Estatística II, Exame época especial 2017/07/21 RESOLUÇÃO

1 X_1 e X_2 são dois acontecimentos com probabilidade positiva de um espaço de resultados.

a) Se X_1 e X_2 são incompatíveis, mostre que X_1 e X_2 são dependentes.

Resolução

X_1 e X_2 são incompatíveis ou seja, $\Pr(X_1 \cap X_2) = \Pr(\emptyset) = 0$.

Donde,

$\Pr(X_2|X_1) = \Pr(X_1 \cap X_2)/\Pr(X_1) = 0/\Pr(X_1) = 0 \neq \Pr(X_2)$, porque $\Pr(X_1) > 0$ e $\Pr(X_2) > 0$.

Ou seja, X_1 e X_2 são dependentes.

b) Se $X_1 \subset X_2$, mostre que $\Pr(X_1) \leq \Pr(X_2)$.

Res.

Se $X_1 \subset X_2$, então $X_2 = X_1 \cup (X_2 - X_1)$, com $X_1 \cap (X_2 - X_1) = \emptyset$.

Donde,

$\Pr(X_2) = \Pr[X_1 \cup (X_2 - X_1)] = \Pr(X_1) + \Pr(X_2 - X_1) \geq \Pr(X_1)$, porque $\Pr[X_1 \cap (X_2 - X_1)] = \Pr(\emptyset) = 0$ e $\Pr(X_2 - X_1) \geq 0$ (porque qualquer probabilidade é não negativa.

2 Numa lotaria selecciona-se aleatoriamente, em cada dia, um número inteiro com três algarismos, de 000 até 999, inclusive. Seja X o número seleccionado em qualquer dia. Calcule

a) $\Pr(X \geq 500 \mid X$ é par).

Res.

$$\Pr(X \geq 500 \mid X \text{ par}) = \Pr(X \geq 500 \wedge X \text{ par})/\Pr(X \text{ par}) =$$
$$\Pr(X \in \{500{,}502, \dots ,998\})/\Pr(X \in \{0{,}2, \dots ,998\}) = (250/1000)/(500/1000) = 1/2.$$

b) $E(X)$.

Res.

$$E(X) = \sum_{x=0}^{999} x f_X(x) = \sum_{x=0}^{999} x/1000 = \{[1000 \times (0 + 999)]/2\}/1000 = 499{,}5.$$

(Note-se que $\sum_{x=0}^{999} x$ é uma progressão aritmética.)

3

a) De uma população $\mathcal{N}(8{,}16)$ extrai-se uma amostra casual de dimensão 100. Qual a probabilidade de o módulo da diferença entre a média amostral e a média populacional ser superior a 0,5?

Res.

$\bar{X} \sim \mathcal{N}[E(X), V(X)/100] \equiv \mathcal{N}(8; 0{,}16)$. Donde,

$$\Pr(|\bar{X} - 8| > 0{,}5) = \Pr\left(\left|\frac{\bar{X} - 8}{0{,}4}\right| > \frac{0{,}5}{0{,}4}\right) = \Pr(Z > 1{,}25 \vee Z < -1{,}25),$$

em que $Z \equiv (\bar{X} - 8)/0{,}4 \sim \mathcal{N}(0{,}1)$.

$$\Pr(Z > 1{,}25 \vee Z < -1{,}25) = \Pr(Z > 1{,}25) + \Pr(Z < -1{,}25) = 1 - \Phi(1{,}25) + 1 - \Phi(1{,}25) = 2 - 2 \times 0{,}8944 = 0{,}2112.$$

b) Considere que $Y \sim \mathcal{N}(20{,}25)$. Determine a dimensão mínima de uma amostra iid tal, que $\Pr(18 < \bar{Y} < 22) \geq 0{,}9$ ($\bar{Y}$: média amostral).

Res.

Seja n a dimensão da amostra. Tem-se $\bar{Y} \sim \mathcal{N}[E(Y), V(Y)/n] \equiv \mathcal{N}(20; 25/n)$. Donde,

$$\Pr(18 < \bar{Y} < 22) = \Pr\left(\frac{18-20}{5/\sqrt{n}} < Z < \frac{22-20}{5/\sqrt{n}}\right) = \Pr\left(-0{,}4\sqrt{n} < Z < 0{,}4\sqrt{n}\right) =$$
$$\Phi\left(0{,}4\sqrt{n}\right) - \Phi\left(-0{,}4\sqrt{n}\right) = 2\Phi\left(0{,}4\sqrt{n}\right) - 1,$$

em que $Z \equiv (\bar{Y}-20)/\left(5/\sqrt{n}\right) \sim \mathcal{N}(0{,}1)$.

$$2\Phi\left(0{,}4\sqrt{n}\right) - 1 \geq 0{,}9 \Leftrightarrow \Phi\left(0{,}4\sqrt{n}\right) \geq 0{,}95 \Leftrightarrow 0{,}4\sqrt{n} \geq 1{,}645 \Rightarrow n \geq 16{,}913$$

ou seja, a dimensão mínima da amostra é 17.

4 Pretende-se ter uma ideia da densidade do tráfego em certo local, entre as oito e as nove horas. O número de veículos que aí passa durante este período, X, segue uma distribuição de Poisson com parâmetro λ. Recolhe-se a amostra casual de 9 observações,

$$(95{,}100{,}80{,}70{,}110{,}98{,}97{,}90{,}70).$$

a) Estime por máxima verosimilhança (MV) a média do número de veículos por hora.

Função probabilidade *Poisson* $\quad f_X(x;\lambda) = e^{-\lambda}\lambda^x/x!, \quad x \in \mathcal{N}_0, \quad \lambda = E(X) = V(X) > 0.$

Res.

$$l(\lambda) = \sum_{i=1}^{n} \log\left[e^{-\lambda}\lambda^{x_i}/x_i!\right] = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\log\lambda - \sum_{i=1}^{n}\log(x_i!) \Rightarrow$$
$$l'(\lambda) = -n + \sum_{i=1}^{n} x_i/\lambda = 0 \Leftrightarrow \lambda = \sum_{i=1}^{n} x_i/n = \bar{x}$$

Estimador MV: $\quad \hat{\lambda} = \bar{X}$.

Estimativa MV a partir da amostra: $\quad \bar{x} = (95 + \cdots + 70)/9 = 90.$

Se não resolveu a alínea ***a***), nas alíneas seguintes considere o estimador MV de λ, $\hat{\lambda} = \sum_{i=1}^{n} X_i/n$, obtido a partir de uma amostra casual de dimensão n.

b) Estime por MV a probabilidade de decorrerem dois minutos sem passar qualquer veículo.

Função densidade exponencial negativa $\quad f_T(t;\lambda) = \nu e^{-\nu t}, \quad t > 0, \quad \nu > 0.$

Res.

O exercício pode resolver-se através da densidade exponencial negativa ou da probabilidade Poisson; recorrendo à Poisson:

Y: número de veículos por cada período de dois minutos

$f_Y(y) = \exp(-\lambda/30)\,(\lambda/30)^y/y! \qquad \Pr(Y=0) = \exp(-\lambda/30);$

$$\Pr(\widehat{Y=0}) = \exp\left(-\hat{\lambda}/30\right) = \exp(-90/30) \approx 0{,}0498.$$

c) Mostre que $\hat{\lambda}$ é consistente para λ.

Res.

$$\begin{cases} E\left(\hat{\lambda}\right) = E(\bar{X}) = E(X) = \lambda \\ V\left(\hat{\lambda}\right) = V(\bar{X}) = V(X)/n \to 0 \end{cases} \Rightarrow \operatorname{plim}_{n\to\infty}\hat{\lambda} = \lambda.$$

d) Seja o estimador alternativo de λ, $\tilde{\lambda} = (X_1 + X_n)/2$, em que X_1 e X_n designam, respectivamente, o primeiro e último elementos da amostra. Qual dos estimadores, $\hat{\lambda}$ ou $\tilde{\lambda}$, é mais eficiente?

Res.

$V\left(\hat{\lambda}\right) = V(\bar{X}) = V(X)/n = \lambda/n$

$V\left(\tilde{\lambda}\right) = V[(X_1 + X_n)/2] = [V(X_1) + V(X_n)]/4 = [V(X) + V(X)]/4 = (\lambda + \lambda)/4 = \lambda/2.$

$V\left(\hat{\lambda}\right) \leq V\left(\tilde{\lambda}\right)$, se $n \geq 2$. Se $n = 2$, os dois estimadores coincidem.

5 De uma amostra casual de dimensão 100 de população $N(\mu, 100)$, obtém-se $\bar{x} = 100$.

a) O extremo superior de um intervalo de confiança (IC) para μ é 102,326. Qual o grau de confiança utilizado na construção deste IC?

Res.

O extremo superior do IC é dado por $100 + \left(10/\sqrt{100}\right) \times \xi_{1-\alpha/2} = 100 + \xi_{1-\alpha/2}$, em que $\xi_{1-\alpha/2}$ denota o quantil de ordem $1 - \alpha/2$ da densidade $\mathcal{N}(0,1)$ – porque a variável fulcral é

$$(\bar{X} - \mu)/\sqrt{\sigma^2/n} \sim \mathcal{N}(0,1),$$

com $\sigma^2 = 100$ e $n = 100$.

Grau de confiança:

$$100 + \xi_{1-\alpha/2} = 102{,}326 \Leftrightarrow \xi_{1-\alpha/2} = 2{,}326 \Leftrightarrow 1 - \alpha/2 = 0{,}99 \Leftrightarrow 1 - \alpha = 0{,}98.$$

b) Num grande número de amostras casuais de dimensão 100, qual a percentagem de intervalos de confiança com a amplitude considerada em ***a***), que <u>não</u> contêm o verdadeiro valor de μ?

Res.

A percentagem pedida é o nível de significância (em percentagem): $\alpha \times 100\% = 2\%$.