Statistica - CPS
Corso di Laurea in Informatica
Compito del 14-12-2022

Esercizio 1. (10 punti) I 42 studenti di un corso (dei quali 13 sono ragazze) sono divisi per le esercitazioni in due gruppi di eguale numero per sorteggio. Si indichino rispettivamente con X e Y le v.a. il cui valore è il numero di ragazze presenti nel primo e nel secondo gruppo.

(i) Calcolare la funzione di probabilità (o densità discreta) della v.a. X.

(ii) Provare che le v.a. X e Y sono equidistribuite, ma non sono indipendenti.

(iii) È possibile determinare i valori attesi $E[X]$ e $E[Y]$ senza usare le leggi di X e Y?

Esercizio 2. (10 punti) Sia X una v.a. con densità

$$f(x) = \begin{cases} c\,x^2, & \text{se } -1 < x < 1 \\ 0, & \text{altrimenti} \end{cases}$$

dove c è una opportuna costante.

(i) Dopo aver determinato la costante c che rende la funzione sopra scritta una densità di probabilità, calcolare $E[X]$ e $\mathrm{Var}(X)$.

(ii) Sia data ora una successione $(X_n)_{n\geq 1}$ di v.a. indipendenti, tutte con la stessa densità di X. Calcolare

$$\lim_{n\to\infty} \mathbb{P}\big\{X_1 + \ldots + X_n \leq \sqrt{n}\big\}.$$

Esercizio 3. (10 punti) Il responsabile di una ditta petrolifera afferma che il contenuto medio di zolfo per litro, nella benzina prodotta da quella ditta, non supera 0.15 mg/l; tuttavia l'unione consumatori contesta questa affermazione perché sono stati prelevati 41 campioni che hanno dato valori $x_1, \ldots, x_{41}$ dai quali si ottiene

$$\overline{x} = \frac{x_1 + \cdots + x_{41}}{41} = 0.2\,.$$

Il responsabile afferma che questo dato non è significativo poiché la variabilità era alta: si è infatti ottenuto il valore $\sum_{i\leq 41}(x_i - \overline{x})^2 = 1$. (Si interpretino i valori delle misurazioni come variabili gaussiane con media e varianza ignote.)

(i) Si può accettare l'affermazione del responsabile della ditta (cioè l'ipotesi che il contenuto medio di zolfo non superi 0.15 mg/l)? Impostare un opportuno test e calcolare il relativo p-value.

(ii) Scrivere l'intervallo di fiducia unilatero destro (della forma $[a, +\infty)$) per il contenuto medio di zolfo al livello del 95%.

Svolgimento

Esercizio 1

(i) La variabile X è ovviamente discreta: i suoi valori sono gli interi $0,\ 1, \dots,\ 13$ e per k compreso tra questi valori si ha

$$\mathbb{P}\big\{X = k\big\} = \frac{\binom{13}{k}\binom{29}{21-k}}{\binom{42}{21}}$$

(infatti è come se scegliessi le k ragazze dal sottinsieme delle 13 ragazze e i $(21-k)$ ragazzi dal restante sottinsieme dei 29 ragazzi).

(ii) Il fatto che le variabili X e Y siano equidistribuite è del tutto intuitivo per ragioni di simmetria; possiamo tuttavia darne una dimostrazione formale.

Notiamo preliminarmente che $X+Y = 13$ e quindi $\mathbb{P}\{Y = k\} = \mathbb{P}\{X = 13-k\}$: di conseguenza affermare che X e Y sono equidistribuite equivale ad affermare $\mathbb{P}\{X = k\} = \mathbb{P}\{X = 13-k\}$. La verifica di questa eguaglianza è facile e segue dalle due eguaglianze (entrambe immediate)

$$\binom{13}{k} = \binom{13}{13-k} \qquad \text{e} \qquad \binom{29}{21-k} = \binom{29}{21-13+k}$$

È evidente che X e Y non possono essere indipendenti poiché $Y = 13 - X$: ad esempio si ha $\mathbb{P}\{X = 5\} \neq 0$ e $\mathbb{P}\{Y = 5\} \neq 0$ mentre $\mathbb{P}\{X = 5\,,\, Y = 5\} = 0$ e quindi non può essere verificata l'eguaglianza

$$\mathbb{P}\{X = 5\,,\, Y = 5\} = \mathbb{P}\{X = 5\}\,.\,\mathbb{P}\{Y = 5\}$$

(iii) Poiché X ed Y sono equidistribuite (e quindi $E[X] = E[Y]$) e $X+Y = 13$ si ha $E[X+Y] = 2\,E[X] = 13$ e quindi $E[X] = E[Y] = 6.5$.

Esercizio 2

(i) Cominciamo a calcolare $\int_{-1}^{1} x^2\,dx = \frac{x^3}{3}\Big|_{-1}^{1} = \frac{2}{3}$ e di conseguenza $c = 3/2$.
Si ha poi $E[X] = 3/2\,.\,\int_{-1}^{1} x^3\,dx = 0$ e
$Var(X) = E[X^2] = 3/2\,.\,\int_{-1}^{1} x^4\,dx = \frac{3}{2}\,.\,\frac{2}{5} = 3/5$.

(ii) Questa è una evidente applicazione del Teorema Limite Centrale, in base al quale

$$\lim_{n\to\infty} \mathbb{P}\Big\{a \leq \frac{X_1 + \cdots + X_n}{\sqrt{3/5}\,\sqrt{n}} \leq b\Big\} = \Phi(b) - \Phi(a)$$

Di conseguenza

$$\lim_{n\to\infty} \mathbb{P}\Big\{X_1 + \cdots + X_n \leq \sqrt{n}\Big\} = \lim_{n\to\infty} \mathbb{P}\Big\{\frac{X_1 + \cdots + X_n}{\sqrt{3/5}\,\sqrt{n}} \leq \sqrt{\frac{5}{3}}\Big\} \approx \Phi(1.29) \approx 0.9$$

Esercizio 3

(i) Facciamo dei conti preliminari: $\overline{X}(\omega) = 0.2$ e $S(\omega) = \sqrt{\frac{\sum_{i\leq 41}(x_i - \overline{x})^2}{40}} = \sqrt{1/40} = 0.158$. Inoltre (se $m = 0.15$), la variabile $\sqrt{40}\,\dfrac{\overline{X} - 0.15}{S}$ ha densità di Student $T(40)$.

Considerando il test dell'ipotesi $\mathcal{H}_0)\ m \leq 0.15$ contro $\mathcal{H}_1)\ m > 0.15$, la formula per il *p-value* è

$$\mathbb{P}_{0.15}\Big\{\sqrt{41}\,\frac{\overline{X} - 0.15}{S} > \sqrt{41}\,\frac{\overline{x} - 0.15}{s}\Big\} = 1 - F_{40}\Big(\sqrt{41}\frac{0.05}{0.158}\Big) = 1 - F_{40}\big(2.026\big)$$

Se si approssima la c.d.f. di Student F_{40} con la c.d.f. gaussiana standard Φ, il conto del *p-value* risulta $1 - \Phi(2.026) = 0.022$.

Tuttavia l'approssimazione gaussiana non è molto precisa perché il numero 40 è troppo basso; guardando con attenzione la tavola dei quantili della variabile di Student $T(40)$, si trova che il quantile $\tau_{(0.975\,,\,40)}$ è 2.0211 (molto vicino a 2.026) e si arriva ad una valutazione del *p-value* (approssimata ma molto più precisa) di 0.025. Si tratta in entrambi i casi di un valore molto basso, e quindi l'affermazione del responsabile della ditta è *assolutamente da scartare.*

(ii) Qui si tratta semplicemente di applicare la formula dell'intervallo di fiducia unilatero destro, tenendo conto del fatto che $\overline{X}(\omega) = 0.2$, $S(\omega) = 0.158$ e $\tau_{(0.95,40)} = 1.683$.

L'intervallo, la cui formula è $\Big[\overline{X}(\omega) - \frac{S(\omega)}{\sqrt{41}}\,\tau_{(.95,40)}\,,\ +\infty\Big)$, diventa $\big[0.2 - \frac{0.158}{\sqrt{41}}\,1.683\,,\ +\infty\big) = \big[0.158\,,\ +\infty\big)$.