Probabilità e Statistica (Informatica) 2024/25, Foglio II

27 marzo 2025

Sia $(x_i)_{i\in\{1,\dots,n\}}$ un campione di numerosità n a valori in un insieme non-vuoto $\mathcal{X}$. Definiamo $p\colon \mathcal{X} \to [0,1]$ tramite

$$p(x) \doteq \frac{|\{i \in \{1,\dots,n\} : x_i = x\}|}{n}, \quad x \in \mathcal{X}.$$

Allora p è una densità discreta su $\mathcal{X}$, la *densità campionaria* di $(x_i)_{i\in\{1,\dots,n\}}$. Nota che $p(x)$ dà la frequenza relativa del valore $x \in \mathcal{X}$.

Esercizio 1. Si crei un campione di numerosità 60 registrando i numeri segnati da un dado a sei facce regolare ed equilibrato lanciato sessanta volte. Si calcolino e si rappresentino graficamente le densità campionarie dei sotto-campioni (partendo dal primo dato) di numerosità rispettivamente 6, 10, 30, e 60. Si confrontino i risultati con la densità discreta teorica.

Esercizio 2. Si considerino i due campioni indicati con Tmin e Ptot nel file `Meteo_Chioggia60.ods`. Per ciascuno dei due campioni e $\alpha \in \{2, 3, 5\}$, si calcoli la proporzione dei dati che si trovano nell'intervallo aperto $(\bar{x} - \alpha s, \bar{x} + \alpha s)$, dove $\bar{x}$ indica la media campionaria e s la deviazione standard campionaria del rispettivo campione. Si confrontino i risultati con le stime che si ottengono dalla disuguaglianza di Chebyshev.

Esercizio 3. Per $n \in \mathbb{N}$, si consideri il campione (artificiale) $(x_i^{(n)})_{i\in\{1,\dots,2n\}}$ di numerosità $2n$ dato da $x_i^{(n)} \doteq (-1)^i$. Per ogni $n \in \mathbb{N}$, ogni $\alpha > 0$, si calcoli la proporzione dei dati che si trovano nell'intervallo aperto $(\bar{x}^{(n)} - \alpha s_n, \bar{x}^{(n)} + \alpha s_n)$, ove $\bar{x}^{(n)}$ indica la media campionaria e s_n la deviazione standard campionaria del campione $(x_i^{(n)})_{i\in\{1,\dots,2n\}}$. Si confrontino i risultati con le stime che si ottengono dalla disuguaglianza di Chebyshev.

Esercizio 4. Da un'urna che contiene 5 palline rosse, 6 palline bianche e 8 palline verdi vengono estratte 3 palline *senza reinserimento*. Si calcoli la probabilità che le palline estratte

(a) abbiano tutte lo stesso colore;

(b) abbiano tre colori diversi.

Si risponda alle stesse domande nel caso di estrazioni *con reinserimento*.

Esercizio 5. Per l'esperimento del lancio di due monete non necessariamente equilibrate, scegliamo

$$\Omega \doteq \{T, C\}^2 = \{(C, C), (C, T), (T, C), (T, T)\}$$

come spazio campionario, e la σ-algebra totale come sistema degli eventi. Così l'evento $\{(C, T)\}$, ad esempio, rappresenta l'affermazione "è uscita croce per la prima moneta e testa per la seconda".

Si trovino tutte le densità discrete su Ω tali che, rispetto alla misura di probabilità indotta, la prima moneta risulta equilibrata, mentre la seconda segna croce con probabilità due terzi. Rispetto a quale di quelle densità gli eventi corrispondenti a "è uscita croce per la prima moneta" ed "è uscita croce per la seconda moneta" sono indipendenti?

Esercizio 6. Lanciamo una moneta non necessariamente equilibrata dieci volte. Denotiamo con $p \in [0, 1]$ la probabilità di ottenere testa in un singolo lancio. Per quale valore di p diventa massima la probabilità di ottenere esattamente quattro volte testa?

Più in generale, lanciamo una moneta non necessariamente equilibrata n volte. Sia $k \in \{0, \dots, n\}$. Se $p \in [0, 1]$ denota la probabilità di ottenere testa in un singolo lancio, per quale valore di p diventa massima la probabilità di ottenere esattamente k volte testa?

Esercizio 7. Per l'esempio del test clinico visto a lezione, si trovi uno spazio di probabilità discreto che descriva il corrispondente esperimento aleatorio e si definiscano gli eventi d'interesse come sottoinsiemi dello spazio campionario scelto.

Esercizio 8. Sia dia un esempio di uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ ed eventi $A_1, A_2, B \in \mathcal{F}$ tali che $\mathbf{P}(B) > 0$, $\mathbf{P}(A_1|B) > \mathbf{P}(A_1)$ e $\mathbf{P}(A_2|B) < \mathbf{P}(A_2)$.

Esercizio 9. Si verifichi se esiste o meno uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ con eventi $A, B, C \in \mathcal{F}$ tali che

$$\mathbf{P}(A) = \frac{5}{12}, \quad \mathbf{P}(B) = \frac{1}{3}, \quad \mathbf{P}(C) = \frac{1}{4},$$
$$\mathbf{P}(A|B) = \frac{1}{3}, \quad \mathbf{P}(A|C) = \frac{1}{3}, \quad \mathbf{P}(B|C) = \frac{1}{2}, \quad \mathbf{P}(A \cup B \cup C) = \frac{2}{3}.$$

Esercizio 10. Sia $\Omega \doteq \{0,1\}^3$, e sia $\mathbf{P}$ la misura uniforme su Ω. Poniamo

$$A \doteq \{\omega \in \Omega : \omega_3 = 0\}, \qquad B \doteq \{\omega \in \Omega : \omega_1 = 1\},$$
$$C \doteq \{(0,0,0),(0,1,0),(1,0,0),(1,0,1)\}.$$

Si calcolino le probabilità delle varie intersezioni e si determinino le relazioni di indipendenza tra A, B, C. Si dia poi un'interpretazione di questi eventi in termini dell'esperimento aleatorio del lancio di tre monete.

Esercizio 11 (Ross, Problema 3.38)**.** "Due palline vengono tinte con vernice nera o dorata, ciascuna con probabilità $1/2$ e indipendentemente l'una dall'altra. Esse vengono poi inserite in un'urna.

(a) Supponi di sapere per certo che la vernice dorata sia stata usata (e quindi vi è almeno una pallina di questo colore). Calcola la probabilità condizionata che entrambe le palline siano dorate.

(b) Supponi adesso che l'urna venga scossa violentemente, e ne esca una pallina dorata. Qual è la probabilità condizionata che anche l'altra pallina lo sia?

(c) Spiega come mai nei due punti precedenti hai ottenuto lo stesso numero / un numero diverso."

Infine, si trovi una spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ che rappresenti l'esperimento aleatorio descritto sopra, e si definiscano gli eventi d'interesse come sottoinsiemi di Ω.

Esercizio 12. Vogliamo stimare il numero di pesci (adulti) che vivono in un laghetto. Indichiamo con X quel numero ignoto. Immaginiamo di catturare r pesci, di marcarli, e di liberarli nel laghetto. Aspettiamo che i pesci marchiati si siano mescolati tra gli altri. Poi catturiamo n pesci e contiamo il numero di esemplari marchiati. Denotiamo con $A_{r,n,i}$ l'evento di aver trovato esattamente i pesci marchiati tra gli n presi alla seconda cattura. Per $r, n, N \in \mathbb{N}$ con $r \vee n \leq N$, $i \in \{0, \ldots, r \wedge n\}$, si calcoli la probabilità condizionale

$$c_{r,n,i}(N) \doteq \mathbf{P}\left(A_{r,n,i} \mid X = N\right).$$

Nel caso $r = 10$, $n = 20$, $i = 7$, si trovi numericamente il numero N che massimizzi $c_{r,n,i}(N) = c_{10,20,7}(N)$. Si dia infine una formula per stimare N in termini di r, n, i.

Esercizio 13 (Caravenna - Dai Pra, Esercizio 1.41)**.** "Si consideri il seguente modello di distribuzione dei figli nei nuclei familiari. La probabilità che un nucleo familiare scelto a caso abbia (esattamente) n figli, con $n \geq 0$, vale $e^{-\lambda}\frac{\lambda^n}{n!}$, dove λ è un parametro fissato, e ciascun figlio è maschio con probabilità $1/2$, indipendentemente da tutti gli altri. Consideriamo l'evento

$$A_k \doteq \text{“il nucleo familiare scelto (a caso) ha esattamente } k \text{ figli maschi”},$$

per $k \in \mathbb{N}_0$. Si mostri che la probabilità di A_k è uguale a $e^{-\lambda/2}\frac{(\lambda/2)^k}{k!}$."

Esercizio 14. Siano $N \in \mathbb{N}$, $q \in (0,1)$. Sia $p_{Bin(N,q)}$ la densità discreta della distribuzione binomiale di parametri N e q:

$$p_{Bin(N,q)}(k) \doteq \begin{cases} \binom{N}{k} \cdot q^k \cdot (1-q)^{N-k} & \text{se } k \in \{0,\dots,N\}, \\ 0 & \text{altrimenti.} \end{cases}$$

Si mostri che, per $k \in \{0,\dots,N-1\}$,

$$p_{Bin(N,q)}(k+1) = \frac{q}{1-q} \cdot \frac{N-k}{k+1} \cdot p_{Bin(N,q)}(k).$$

Si verifichi inoltre che, per $k \in \{0,\dots,N\}$,

$$p_{Bin(N,q)}(k) = p_{Bin(N,1-q)}(N-k).$$

Esercizio 15 ($*$). In un ballottaggio tra due candidati, A e B, votano $N+M$ persone, con $N \in \mathbb{N}$, $M \in \{0,\dots,N\}$: N elettori sono completamente indecisi e votano a caso, senza preferenza tra A e B, mentre il gruppo di M persone sostiene il candidato A. Vogliamo trovare la probabilità che vinca A.

Possiamo descrivere il comportamento elettorale delle N persone indecise tramite una variabile aleatoria S_N con distribuzione binomiale di parametri N e $1/2$, definita su un opportuno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$; S_N rappresenta il numero di voti che il candidato A riceve dal gruppo delle persone indecise. La probabilità che vinca A è allora data da

$$\mathbf{P}\left(S_N + M > \frac{N+M}{2}\right) = \mathbf{P}\left(S_N > \frac{N-M}{2}\right) = \sum_{k=\lfloor \frac{N-M}{2} \rfloor + 1}^{N} p_{Bin(N,1/2)}(k).$$

Si scriva un programma che calcoli numericamente questa probabilità in funzione di N, M come sopra, in grado di trattare il caso $N+M = 10^6$, $M \leq 5000$.

Per la consegna servono:

- lo svolgimento dell'Esercizio 14 e una giustificazione matematica della procedura utilizzata per il calcolo della probabilità di vittoria elettorale richiesta;
- lo pseudo-codice del programma e il codice commentato in un linguaggio standard come `C++` o Python (il codice anche in un file separato);
- un grafico in formato `pdf` che riporti la probabilità di vittoria elettorale in funzione di M quando $N+M = 10^6$ e M varia da 0 a 5000 (in passi da dieci).

Contatto: Markus Fischer (fischer@math.unipd.it)