theta(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^2+\theta x}{1+\theta} & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

a) Determinare per quali valori di θ questa è la funzione di ripartizione di una variabile con densità.
b) Sono stati osservati i seguenti dati numerici: 0.35, 0.40, 0.31, 0.85, 0.59, 0.60. Immaginando che questi dati siano l'osservazione di un campione statistico con la c.d.f. sopra scritta, dare una stima di θ col metodo dei momenti.

5 Esercizi su Intervalli di fiducia e verifica delle ipotesi

Esercizio 5.1. I risultati dei compiti di Statistica hanno media 21 e varianza 7: quanti studenti almeno devono fare il prossimo compito affinché, con probabilità (approssimativa) di almeno 0.99, la media dei loro voti sia compresa tra 19 e 23?

Esercizio 5.2. Una barra di metallo di produzione industriale deve essere lunga 40 cm, tuttavia sono state effettuate 400 misurazioni e si è trovato un valore medio di 39.6 con deviazione standard di 4: ci sono seri motivi per ritenere che la produzione sia di cattiva qualità (cioè che non si possa assumere che la lunghezza media sia effettivamente 40 cm)?

È possibile risolvere questo problema senza l'uso di strumenti di calcolo ma con la sola tavola della variabile $N(0,1)$?

Esercizio 5.3. Si osserva un campione di 10000 neonati e si trova che di essi 5106 sono femmine: testare l'ipotesi che il sesso dei neonati sia equamente distribuito, trovare il *p-value* e commentare il risultato.

Esercizio 5.4. In un campione di 1000 assi di legno, se ne trovano 350 di lunghezza superiore a 100 cm : determinare un intervallo di fiducia bilatero al 99% per la probabilità che un asse di legno sia più lungo di 100 cm.

Esercizio 5.5. Il peso di un oggetto misurato da una bilancia è il peso vero al quale si sovrappone un errore gaussiano (con media 0 e varianza sconosciuta). Si pesa un oggetto e si raccolgono i seguenti dati:

$3.142\,g \qquad 3.163\,g \qquad 3.155\,g \qquad 3.150\,g \qquad 3.141\,g$

Fermo restando che 5 dati sono un campione troppo piccolo per avere risposte statistiche veramente affidabili, testare l'ipotesi che il peso vero non superi 3.142 , calcolare il *p-value* e commentare il risultato.

Dove i calcoli non risultano praticabili, indicare con quali comandi di **R** si può fornire la soluzione.

Esercizio 5.6. Nel mese di gennaio è stata misurata per ogni giorno la temperatura più bassa, ottenendo una media di -12 (gradi centigradi) con deviazione standard 7.2.

a) Qual è una ragionevole stima della minima temperatura più bassa rilevabile, con un'incertezza sulla media con fiducia al 98%?

b) Determinare un intervallo di fiducia sinistro al 99% per la varianza della temperatura più bassa.

c) Consideriamo l'ipotesi che in realtà la temperatura più bassa sia superiore a -10: scegliere il test opportuno, scrivere il *p-value* e la forma della regione critica al livello 0.04.

d) Se nel successivo mese di febbraio è stata rilevata una media di -14.3 con deviazione standard 6.4, si può accettare l'ipotesi che in realtà le temperature minime abbiano la stessa media (dando per scontato per semplicità che le varianze siano eguali)? Quale test è opportuno pianificare, quale risulta la formula per il *p-value*?

Esercizio 5.7. La temperatura ottimale di estrusione di una fibra artificiale è tra i 280 ed i 290 gradi centigradi: una serie di 64 controlli ha permesso di ricavare una temperatura media (empirica) di 277.6 con deviazione standard 10.3.

a) Dare un valore massimo per la stima della temperatura di estrusione a un livello di fiducia del 98%.

b) Qual è il test più opportuno per valutare se la temperatura ottimale è inferiore a 275? Scrivere la formula per il *p-value* e la forma della regione critica (o di rifiuto) con fiducia al 95% (cioè al livello 0.05).

c) Per valutare se possiamo affermare con ragionevole sicurezza che la deviazione standard non supera 15, considerare il test dell'ipotesi $\mathcal{H}_0\big)\ \sigma \geq 15$: calcolare il *p-value* e scrivere la formula della *potenza* del test al livello 0.025.

Esercizio 5.8. Le temperature massime giornaliere, misurate a Firenze nel luglio 2013, hanno fornito una media di 32.3 con varianzacampionaria 11.9 ; nell'anno successivo, sempre nel mese di luglio, la media è stata di 30.5 con varianza campionaria 13.1.

Un metereologo sostiene però che si tratta di una semplice fluttuazione e che il luglio 2013 non è stato veramente più caldo ed a tale scopo imposta il test dell'ipotesi $\mathcal{H}_0\big)\ m_1 = m_2$ contro l'alternativa $\mathcal{H}_1\big)\ m_1 > m_2$ (dove m_1 ed m_2 sono le medie delle temperature massime rispettivamente nei due anni).

Sviluppare questo test supponendo per semplicità che le temperature massime possano essere rappresentate da variabili gaussiane e che le varianze nei due mesi siano eguali, calcolare con buona approssimazione il *p-value* e trarne le conclusioni.