ESAME DI STATISTICA MATEMATICA A.A. 2017-18 (23.01.2018)

ESERCIZIO 1. Di seguito indichiamo le temperature massime giornaliere in gradi $^\circ C$ per il mese di giugno 2017 a Codroipo con un diagramma stem and leaf:

1 | 6, 7, 8.5, 8.5, 9, 9,

2 | 1, 1, 2, 2, 2.5, 3, 3, 3, 3, 4.5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6,

3 | 0, 0, 0, 0

Determinare la moda, la mediana campionaria, il primo quartile ed il terzo quartile.

ESERCIZIO 2. Sia $X_1, X_2, \ldots, X_n$ un campione causale con X_i variabile aleatoria gaussiana di media 2 e varianza 9. Sia $\bar{X}_n$ la media campionaria e sia S_n^2 la varianza campio– naria. Per $n = 10$ calcolare $\mathbb{E}[(\bar{X}_n - 2 + S_n)^2]$

ESERCIZIO 3. Supponiamo che la distribuzione di probabilità associata alla popolazione sia data da $p(x)dx$. La suddetta densità di probabilità $p(x)$ è data da

$$p(x) = \begin{cases} c\, x^\beta & \text{se } x \in [0, a]\,, \\ 0 & \text{altrimenti}\,, \end{cases}$$

dove $a > 0$ e $\beta > -1$ sono parametri non noti, mentre c è la costante di normalizzazione. Determinare gli stimatori $\hat{a}$ e $\hat{\beta}$ rispettivamente di a e β con il metodo dei momenti.

ESERCIZIO 4. Si suppone che il tempo di vita di un batterio (in secondi) sia uniformemente distribuito nell'intervallo $[0, a]$ e che la distribuzione a priori del parametro a sia uniforme sull'intervallo $[0, 10]$. Determinare lo stimatore bayesiano del parametro a.

ESERCIZIO 5. Supponiamo che la distribuzione delle popolazione sia la distribuzione uniforme sull'intervallo $[0, a]$, con $a > 0$ parametro non noto. Tramite una trasformazione che stabilizza la varianza, trovare un intervallo di confidenza asintotico per il parametro a con livello di confidenza pari a 95%.

ESERCIZIO 6. Si lancia un dado 1000 volte. I valori 1,2,3,4,5,6 escono con le seguenti frequenze assolute:

$$\begin{array}{ll} 1 \to 150 & 2 \to 160 \\ 3 \to 250 & 4 \to 140 \\ 5 \to 130 & 6 \to 170 \end{array}$$

Tramite il test χ^2 di Pearson con ampiezza asintotica 95% dire se accettiamo l'ipotesi nulla " il dado è onesto".

Traccia delle soluzioni

ESERCIZIO 1. Non vi è moda (25 e 26 sono i valori modali ma non vi è un valore che appare più di tutti gli altri).
Abbiamo $n = 30$ dati numerici.
$n(0.25) = 7.5$ quindi il primo quartile è l'ottavo valore in ordine crescente, cioè 21.
$n(0.5) = 15$ quindi il secondo quartile, detto anche mediana campionaria, è la media aritmetica tra il 15-esimo valore e il 16-esimo valore, cioè $(23 + 24.5)/2 = 23.75$.
$n(0.75) = 22.5$ quindi il terzo quartile è il 23–esimo valore, cioè 26.

ESERCIZIO 2. $\bar{X}_n$ ha legge $\mathcal{N}(2, 9/n)$, $\mathbb{E}[S_n] = 9$, S_n e $\bar{X}_n$ sono indipendenti. Abbiamo quindi (da notare che nell'ultima uguaglianza abbiamo usato la suddetta indipendenza)
$\mathbb{E}[(\bar{X}_n-2+S_n)^2] = \mathbb{E}[(\bar{X}_n-2)^2+2(\bar{X}_n-2)S_n+S_n^2] = \mathbb{E}[(\bar{X}_n-2)^2]+2\mathbb{E}[\bar{X}_n-2]\mathbb{E}[S_n]+\mathbb{E}[S_n^2]$.
$\bar{X}_n - 2 \sim \mathcal{N}(0, 9/10) \Longrightarrow \mathbb{E}[(\bar{X}_n - 2)^2] = 9/10$.
$2\mathbb{E}[\bar{X}_n - 2]\mathbb{E}[S_n] = 2 \cdot 0 \cdot \mathbb{E}[S_n] = 0$; $\mathbb{E}[S_n^2] = 9$. Concludiamo che $\mathbb{E}[(\bar{X}_n - 2 + S_n)^2] = 9/10 + 9 = 9.9$.

ESERCIZIO 3. Siccome (avendo $\beta > -1$) $\int_0^a x^\beta dx = \frac{a^{\beta+1}}{\beta+1}$, abbiamo $c = \frac{\beta+1}{a^{\beta+1}}$. Chiamato μ_i il momento i-esimo della distribuzione abbiamo

$$\mu_1 = c\frac{a^{\beta+2}}{\beta+2} = a\frac{\beta+1}{\beta+2}$$
$$\mu_2 = c\frac{a^{\beta+3}}{\beta+3} = a^2\frac{\beta+1}{\beta+3}.$$

Chiamati M_1, M_2 i momenti campionari (omettiamo indice n) deve essere

$$M_1 = \hat{a}\frac{\hat{\beta}+1}{\hat{\beta}+2}$$
$$M_2 = \hat{a}^2\frac{\hat{\beta}+1}{\hat{\beta}+3}.$$

Scrivo la prima come $\hat{a} = M_1\frac{\hat{\beta}+2}{\hat{\beta}+1}$. Quindi la seconda diviene

$$M_2 = M_1^2\left(\frac{\hat{\beta}+2}{\hat{\beta}+1}\right)^2\frac{\hat{\beta}+1}{\hat{\beta}+3}, \qquad \text{equivalentemente} \qquad M_2 = M_1^2\frac{(\hat{\beta}+2)^2}{(\hat{\beta}+1)(\hat{\beta}+3)}$$

Pongo $\gamma = M_1^2/M_2$ (sappiamo che $0 < \gamma < 1$ a.s.) ottengo due soluzioni $\hat{\beta} = -2 \pm \sqrt{\frac{1}{1-\gamma}}$.
Siccome lavoriamo con $\beta > -1$, tengo solo la soluzione $\hat{\beta} := -2 + \sqrt{\frac{1}{1-\gamma}}$ che è maggiore di -1 (recall $\gamma \in (0,1)$ a.s.). La conclusione è

$$\hat{\beta} := -2 + \sqrt{\frac{1}{1 - M_1^2/M_2}} = -2 + \sqrt{\frac{M_2}{M_2 - M_1^2}}$$

e

$$\hat{a} = M_1\frac{\hat{\beta}+2}{\hat{\beta}+1} = M_1\frac{\sqrt{\frac{M_2}{M_2-M_1^2}}}{-1+\sqrt{\frac{M_2}{M_2-M_1^2}}} = \frac{\sqrt{M_1^2 M_2}}{\sqrt{M_2} - \sqrt{M_2 - M_1^2}}$$

ESERCIZIO 4. Distribuzione a priori di a è $p(a)da = \frac{1}{10}\mathbb{1}(a \in [0,10])\,da$. Abbiamo

$$p(x_1, x_2, \ldots, x_n|a) = \frac{1}{a^n}\mathbb{1}(0 \leq x_1, \ldots, x_n \leq a)$$

Per Bayes

$$p(a|x_1, x_2, \ldots, x_n) = Cp(x_1, x_2, \ldots, x_n|a)p(a)$$

con C constante di normalizzazione. Quindi

$$p(a|x_1, x_2, \ldots, x_n) = C\frac{1}{a^n}\mathbb{1}(0 \leq x_1, \ldots, x_n \leq a)\frac{1}{10}\mathbb{1}(a \in [0,10])$$

Limitiamoci a $x_1, \ldots, x_n \in [0,10]$ (cose diverse non ne osserveremo) e poniamo $M_n := \max\{x_1, x_2, \ldots, x_n\}$, allora abbiamo

$$p(a|x_1, x_2, \ldots, x_n) = \bar{C}\frac{1}{a^n}\mathbb{1}(a \geq M_n)\mathbb{1}(a \in [0,10]) = \bar{C}\frac{1}{a^n}\mathbb{1}(M_n \leq a \leq 10)$$

con $\bar{C}$ costante di normalizzazione. Quindi, per $n \geq 2$,

$$\bar{C}^{-1} = \int_{M_n}^{10} \frac{1}{a^n}da = \frac{M_n^{1-n}}{n-1} - \frac{10^{1-n}}{n-1}\,.$$

Sopra, e di seguito, abbiamo assunto $M_n > 0$, evento verp quasi certamente. Siccome, per $n \geq 3$,

$$\int_{M_n}^{10} \frac{a}{a^n}da = \frac{M_n^{2-n}}{n-2} - \frac{10^{2-n}}{n-2}\,,$$

concludiamo che, per $n \geq 3$,

$$E[a|X_1, \ldots, X_n] = \bar{C}\left[\frac{M_n^{2-n}}{n-2} - \frac{10^{2-n}}{n-2}\right] = \frac{\frac{M_n^{2-n}}{n-2} - \frac{10^{2-n}}{n-2}}{\frac{M_n^{1-n}}{n-1} - \frac{10^{1-n}}{n-1}}\,, \qquad M_n := \max\{X_1, X_2, \ldots, X_n\}\,.$$

ESERCIZIO 5. Calcolo varianza di X_1. Ho $\mathbb{E}[X_1] = a/2$ e

$$\mathbb{E}[X_1^2] = \frac{1}{a}\int_0^a x^2dx = \frac{1}{a}\frac{a^3}{3} = \frac{a^2}{3}\,.$$

Quindi $Var[X_1] = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = a^2/3 - a^2/4 = a^2/12$. CLT ci dice che

$$\sqrt{n}[\bar{X}_n - \frac{a}{2}] \Longrightarrow \mathcal{N}(0, a^2/12)$$

in legge. Metodo delta ci dice che, con f derivabile in $a/2$, vale

$$\sqrt{n}[f(\bar{X}_n) - f(\frac{a}{2})] \Longrightarrow \mathcal{N}(0, f'(a/2)^2a^2/12)\,.$$

Cerco f tale che $f'(a/2)^2a^2/12 = 1$, quindi $f'(a/2)^2 = 12/a^2 = 3/(a/2)^2$. Basta avere $f'(x) = \sqrt{3}/x$. Prendo $f(x) = \sqrt{3}\ln x$ e ottengo

$$\sqrt{n}[\sqrt{3}\log(\bar{X}_n) - \sqrt{3}\log(a/2)] \Longrightarrow \mathcal{N}(0,1)\,.$$

Per n grande ho quindi

$$\mathbb{P}[\sqrt{n}|\sqrt{3}\log(\bar{X}_n) - \sqrt{3}\log(a/2)| \leq z_{\alpha/2}] \approx \mathbb{P}(|Z| \leq z_{\alpha/2}) = 1 - 2\mathbb{P}(Z > z_{\alpha/2}) = 1 - \alpha$$

Quindi

$$\mathbb{P}\Big[\log(a/2) \in [\log(\bar{X}_n) - \frac{z_{\alpha/2}}{\sqrt{3n}}, \log(\bar{X}_n/2) + \frac{z_{\alpha/2}}{\sqrt{3n}}]\Big] \approx 1 - \alpha\,,$$

quindi

$$\mathbb{P}\Big[a \in [2\bar{X}_n e^{-\frac{z_{\alpha/2}}{\sqrt{3n}}}, 2\bar{X}_n e^{\frac{z_{\alpha/2}}{\sqrt{3n}}}]\Big] \approx 1 - \alpha\,.$$

Prendo $\alpha = 0.05$, ho $z_{\alpha/2} = z_{0.025} \approx 1.96$. L'intervallo cercato è quindi $[2\bar{X}_n e^{-\frac{1.96}{\sqrt{3n}}}, 2\bar{X}_n e^{\frac{1.96}{\sqrt{3n}}}]$.

ESERCIZIO 6. Poniamo $p_{0,j} = 1/6$ e $n = 1000$. SiaN_i il numero di volte in qui nei 1000 lanci è uscita la faccia i. Sia T la statistica di Pearson:

$$\begin{aligned}
T &= \sum_{j=1}^{6} \frac{(N_j - np_{0,j})^2}{np_{0,j}} = \frac{1000}{6}\sum_{j=1}^{6}(\frac{N_j}{np_{0,j}} - 1)^2 = \frac{1000}{6}\sum_{j=1}^{6}(6\frac{N_j}{1000} - 1)^2 \\
&= \frac{1000}{6}\Big[(6\cdot 0,15 - 1)^2 + (6\cdot 0.16 - 1)^2 + (6\cdot 0.25 - 1)^2 \\
&+ (6\cdot 0.14 - 1)^2 + (6\cdot 0.13 - 1)^2 + (6\cdot 0.17 - 1)^2\Big] \\
&= \frac{1000}{6}\Big[(0.1)^2 + (0.04)^2 + (0,5)^2 + (0.16)^2 + (0.22)^2 + (0.02)^2\Big] = \frac{1000}{6}0.336 = 56\,.
\end{aligned}$$

Posto $\alpha = 0.95$ abbiamo: rifiuto l'ipotesi nulla se e solo se $T > \chi^2_{5,\alpha}$. Siccome $\chi^2_{5,0.95} = 1.145$, rifiutiamo l'ipotesi nulla.

A parte la matematica consistente, il testo piu'ragionevole sarebbe stato "con ampiezza asintotica del 5%" Allora rifiutavamo l'ipotesi nulla se e solo se $T > \chi^2_{5,0.05} = 11.070$. Quindi anche in questo caso avrei rifiutato H_0