Estatística e Probabilidade - Avaliação Presencial 01 (Gabarito) - 2022/01

Prof. Hugo Carvalho

28/06/2022

Questão 1:

a) X e Y não são independentes, pois não há como fatorar $f(x, y)$ como produto de uma função que depende somente de x por uma função que depende somente de y: de fato, caso isso pudesse ser feito, teríamos um termo de produto cruzado xy que não existe na expressão dada na questão. Uma outar forma de mostrar isso envolveria calcular as distribuições marginais $f_X(x)$ e $f_Y(y)$ e verificar que o produto delas não é igual a $f(x, y)$.

b) Para calcular tal probabilidade, precisamos primeiro encontrar a distribuição condicional $f(x|y_0)$, que é dada pela expressão

$$f(x|y_0) = \frac{f(x, y_0)}{f_Y(y_0)}.$$

Comecemos calculando a distribuição marginal de Y:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\ dx \\ &= \int_0^1 \frac{2}{5}(2x + 3y)\ dx \\ &= \frac{2}{5}\left(x^2 + 3xy\right)\Big|_{x=0}^{x=1} \\ &= \frac{2}{5}(1 + 3y),\ \text{para } 0 \leq y \leq 1. \end{aligned}$$

Portanto, temos que

$$f(x|y_0) = \frac{\frac{2}{5}(2x + 3y_0)}{\frac{2}{5}(1 + 3y_0)},$$

que após substituirmos $y_0 = 2/3$ nos dá que

$$f(x|y_0 = 2/3) = \frac{1}{3}(2x + 2),\ \text{para } 0 \leq x \leq 1.$$

Dessa forma, a probabilidade desejada é dada por

$$\begin{aligned} \mathbb{P}((X|Y = 2/3) > 0{,}8) &= \int_{0{,}8}^{1} f(x|y_0 = 2/3)\ dx \\ &= \int_{0{,}8}^{1} \frac{1}{3}(2x + 2)\ dx \\ &= \frac{1}{3}(x^2 + 2x)\Big|_{x=0{,}8}^{x=1} \\ &\approx 0{,}253. \end{aligned}$$

Questão 2:

a) Podemos mencionar diversas hipóteses para a validade do modelo, dentre as quais: independência entre a chegada dos consumidores, visto que há um produto de probabilidades individuais no modelo proposto; em particular, não chegam dois consumidores precisamente "juntos" na loja; como não é especificada a hora em particular, é assumido que o comportamento dos consumidores é homogêneo ao longo do dia ou que o modelo é válido somente em determinados horários; o modelo impede que mais de 3.600 pessoas cheguem à loja em determinada hora. Frente a isso, tal modelo parece ser adequado para descrever a situação desejada pelo comerciante, porém pode-se pensar que alguns aspectos negativos graves dele são: a hipótese da independência (que pode ser sensivelmente amenizada aumentando-se o valor de n); dificuldade computacional de se calcular $n!$ e/ou potências de p e $1-p$. Dessa forma, apesar do modelo ser teoricamente adequado para a situação, do ponto de vista prático ele pode ser de difícil utilização, o que torna o seu custo maior que o benefício da modelagem. Como já vimos indícios acima, aumentar n terá o impacto da maior dificuldade de calcular $n!$, e apesar de tal ideia ser interessante para refinar o modelo, do ponto de vista computacional ela não se mostra muito adequada.

b) Podemos pensar, intuitivamente, que se n é "grande" e portanto p é "pequeno", estamos dividindo uma hora em intervalos contíguos muito pequenos e que em cada um desses intervalos há uma pequena probabilidade de chegar um consumidor. Dessa forma, se chegam x consumidores ao longo de uma hora, como uma baixa probabilidade está distribuída ao longo de diversos *slots*, não parece razoável assumir uma discrepância muito grande entre as probabilidades de chegada de x e $x+1$. Outra forma de enxergar isso é um argumento por contradição: se as probabilidades de chegada de x e $x+1$ consumidores não se relacionam de modo simples entre si e variam muito, isso não parece ser algo compatível com o cenário que desejamos modelar. Mostremos agora isso mais formalmente, como a questão pede:

$$\begin{aligned}
\frac{\mathbb{P}(X=x+1)}{\mathbb{P}(X=x)} &= \frac{\binom{n}{x+1}p^{x+1}(1-p)^{n-x-1}}{\binom{n}{x}p^x(1-p)^{n-x}} \\
&= \frac{\frac{n!}{(x+1)!(n-x-1)!}p^{x+1}(1-p)^{n-x-1}}{\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}} \\
&= \frac{\frac{n!}{(x+1)x!(n-x-1)!}p^x p(1-p)^{n-x-1}}{\frac{n!}{x!(n-x)(n-x-1)!}p^x(1-p)^{n-x-1}(1-p)} \\
&= \frac{n-x}{x+1}\frac{p}{(1-p)} \\
&\approx \frac{np}{x+1},
\end{aligned}$$

onde na última passagem usamos a aproximação $n-x \approx n$ (justificada pelo fato que usualmente teremos $x \ll n$) e que $1-p \approx 1$ (justificada pois $p \ll 1$).

c) Fazendo $x=0$, temos que:

$$\mathbb{P}(X=1) = \lambda\mathbb{P}(X=0).$$

Agora, fazendo $x=1$, temos:

$$\mathbb{P}(X=2) = \frac{\lambda}{2}\mathbb{P}(X=1) = \frac{\lambda^2}{2}\mathbb{P}(X=0).$$

Substituindo $x=2$, verificamos que:

$$\mathbb{P}(X=3) = \frac{\lambda}{3}\mathbb{P}(X=2) = \frac{\lambda^3}{3!}\mathbb{P}(X=0).$$

Indutivamente, vemos que vale a fórmula geral

$$\mathbb{P}(X=x) = \frac{\lambda^x}{x!}\mathbb{P}(X=0).$$

d) A soma de todas as probabilidades deve ser igual a 1. Portanto:

$$\begin{aligned}
1 &= \sum_{x=0}^{\infty}\mathbb{P}(X=x) \\
&+ \sum_{x=0}^{\infty}\frac{\lambda^x}{x!}\mathbb{P}(X=0) \\
&= \mathbb{P}(X=0)\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} \\
&= \mathbb{P}(X=0)e^{\lambda},
\end{aligned}$$

que nos dá diretamente que $\mathbb{P}(X=0)=e^{-\lambda}$. Dessa forma, o modelo de Poisson está completamente deduzido. Podemos apontar uma vantagem como o fato de ser computacionalmente mais plausível que o primeiro modelo proposto pelo comerciante, e duas desvantagens como sendo o fato de admitir uma quantidade irreal de pessoas entrando na loja no intervalo de q hora e uma certa homogeneidade da entrada de pessoas ao longo do tempo.

Questão 3:

a) Um elemento do evento $\{Y \leq t\}$ configura uma situação onde o tempo de chegada de dois consumidores consecutivos na loja é menor que ou igual à t. Como começou-se a contar o tempo a partir de $t_0 = 0$ e $t > 0$, isso significa que entrou ao menos uma pessoa na loja no intervalo de tempo $(0, t]$. Como a variável aleatória X_t conta exatamente a quantidade de consumidores que entrou na loja em tal intervalo de tempo, têm-se que esse elemento também está no evento $\{X_t \geq 1\}$. Reciprocamente, um elemento do evento $\{X_t \geq 1\}$ representa uma situação na qual entrou ao menos um consumidor na loja no intervalo de tempo $(0, t]$. Como chegou um consumidor exatamente no tempo $t_0 = 0$, isso significa que dois consumidores chegaram na loja em um intervalo de tempo de tamanho no máximo t, de modo que tal situação também está abarcada no evento $\{Y \leq t\}$.

b) A questão anterior nos indica que processos de contagem podem ser bem modelados por distribuições de Poisson, cujo parâmetro representa uma ocorrência média por unidade de tempo. Se interpretamos λ como a taxa média de chegada de consumidores no intervalo de tempo de uma hora, ao considerarmos um intervalo de t horas parece razoável admitir que a taxa de chegada de consumidores será de $t\lambda$. Assim, concluímos que X_t tem distribuição de Poisson (por também representar um processo de contagem) com parâmetro $t\lambda$ (pela discussão feita acima). Portanto, pelo resultado do item a), temos que:

$$\begin{aligned}\mathbb{P}(Y \leq t) &= \mathbb{P}(X_t \geq 1) \\ &= 1 - \mathbb{P}(X_t = 0) \\ &= 1 - e^{-\lambda t}.\end{aligned}$$

c) A função de probabilidade acumulada de Y, denotada por $F_Y(t)$, é dada por:

$$F_Y(t) = \mathbb{P}(Y \leq t),$$

por definição. Como tal quantidade foi calculada acima e é igual a $1 - e^{-\lambda t}$, para obtermos a função densidade de probabilidade de Y derivamos F_Y e obtemos que:

$$\begin{aligned}f_Y(t) &= F'_Y(t) \\ &= \frac{d}{dt} 1 - e^{-\lambda t} \\ &= \lambda e^{-\lambda t},\end{aligned}$$

para $t \geq 0$, e 0 caso contrário.