UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Avaliação Final de Probabilidade e Estatística **02-07-2019**

Atenção: Não serão aceitas respostas sem justificativa.
Resolver as questões nos espaços apropriados.

Q1) Sejam (X_1, X_2) a variável aleatória bidimensional cuja distribuição conjuntadada é dada pela tabela:

$X_1 \setminus X_2$	-1	0	1	$p_1(x_1)$
-1	0,2	0,0	0,2	0,4
0	0,0	0,2	0,0	0,2
-1	0,2	0,0	0,2	0,4
$p_2(x_2)$	0,4	0,2	0,4	1

(a) Calcule E(X_i) e Var(X_i), sendo i=1,2.

(b) Calcule E$(X_1 X_2)$, Cov(X_1, X_2) e Var$(X_1 - X_2)$. As v.a.'s X_1 e X_2 são independentes?

(c) O que se pode dizer sobre a veracidade da afirmação:
"Se X_1 e X_2 são independentes, então Cov(X_1, X_2)=0"? Justifique.

(d) O que se pode dizer com relação a esta afirmação:
"Se Cov(X_1, X_2)=0, então X_1 e X_2 são independentes"? Justifique.

Q2) A quantidade de partículas emitidas por uma fonte radioativa durante um período de uma hora é uma variável aleatória de Poisson com parâmetro λ, a princípio desconhecido. Sabe-se que a probabilidade de nenhuma partícula ser emitida durante esse intervalo de tempo é de 0,00673.

(a) Em média, quantas partículas são emitidas por hora? Escreva também a expressão da função densidade de probabilidade de T, o tempo em horas entre duas emissões consecutivas.

(b) Sabendo que em 20 minutos após uma emissão nenhuma outra partícula foi emitida, qual a probabilidade de que precisamos esperar mais meia hora pela próxima emissão?

(c) Sejam $X_1, \ldots, X_{24}$ os números de partículas emitidas pela fonte ao longo das 24 horas de um determinado dia, e admita que as quantidades de partículas emitidas em horas distintas são independentes entre sí. Qual é a probabilidade de, ao longo dessas 24 horas, observarmos a emissão de no máximo 100 partículas?

Q3) Sejam $\bar{X}_1$ a média de uma amostra aleatória de dimensão n extraída de uma população normal de valor esperado μ e variância σ_1^2, e $\bar{X}_2$ a média de uma amostra aleatória de dimensão n, independente da primeira, extraída de uma população normal de valor esperado μ e variância σ_2^2. Mostre que:

(a) Se $w \in [0, 1]$, $w\bar{X}_1 + (1 - w)\bar{X}_2$ é um estimador não viesado de μ.

(b) A variância do estimador em (a) é mínima quando $w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$

Q4) Em uma linha de produção, o engenheiro responsável pelo controle de qualidade deseja estimar a proporção p de itens defeituosos produzidos. Para isso, seleciona aleatoriamente uma amostra de 50 itens, observando 4 itens defeituosos. Historicamente, mesmo em ocasiões de descontrole na linha de produção, a proporção de itens defeituosos jamais superou 15% e o engenheiro acredita que esse limiar continue a valer.

(a) Construa um intervalo não conservativo ao nível de confiança 95% para a proporção de itens defeituosos na linha de produção.

(b) Qual deve ser o tamanho de amostra coletada para que, ao se utilizar a proporção amostral de itens defeituosos ($\hat{p}$) como estimador da proporção de itens defeituosos (p) na linha de produção, o erro absoluto de estimação $|\hat{p} - p|$ seja inferior a 0,05, com probabilidade 95%?

(c) Com base nos resultados experimentais calcule o p-valor correspondente ao teste da hipótese H_0 de que a proporção de itens defeituosos não supera 7% contra a alternativa H_1 de que supera 7%.

Boa Prova!

Solução

Q1) (a) $E(X_i) = -1 \times 0,4 + 0 \times 0,2 + 1 \times 0,4 = 0$
$Var(X_i) = E(X_i^2) = (-1)^2 \times 0,4 + 0^2 \times 0,2 + 1^2 \times 0,4 = 0,8$

(b) $E(X_1X_2)) = 1 \times 0,2 + (-1) \times 0,2 + 1 \times 0,2 + (-1) \times 0,2 = 0$
$Cov(X_1, X_2)) = E(X_1X_2)) - E(X_1)E(X_2)) = 0$
$Var(X_1 - X_2)) = Var(X_1) + Var(X_2)) - 2Cov(X_1, X_2)) = 1,6$
X_1 e X_2 não são v.a.'s independentes! Pois, p(-1, -1)=0,2 $\neq$ p(-1) p(-1) = 0,4 0,4=0,16

(c) Como, $Cov(X, Y) = E(XY) - E(X)E(Y)$, quando X e Y são independentes, $Cov(X, Y) = 0$.

(d) Nos itens (a) e (b) acima vimos um contra exemplo que mostra que a recíproca da afirmação do item (c) não é verdadeira. X_1 e X_2 não são independentes, mas a Covariância é zero.

Q2) (a) Seja X o número de partículas emitidas pela fonte radioativa em um período de uma hora. Primeiramente calculemos o parâmetro λ. Temos que

$$0,00673 = P(X = 0) = e^{-\lambda} \Rightarrow \lambda = -\ln(0,00673) = 5{,}00,$$

na precisão de duas casas decimais. Como $E[X] = \lambda$, em média 5 partículas são emitidas por hora. Sabemos que $T \sim \text{Exp}(\lambda)$, de modo que a densidade de T é dada por$f_T(t) = 5e^{-5t}$, para $t > 0$ e $f_T(t) = 0$, se $t < 0$.

(b) Pela propriedade da perda de memória da exponencial, temos que:

$$\begin{aligned} P(T > 5/6|T > 1/3) &= P(T > 1/2) \\ &= \int_{1/2}^{\infty} 5e^{-5t}\, dt = 0{,}08 = 8\%. \end{aligned}$$

(c) Seja $Y = X_1 + \ldots, +X_{24}$ o número de partículas emitidas ao longo do dia. Sabemos que $Y \sim \text{Poi}(5 \times 24) = \text{Poi}(120)$, porém é inviável trabalhar com tal distribuição. Por isso, com base no TCL, a aproximamos por uma v.a. $\tilde{Y}$, com distribuição Normal de média $\mu = \lambda = 120$ e variância $\sigma^2 = \lambda = 120$. Usando a correção para a continuidade, temos que:

$$\begin{aligned} P(Y \leq 100) &= (Y \leq 100,5) \\ &\approx P(\tilde{Y} \leq 100,5) = P\left(\frac{\tilde{Y} - 120}{\sqrt{120}} < \frac{100,5 - 120}{\sqrt{120}}\right) \\ &= \Phi(-1,78) = 1 - 0,9625 = 0,0375. \end{aligned}$$

Q3) (a) $w\bar{X}_1 + (1-w)\bar{X}_2$ é não viesado se $E[w\bar{X}_1 + (1-w)\bar{X}_2] = \mu$. Por hipótese do exercício $E(\bar{X}_1) = E(\bar{X}_2) = \mu$. Então, temos que $E[w\bar{X}_1 + (1-w)\bar{X}_2] = wE(\bar{X}_1) + (1-w)E(\bar{X}_2) = w\mu + (1-w)\mu = \mu$. Logo, $w\bar{X}_1 + (1-w)\bar{X}_2$ é não viesado.

(b) A variância de $w\bar{X}_1 + (1-w)\bar{X}_2$ é dada por

$$V(w) = Var(w\bar{X}_1 + (1-w)\bar{X}_2) = w^2 Var(\bar{X}_1) + (1-w)^2 Var(\bar{X}_2) = w^2\frac{\sigma_1^2}{n} + (1-w)^2\frac{\sigma_2^2}{n}$$

Igualando a zero a derivada dessa expressão com relação a w, conclui-se que $Var(w\bar{X}_1 + (1-w)\bar{X}_2)$ será mínima se

$$\frac{dV}{dw} = 2w\frac{\sigma_1^2}{n} - 2(1-w)\frac{\sigma_2^2}{n} = 0$$

ou equivalentemente se

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

A garantia de que o ponto é de mínimo está no fato da segunda derivada ser sempre positiva:

$$\frac{d^2V}{dw^2} = 2\frac{\sigma_1^2}{n} + 2\frac{\sigma_2^2}{n} > 0$$

Q4) (a) Os extremos do Intervalo de Confiança serão dados por:

$4/50 \pm 1,96\sqrt{\frac{4/50 \times 46/50}{50}}$, ou seja

$0,08 \pm 0,0752 = (0,0048; 0,1552)$.

(b) Para que $P[|\hat{p} - p| < 0,05] = 0,95$,

o tamanho amostral deve ser

$n = (\frac{z_{0,975}}{0,05})^2 p(1-p)$.

Como p≤0,15, podemos fazer

$n = (\frac{1,96}{0,05})^2 \times 0,15 \times 0,85 = 195,92 \approx 196$

(c) p-valor=1-$\Phi\left(\frac{0,08-0,07}{\sqrt{\frac{0,07 \times 0,93}{50}}}\right) = 1 - \Phi(0,28) = 0,3897$. Com o p-valor tão grande não há evidência para a rejeição da hipótese H_0.