Estatística e Probabilidade - Avaliação “Presencial” 02 - 2022/02 - Gabarito

Prof. Hugo Carvalho

02/01/2023

Questão 1:

a) Encontremos a função de probabilidade de $X+Y$. Primeiramente, note que $X+Y$ pode assumir os valores $0, 1, 2, \ldots$. Dado z em tal conjunto, queremos estudar a probabilidade do evento $\{X+Y=z\}$. Note que, sendo $X=x$, Y necessariamente deverá assumir o valor $z-x$, para $x=0,1,\ldots,z$. Dessa forma, o evento $\{X+Y=z\}$ pode ser reescrito como $\bigcup_{x=0}^{z}\{X=x \text{ e } Y=z-x\}$, de modo a nos permitir utilizar a independência entre X e Y, sendo os conjuntos dessa união dois-a-dois disjuntos. Temos então que:

$$\begin{aligned}
\mathbb{P}(X+Y=z) &= \mathbb{P}\left(\bigcup_{x=0}^{z}\{X=x \text{ e } Y=z-x\}\right) \\
&= \sum_{x=0}^{z}\mathbb{P}(X=x \text{ e } Y=z-x) \\
&= \sum_{x=0}^{z}\mathbb{P}(X=x)\mathbb{P}(Y=z-x) \\
&= \sum_{x=0}^{z}\left[e^{-\lambda_1}\frac{\lambda_1^x}{x!}\right]\left[e^{-\lambda_2}\frac{\lambda_2^{z-x}}{(z-x)!}\right] \\
&= e^{-(\lambda_1+\lambda_2)}\sum_{x=0}^{z}\frac{1}{x!(z-x)!}\lambda_1^x\lambda_2^{z-x} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}\sum_{x=0}^{z}\frac{z!}{x!(z-x)!}\lambda_1^x\lambda_2^{z-x} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}(\lambda_1+\lambda_2)^z,
\end{aligned}$$

onde na última igualdade utilizamos a expansão binomial. Reconhecemos essa fórmula como sendo a função de probabilidade de uma Poisson de parâmetro $\lambda_1+\lambda_2$, sendo essa a distribuição de X_1+X_2.

b) Utilizando a definição de probabilidade condicional, temos que:

$$\begin{aligned}
\mathbb{P}(X=x|X+Y=z) &= \frac{\mathbb{P}(X=x \text{ e } X+Y=z)}{\mathbb{P}(X+Y=z)} \\
&= \frac{\mathbb{P}(X=x \text{ e } Y=z-x)}{\mathbb{P}(X+Y=z)} \\
&= \frac{\mathbb{P}(X=x)\mathbb{P}(Y=z-x)}{\mathbb{P}(X+Y=z)} \\
&= \frac{\left[e^{-\lambda_1}\frac{\lambda_1^x}{x!}\right]\left[e^{-\lambda_2}\frac{\lambda_2^{z-x}}{(z-x)!}\right]}{\left[e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^z}{z!}\right]} \\
&= \frac{z!}{x!(z-x)!}\frac{\lambda_1^x\lambda_2^{z-x}}{(\lambda_1+\lambda_2)^{z\pm x}} \\
&= \frac{z!}{x!(z-x)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{z-x} \\
&= \frac{z!}{x!(z-x)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{z-x},
\end{aligned}$$

onde na segunda igualdade reescrevemos o evento $\{X = x \text{ e } X + Y = z\}$ como $\{X = x \text{ e } Y = z - x\}$, e na terceira igualdade usamos a independência entre X e Y. Note que reconhecemos tal expressão como a função de probabilidade de uma distribuição Binomial de parâmetros z e $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$.

c) *(Resposta livre, sendo esse um modelo)* Para ilustrar, sejam X e Y o número de pessoas que entram em determinada loja em um período de uma hora e que compram algum produto ou só "dão uma olhadinha", respectivamente. Dessa forma, $X + Y$ é o número total de pessoas que entra nessa loja nesse mesmo intervalo de tempo. Como X, Y, e $X + Y$ contam eventos da mesma natureza e estamos assumindo que X e Y são Poisson, parece razoável que $X + Y$ também o seja. Como os parâmetros de X e Y podem ser interpretados como a quantidade média de pessoas que entram na loja por hora, é razoável que sua soma seja o parâmetro de $X+Y$. Nesse mesmo contexto, o comerciante está interessado mais nas pessoas que de fato fazem alguma compra, sendo a entrada de tal cliente um evento que ele caracteriza como "sucesso". Assim, $X|X + Y$ conta o número de sucessos em $X + Y$ observações de indivíduos que entram na loja, de modo que a Binomial parece um modelo razoável. Note que o parâmetro $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ mede a proporção de pessoas que de fato compram algo, quantidade razoável de ser chamada de "probabilidade de sucesso".

Questão 2:

a) Note, primeiramente, que $[X]$ assume os valores $0, 1, 2, \ldots$. Portanto, para que tenhamos $[X] = n$ devemos ter $X \in [n, n + 1)$, para $n = 0, 1, 2, \ldots$. Portanto, temos que:

$$\begin{aligned}
\mathbb{P}([X] = n) &= \mathbb{P}(n \leq X < n + 1) \\
&= \int_n^{n+1} \lambda e^{-\lambda t}\, dt \\
&= -e^{-\lambda t}\Big|_{t=n}^{t=n+1} \\
&= -[e^{-\lambda(n+1)} - e^{-\lambda n}] \\
&= (e^{-\lambda})^n(1 - e^{-\lambda}),
\end{aligned}$$

para $n = 0, 1, 2, 3, \ldots$.

b) *(Resposta livre, sendo esse um modelo)* Note que o resultado encontrado no item a) é a função de probabilidade de uma distribuição Geométrica de parâmetro $p = 1 - e^{-\lambda}$. Portanto, $[X] \sim \text{Geo}(1 - e^{-\lambda})$. Se pensarmos que X é o tempo transcorrido entre duas ocorrências sucessivas de um evento de interesse, para que tenhamos a sua parte inteira igual a n, devemos ter a ocorrência de n intervalos sucessivos de uma unidade de tempo sem a ocorrência de tal evento, e isso tem probabilidade $e^{-\lambda n}$, sendo tal fato argumentado por uma integração ou pela propriedade da perda de memória da Exponencial. Seguidamente, no próximo intervalo de uma unidade de tempo necessariamente deverá ocorrer ao menos um evento de interesse, e a probabilidade disso ocorrer, novamente pela perda de memória, é dada por $\mathbb{P}(0 \leq X \leq 1)$, justamente o parâmetro p de $[X]$.

Questão 3:

a) Usando as propriedades vistas em aula, temos que:

$$\begin{aligned}
\mathbb{E}[\overline{X}_n] &= \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] \\
&= \frac{\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]}{n} \\
&= \frac{\mu + \cdots + \mu}{n} \\
&= \mu.
\end{aligned}$$

$$
\begin{aligned}
\mathbb{V}(\overline{X}_n) &= \mathbb{V}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{\mathbb{V}(X_1) + \cdots + \mathbb{E}(X_n)}{n^2} \\
&= \frac{\sigma^2 + \cdots + \sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}.
\end{aligned}
$$

A igualdade $\mathbb{E}[\overline{X}_n] = \mu$ significa que "em média, nosso estimador acerta". Ou seja, se a coleta de dados for feita diversas vezes, da mesma população e seguindo a mesma metodologia, apesar de possivelmente obtermos estimativas diferentes para μ, "em média" elas estarão certas. Do ponto de vista de $\mathbb{V}(\overline{X}_n)$, temos que tal quantidade se aproxima de 0 a medida que n cresce. Isso significa que "quanto mais observações, melhor", um comportamento esperado de um estimador. Portanto, essas propriedades respaldam $\overline{X}_n$ como um "bom" estimador para μ.

b) Pelo Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas, temos que:

$$
\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \approx N(0,1).
$$

Com isso, podemos aproximar a probabilidade de interesse conforme segue abaixo:

$$
\begin{aligned}
\mathbb{P}\left(|\overline{X}_n - \mu| > c\right) &= \mathbb{P}\left(\left|\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right| > \frac{c}{\sigma/\sqrt{n}}\right) \\
&= \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{c}{\sigma/\sqrt{n}}\right) + \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < -\frac{c}{\sigma/\sqrt{n}}\right) \\
&= \left[1 - \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{c}{\sigma/\sqrt{n}}\right)\right] + \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < -\frac{c}{\sigma/\sqrt{n}}\right) \\
&\approx \left[1 - \Phi\left(\frac{c}{\sigma/\sqrt{n}}\right)\right] + \Phi\left(-\frac{c}{\sigma/\sqrt{n}}\right) \\
&= 2\left[1 - \Phi\left(\frac{c}{\sigma/\sqrt{n}}\right)\right] \\
&= 2\left[1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right].
\end{aligned}
$$

c) Fazendo um paralelo com o que foi feito no item b), queremos encontrar o menor valor de n que satisfaça $\mathbb{P}\left(|\overline{X}_n - \mu| < 0{,}01\right) = 0{,}99$, ou equivalentemente, $\mathbb{P}\left(|\overline{X}_n - \mu| > 0{,}01\right) = 1 - 0{,}99 = 0{,}01$. Do item b), sabemos que a probabilidade $\mathbb{P}\left(|\overline{X}_n - \mu| > 0{,}01\right)$ é dada, aproximadamente, por $2\left[1 - \Phi\left(\frac{0{,}01\sqrt{n}}{1{,}1}\right)\right]$, e queremos que tal valor seja igual a 0,01. Resolvendo então a equação $2\left[1 - \Phi\left(\frac{0{,}01\sqrt{n}}{1{,}1}\right)\right] = 0{,}01$ para isolar Φ, temos que $\Phi\left(\frac{0{,}01\sqrt{n}}{1{,}1}\right) = 0{,}995$. Consultando a página do GeoGebra referenciada na dica, tal igualdade é satisfeita para $\frac{0{,}01\sqrt{n}}{1{,}1} = 2{,}5758$. Resolvendo para n, temos que um valor de 80.280,42, de modo que o tamanho mínimo da amostra deverá ser de 80.281 para satisfazer o requerimento desejado.

Questão 4: (QUESTÃO BÔNUS)

a) Primeiramente, note que a soma dos p_k sempre será 1, independente do valor de α:

$$\begin{aligned}\sum_{k=1}^{8} p_k &= \frac{8\alpha}{9} + 3\left(\frac{7-16\alpha}{27}\right) + 3\left(\frac{1+8\alpha}{27}\right) + \frac{1}{9} \\ &= \frac{24\alpha + 21 - 48\alpha + 3 + 24\alpha + 3}{27} \\ &= \frac{27}{27} \\ &= 1.\end{aligned}$$

Dessa forma, precisamos estudar quando que os p_k são não-negativos:

$$\begin{aligned}p_1 \geq 0 &\iff \alpha \geq 0 \\ p_2, p_3, p_4 \geq 0 &\iff \alpha \leq \frac{7}{16} \\ p_5, p_6, p_7 \geq 0 &\iff \alpha \geq -\frac{1}{8}.\end{aligned}$$

Portanto, o intervalo que faz com que todos ps p_k sejam não-negativos é o intervalo $\left[0, \frac{7}{16}\right]$. Como a questão versa sobre independência, é interessante excluir eventos de probabilidade zero, portanto o intervalo mais conveniente é $\left(0, \frac{7}{16}\right)$. Note que como o enunciado deste item pede somente um intervalo válido para α, ambas as respostas acima serão consideradas corretas.

b) Para termos independência, devemos ter $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$, para $1 \leq i < j \leq 3$ e $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$. Verifiquemos se tais condições são válidas. Primeiramente, encontremos quem são as interseções:

$$\begin{aligned}A_1 \cap A_2 &= \{5, 6, 8\} \\ A_1 \cap A_3 &= \{6, 8\} \\ A_2 \cap A_3 &= \{6, 8\} \\ A_1 \cap A_2 \cap A_3 &= \{6, 8\}.\end{aligned}$$

Calculando as probabilidades, temos que:

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= p_5 + p_6 + p_8 = \frac{5+16\alpha}{27} \\ \mathbb{P}(A_1 \cap A_3) &= p_6 + p_8 = \frac{4+8\alpha}{27} \\ \mathbb{P}(A_2 \cap A_3) &= p_6 + p_8 = \frac{4+8\alpha}{27} \\ \mathbb{P}(A_1 \cap A_2 \cap A_3) &= p_6 + p_8 = \frac{4+8\alpha}{27}.\end{aligned}$$

Finalmente, calculando as probabilidades individuais:

$$\begin{aligned}\mathbb{P}(A_1) &= p_2 + p_5 + p_6 + p_8 = \frac{4}{9} \\ \mathbb{P}(A_2) &= p_3 + p_5 + p_6 + p_8 = \frac{4}{9} \\ \mathbb{P}(A_3) &= p_4 + p_6 + p_7 + p_8 = \frac{4}{9}.\end{aligned}$$

Considerando, primeiramente, a igualdade $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$, devemos ter:

$$\frac{4+8\alpha}{27} = \left(\frac{4}{9}\right)^3 = \frac{64}{729} \Rightarrow 1 + 2\alpha = \frac{16}{27} \Rightarrow \alpha = -\frac{11}{54},$$

um valor que não é permitido para α, pois implicará valores de p_5, p_6 e p_7 negativos. Dessa forma, não há como os eventos A_1, A_2 e A_3 serem independentes. Note que não foram estudadas igualdades do tipo $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$, para $i \neq j$, pois o caso estudado acima já falha.