**Statistica - CPS**
**Corso di Laurea in Informatica**
**Compito del 27-06-2022**

**Esercizio 1 (10 punti)**

Una ditta produce certi componenti elettronici dei quali circa il 20% sono difettosi: questi componenti sono esportati in scatole da 400 pezzi e la ditta si impegna a sostituire integralmente la scatola se il numero di pezzi difettosi è superiore a 90.

(i) Qual è (approssimativamente) la probabilità che la ditta debba sostituire una scatola?

(ii) Se si vuole che la probabilità di dover sostituire una scatola sia inferiore a $0.05$, come deve migliorare la produzione (cioè di quanto deve -approssimativamente- scendere la percentuale di pezzi difettosi)?

**Esercizio 2 (10 punti)**

È assegnata la funzione

$$f(x) = \begin{cases} -x & \text{per } -1 \le x < 0 \\ ax + b & \text{per } 0 \le x < 1 \\ 0 & \text{altrove} \end{cases}$$

dove $a$ e $b$ sono due parametri reali.

(i) Trovare tutti i valori di $a$ e $b$ per i quali $f$ è una densità di probabilità.

(ii) Sia $X$ una v.a. avente densità $f$ (con $a$ e $b$ che soddisfano alle condizioni trovate nel punto precedente): calcolare $a$ e $b$ in modo che si abbia $E[X] = 0$.

(iii) In funzione di $a$ e $b$, calcolare $\mathbf{P}\Big\{|X| \ge \frac{1}{2} \Big| X \le 0\Big\}$.

**Esercizio 3 (10 punti)**

Si hanno a disposizione 16 osservazioni indipendenti di una v.a. Gaussiana con media $\mu$ sconosciuta e varianza nota eguale a 36; con queste oservazioni si vuole effettuare il test dell'ipotesi $H_0\big)\ \mu = 30$ contro $H_1\big)\ \mu \neq 30$.

Si decide di respingere l'ipotesi $H_0\big)$ se la media campionaria delle 16 osservazioni cade al di fuori dell'intervallo $\big(26.91\,,\,33.09\big)$.

(i) A quale livello viene effettuato il test?

(ii) Se le osservazioni fossero 25 e il test venisse effettuato ancora allo stesso livello del punto (i), quale sarebbe la regione critica?

**Una soluzione:**

**Esercizio 1** i) Se si indica con $X$ la v.a. che conta il numero di pezzi difettosi, $X$ è Binomiale di parametri 400 e 0.2 e, per il Teorema Limite Centrale,

$\dfrac{X - 400 \times 0.2}{\sqrt{400 \times 0.2 \times 0.8}} = \dfrac{X - 80}{20\sqrt{0.2 \times 0.8}} = \dfrac{X - 80}{8}$ si approssima con una variabile Gaussiana standard.

Di conseguenza

$$\mathbf{P}\big\{X \geq 90\big\} = \mathbf{P}\big\{\frac{X-80}{8} \geq \frac{90-80}{8}\big\} \approx 1 - \Phi(1.25) = 0.106$$

ii) Ripetendo gli stessi conti con un generico numero $p$ al posto di $0.2$, e poiché per una variabile $Y$ con densità $N(0,1)$ si ha $\mathbf{P}\big\{Y > d\big\} = 0.05$ se $d = q_{0.95} = 1.65$, si deve individuare un numero $p$ (da cercare inferiore a 0.2) tale che $\frac{90 - 400 \times p}{20\sqrt{p(1-p)}} \approx 1.65$.

Per tentativi, prendendo $p = 0.19$ si ottiene il valore 1.78 che è leggermente superiore a quello cercato: dunque già riuscendo a migliorare la produzione in modo che la percentuale di pezzi difettosi non superi il 19 % si ottiene il risultato voluto.

**Esercizio 2**

(i) Cominciamo a imporre che l'integrale della funzione $f$ sia eguale ad 1: si ha

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_{-1}^{0} (-x)\,\mathrm{d}x + \int_{0}^{1} (ax+b)\,\mathrm{d}x = \frac{1}{2} + \frac{a}{2} + b = 1$$

e da qui si ricava l'eguaglianza $a+2b = 1$.

Tuttavia la funzione $f$ deve anche essere a valori positivi: prendendo $x = 0$ si ottiene $b \geq 0$ e prendendo $x = 1$ si ottiene $a+b \geq 0$.

Mettendo insieme queste diseguaglianze ed eguaglianze, si ottiene $0 \leq b \leq 1$ e $a = 1–2b$.

(ii) Calcolando il valore atteso si ha

$$\mathbf{E}[X] = \int_{-1}^{0} -x^2\,\mathrm{d}x + \int_{0}^{1} (ax^2 + b)\,\mathrm{d}x = -\frac{1}{3} + \frac{a}{3} + \frac{b}{2} = \frac{-2b}{3} + \frac{b}{2}$$

ed è evidente che affinché $\mathbf{E}[X]$ sia eguale a 0 si deve avere $b = 0$ e $a = 1$.

(iii) Notiamo che l'insieme $\big\{|X| \geq 1/2\big\} \cap \big\{X \leq 0\big\} = \big\{|X| \geq 1/2\,,\ X \leq 0\big\}$ coincide con $\big\{-1 \leq X \leq -1/2\big\}$ (infatti la variabile prende valori tra -1 e 1 poiché la densità è diversa da 0 solo in quell'intervallo): si ha pertanto

$$\mathbf{P}\Big\{|X| \geq \frac{1}{2}\Big| X \leq 0\Big\} = \frac{\mathbf{P}\big\{-1 \leq X \leq -1/2\big\}}{\mathbf{P}\big\{-1 \leq X \leq 0\big\}} = \frac{\int_{-1}^{-1/2}(-x)\,\mathrm{d}x}{\int_{-1}^{0}(-x)\,\mathrm{d}x} =$$

$$= \frac{-\frac{1}{8} + \frac{1}{2}}{\frac{1}{2}} = \frac{3}{4}$$

e come si vede questa probabilità condizionata non dipende da $a$ e da $b$.

**Esercizio 3**

(i) Il test dell'ipotesi $\mathcal{H}_0\big)\ \mu = \mu_0$ (sulla media di un campione gaussiano con varianza nota) ha una regione critica della forma $\Big\{\,\big|\overline{X} - \mu_0\big| > d\Big\}$ dove $d = \dfrac{\sigma}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}$.

In questo caso, essendo $\mu_0 = 30$, si ottiene per $d$ il valore $3.09$, e poichè $\sigma = 6$ e $\sqrt{n} = \sqrt{16} = 4$, si ricava $q_{1-\frac{\alpha}{2}} = 3.09 \times \frac{2}{3} = 2.06$.

Poiché $\Phi(2.06) = 0.98$, dall'equazione $1 - \frac{\alpha}{2} = 0.98$ si ottiene $\alpha = 0.04$, cioè il livello del test è del $0.04$.

(ii) Ricordiamo sempre l'espressione del numero $d = \dfrac{\sigma}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}$ : di conseguenza se la numerosità del campione passa da 16 a 25 (tutto il resto restando invariato) il numero del dato precedente (cioè 3.09) deve essere moltiplicato per $\sqrt{16} = 4$ e diviso per $\sqrt{25} = 5$, e si ottiene $2.472$.

Più precisamente, la regione critica diventa $\left\{ \left|\overline{X} - 30\right| > 2.472 \right\}$, cioè si respinge l'ipotesi se la media campionaria delle 25 osservazioni cade al di fuori dell'intervallo $\big(27.528\,,\, 32.472\big)$