Politecnico di Milano - Scuola di Ingegneria Industriale e dell'Informazione

PRIMA PROVA IN ITINERE DI STATISTICA PER INGEGNERIA ENERGETICA
27 novembre 2014

©I diritti d'autore sono riservati. Ogni sfruttamento commerciale non autorizzato sarà perseguito.

Cognome, Nome e Numero di matricola:

Problema 1. Nel 1929 l'astronomo americano Edwin Hubble trova la famosa relazione tra la distanza d di una galassia e la sua velocità v di allontanamento dalla Via Lattea: $v = h \cdot d$. La cosiddetta legge di Hubble contiene una costante che chiamiamo h. Di seguito si riportano alcune misurazioni di h, in km/s Mpc, effettuate a partire dal 1929 fino ai giorni nostri, suddivise per classi.

Classi	h
$[0-10)$	2
$[10-20)$	0
$[20-30)$	2
$[30-40)$	2
$[40-50)$	6
$[50-60)$	7
$[60-70)$	5
$[70-80)$	6
$[80-90)$	3
$[90-100)$	2
$[100-110)$	1

(a) Si calcoli la tabella di distribuzione delle frequenze relative, delle densità e delle frequenze cumulate.

(b) Si costruisca l'istogramma dei dati con le classi assegnate.

(c) L'istogramma al punto (b) risulta soddisfacente? In caso negativo, modificarlo opportunamente.

(d) Valutare la media dei dati raccolti con la miglior approssimazione possibile.

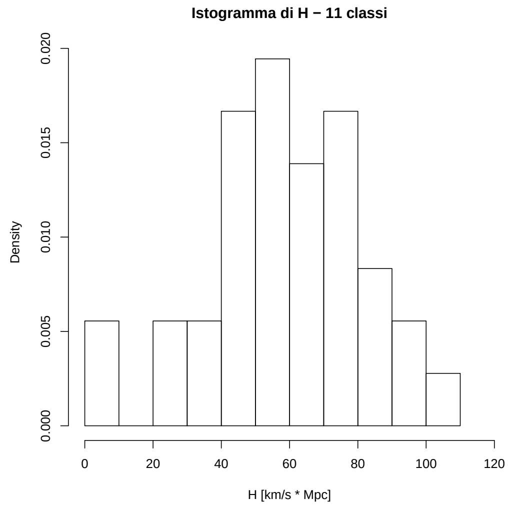
(e) Valutare il 75-esimo percentile dei dati raccolti con la miglior approssimazione possibile.

Risultati.

(a)

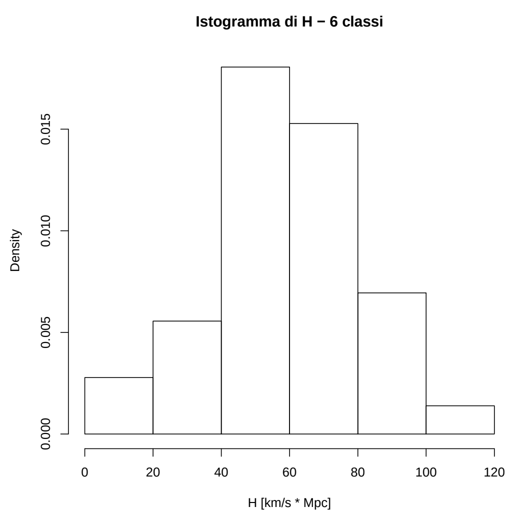
Classi	F. Ass.	F. Rel.	Densità	F. Cum.
$[0,10)$	2	0.0556	0.0056	0.0556
$[10-20)$	0	0	0	0.0556
$[20,30)$	2	0.0556	0.0056	0.1111
$[30,40)$	2	0.0556	0.0056	0.1667
$[40,50)$	6	0.1667	0.0167	0.3333
$[50-60)$	7	0.1944	0.0194	0.5278
$[60-70)$	5	0.1389	0.0139	0.6667
$[70-80)$	6	0.1667	0.0167	0.8333
$[80-90)$	3	0.0833	0.0083	0.9167
$[90-100)$	2	0.0556	0.0056	0.9722
$[100-110)$	1	0.0278	0.0028	1

(b)



(c) L'istogramma appare eccessivamente frastagliato, come se ci fossero troppe classi. In effetti con $n = 36$ dati, le solite regole empiriche ($\sqrt{n}$ e $1 + \log_2 n$) portano ad un istogramma con 6 classi anziché 11. Accorpando quindi le classi fornite si trova

Classi	F. Ass.	F. Rel.	Densità	F. Cum.
$[0, 20)$	2	0.0556	0.0028	0.0556
$[20 - 40)$	4	0.1111	0.0056	0.1667
$[40, 60)$	13	0.3611	0.0181	0.5278
$[60, 80)$	11	0.3056	0.0153	0.8333
$[80, 10)$	5	0.1389	0.0069	0.9722
$[100 - 120)$	1	0.0278	0.0014	1



(d) La migliore approssimazione della media dei dati è data chiaramente dalle classi iniziali, per cui

$$\overline{h}_{36} \simeq (2 \cdot 5 + 0 \cdot 15 + 2 \cdot 25 + 2 \cdot 35 + 6 \cdot 45 + 7 \cdot 55 + 5 \cdot 65 + 6 \cdot 75 + 3 \cdot 85 + 2 \cdot 95 + 1 \cdot 105)/36 = 58.61$$

(e) La migliore approssimazione del terzo quartile è data chiaramente dalle classi iniziali. Guardando la frequenza cumulata al punto a), vedo che $Q_3 \in [70 - 80)$. Pertanto

$$0.75 \simeq 0.6667 + (Q_3 - 70) \cdot 0,0167 \implies Q_3 \simeq 75.$$

Problema 2. Sia X una variabile aleatoria continua con densità

$$f_X(x) = \begin{cases} \alpha(x+1)^3, & -1 \le x \le 1, \\ 0, & \text{altrimenti.} \end{cases}$$

(a) Trovare la costante α e disegnare un grafico qualitativo di f_X.

(b) Trovare la funzione di ripartizione F_X e disegnarne un grafico qualitativo.

(c) Trovare la probabilità che X sia positivo.

(d) Vi aspettate una mediana positiva o negativa? Più grande o più piccola della media? Perché?

(e) Calcolare la mediana di X.

(f) Calcolare il valore atteso di X.

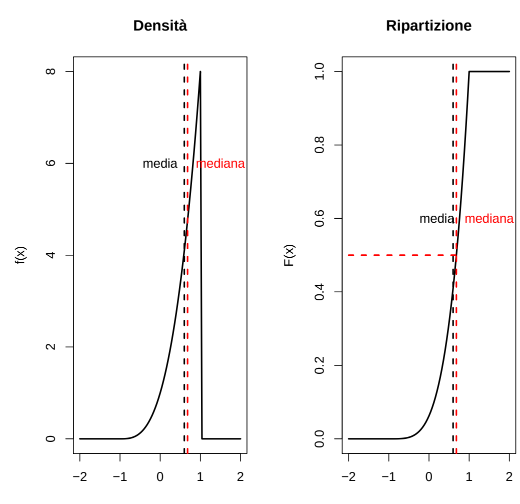
Risultati.

(a) $\begin{cases} \alpha \ge 0, \\ 1 = \int_{-1}^{1} \alpha(x+1)^3 \, \mathrm{d}x = \alpha \left. \dfrac{(x+1)^4}{4} \right|_{-1}^{1} \end{cases} \implies \alpha = \mathbf{1/4}.$

(b) Per $-1 \le x \le 1$ si ha $F_X(x) = \int_{-1}^{x} \frac{1}{4}(t+1)^3 \, \mathrm{d}t = \frac{(x+1)^4}{16}$, per cui

$$F_X(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{16}(x+1)^4, & -1 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

Grafici:



(c) $\mathbb{P}(X > 0) = 1 - \mathbb{P}(X \le 0) = 1 - F_X(0) = 1 - \frac{1}{16} = 0.9375.$

(d) $m > 0$ perchè $\mathbb{P}(X \le 0) < 0.5$, mentre $m < \mathbb{E}[X]$ perché f_X è asimmetrica con coda a sinistra.

(e) $F_X(m) = 0.5 \implies \frac{1}{16}(m+1)^4 = \frac{1}{2} \implies m+1 = \sqrt[4]{8} \implies m = 0.6817.$

(f) $\mathbb{E}[X] = \int_{-1}^{1} x \cdot \frac{1}{4}(x+1)^3 \, \mathrm{d}x = \left[\frac{1}{4}\frac{x^5}{5} + \frac{3}{4}\frac{x^4}{4} + \frac{3}{4}\frac{x^3}{3} + \frac{1}{4}\frac{x^2}{2} \right] \Big|_{-1}^{1} = \frac{1}{4}\left[\frac{1}{5} + \frac{1}{5}\right] + \frac{1}{4}[1+1] = \frac{6}{10} = 0.6.$

Problema 3. Una compagnia aerea copre una certa tratta con aerei da 330 posti. L'esperienza ha mostrato che, nonostante vengano sempre venduti tutti i biglietti di ogni volo, l'8% delle persone poi non si presenta alla partenza. Con lo scopo di ridurre il numero di posti non occupati, la compagnia vende 363 biglietti per ogni volo, superando così del 10% il numero dei posti disponibili.

(a) Qual è la probabilità che si presentino alla partenza tutti i 363 acquirenti?

(b) Qual è la probabilità che un aereo decolli con 5 o più posti vuoti?

(c) Qual è il minimo numero di biglietti da vendere, per volo, per ridurre tale probabilità sotto il 2.5%?

(d) Per tale numero n di biglietti venduti, calcolare la probabilità di lasciare a terra qualche acquirente presentatosi in aeroporto.

Risultati. Sia X il numero di acquirenti su 363 che si presentano per il volo, per cui $X \sim B(363, 0.92)$. Inoltre, poichè $363 \cdot 0.08 = 29.04 > 5$ e $363 \cdot 0.92 = 333.96 > 5$, vale l'approssimazione normale

$$X \sim B(363, 0.92) \simeq N(333.96\,,\,26.7168).$$

(a) $\mathbb{P}(X = 363) = 0.92^{363} = 7.2 \cdot 10^{-14}$.

(b) $\mathbb{P}(X \leq 325) = \mathbb{P}(X \leq 325.5) \simeq \Phi\left(\frac{325.5-333.96}{\sqrt{26.7168}}\right) = \Phi(-1.64) = 1 - \Phi(1.64) = 0.0508 = 5.08\%$.

Senza approssimazione normale, $\mathbb{P}(X \leq 325) = 0.0550 = 5.5\%$.

(c) Sia Y il numero di acquirenti su n che si presentano per il volo, per cui $Y \sim B(n, 0.92)$. Vogliamo minimo n tale che

$$\mathbb{P}(Y \leq 325) < 0.025$$

Per il punto precedete sarà $n > 363$, per cui

$$Y \sim B(n, 0.92) \simeq N(0.92n\,,\,0.0736n).$$

Allora

$$0.025 > \mathbb{P}(Y \leq 325) = \Phi\left(\frac{325.5 - 0.92n}{\sqrt{0.0736n}}\right)$$

$$\frac{325.5 - 0.92n}{\sqrt{0.0736n}} < -z_{0.025} = -1.96$$

$$n > 364.84$$

quindi $n = 365$.

(d) $\mathbb{P}(Y > 330) \simeq 1 - \Phi\left(\frac{330.5-365\cdot 0.92}{\sqrt{365\cdot 0.92\cdot 0.08}}\right) = 0.8467 = 84.67\%$.

Senza approssimazione normale, ancora $\mathbb{P}(Y > 330) = 0.8467 = 84.67\%$.