GABARITO - Primeira Prova
Fundamentos Estatísticos para Ciência dos Dados

16/04/2015

1. Qual das seguintes funções corresponde a uma função de distribuição acumulada (fda) de probabilidade? Para aquelas não são, explicar o que falha.

$$(A)\ F(x) = \begin{cases} 1 - x^2, & \text{se } x \in (0,1) \\ 0, & \text{caso contrário} \end{cases} \qquad (B)\ F(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1/2, & \text{se } 0 \leq x < 1 \\ 1, & \text{se } 1 \leq x < 2 \end{cases} \qquad (C)\ F(x) = \begin{cases} x^2, & \text{se } x \in (0,2) \\ 0, & \text{caso contrário} \end{cases}$$

RESPOSTA: A função (A) não é uma fda pois é decrescente. A função (B) satisfaz as condições para ser uma fda (não-decrescente e com limites em 0 e 1). E a terceira função, (C), assume valor maior que 1 portanto não é uma fda.

Avaliação: total de 3 pontos – 1 ponto por letra.

Comentário: Uma função de distribuição acumulada não precisa ter integral igual a 1. Logo, integrar as funções não era uma escolha correta.

---

2. A Figura 1 mostra uma rede com quatro nós e cinco arestas. Suponha que cada conexão falha com probabilidade 0.10 e que as falhas de conexões sejam eventos independentes uns dos outros. Calcule a probabilidade de que exista um caminho ativo de B para C.

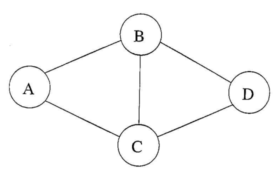

Figura 1: Rede com quatro nós.

RESPOSTA: Existem várias maneiras de resolver este problema. Uma solução simples é a seguinte: A probabilidade de existir um caminho ativo entre B e C é igual a 1 menos a probabilidade de NÃO existir um caminho ativo. Não existir uma caminho ativo significa todos os três caminhos possíveis, BAC, BC e BDC estejam simultaneamente fechados. Como estes três caminhos não possuem arestas comuns, eles abrem ou fecham de forma independente. Denotando $\sim AB$ pela negação do caminho AB estar aberto, temos:

$$\begin{aligned} \mathbb{P}(B \to C) &= 1 - \mathbb{P}(B \nrightarrow C) \\ &= 1 - \mathbb{P}(\sim BAC \cap \sim BC \cap \sim BDC) \\ &= 1 - \mathbb{P}(\sim BAC)\ \mathbb{P}(\sim BC)\ \mathbb{P}(\sim BDC) \\ &= 1 - (1 - \mathbb{P}(BAC))\ (1 - \mathbb{P}(BC))\ (1 - \mathbb{P}(BDC)) \\ &= 1 - (1 - 0.9^2)\ (1 - 0.9)\ (1 - 0.9^2) \\ &= 0.99639 \end{aligned}$$

Outra solução possível: Existem $2^5 = 32$ configurações para a rede. Listar as 9 configurações em que não há caminho ativo entre B e C indicando, para cada uma das arestas, se ela está aberta (1) ou fechada (0):

AB	AC	BC	BD	DC	Probab
0	0	0	0	0	0.1^5
0	1	0	0	0	$0.1^4\ 0.9$
1	0	0	0	0	$0.1^4\ 0.9$
0	0	0	1	0	$0.1^4\ 0.9$
0	0	0	0	1	$0.1^4\ 0.9$
1	0	0	1	0	$0.1^3\ 0.9^2$
0	1	0	1	0	$0.1^3\ 0.9^2$
1	0	0	0	1	$0.1^3\ 0.9^2$
0	1	0	0	1	$0.1^3\ 0.9^2$

Assim, a probabilidade de estar aberto é igual a $1 - 0.1^5 - 4 \times 0.1^4\ 0.9 - 4 \times 0.1^3\ 0.9^2 = 0.99639$.

Avaliação: total de 3 pontos.

Comentário: Calcular a probabilidade de haver caminho de B a C somente pela soma ou média das probabilidades dos caminhos BC, BAC e BDC estarem abertos, desprezando interseções está incorreto.

---

3. Um sistema de despacho dispara mensagens a cada 20 minutos a partir da meia noite. Uma requisição chega nos primeiros 20 minutos após a meia noite. Seja X o tempo aleatório da chegada desta requisição a partir da meia noite (em minutos). Sabe-se que a densidade de X é dada por

$$f(x) = \begin{cases} cx, & \text{se } x \in (0, 20) \\ 0, & \text{caso contrário} \end{cases}$$

Mostre que a constante c da densidade acima é igual a $1/200$. A seguir, obtenha o valor esperado do tempo transcorrido entre a chegada da requisição e a próxima mensagem a ser enviada.

RESPOSTA: Para encontrar a constante c:

$$1 = \int_0^{20} cx\ dx = c\left(\frac{x^2}{2}\ |_0^{20}\right) = c\frac{20^2}{2} = 200c$$

e portanto $c = 1/200$.

O tempo entre a chegada da requisição e o disparo da próxima mensagem após a meia-noite é igual a $20 - X$. O tempo esperado é

$$\mathbb{E}(20 - X) = 20 - \mathbb{E}(X) = 20 - \int_0^{20} x\ \frac{x}{200}\ dx = 20 - \frac{1}{200}\left(\frac{x^3}{3}\ |_0^{20}\right) = 6.667$$

Avaliação: total de 4 pontos – 2 pontos por subitem.

Comentário: Calcular somente a $\mathbb{E}(X)$ não responde corretamente o segundo subitem da questão.

---

4. Um dado bem equilibrado é lançado independentemente. Considera-se que ocorreu sucesso se sair a face 1 ou 2. O dado é lançado sucessivamente e de forma independente até que ocorra o SEGUNDO sucesso. Descreva o espaço amostral deste experimento e atribua probabilidades aos resultados possíveis.

RESPOSTA: O espaço amostral deste experimento é formado pelo número de vezes que o dado é lançado. O menor valor possível é 2, dois sucessos consecutivos. Logo $\Omega = \{2, 3, 4, ...\}$.

Seja S o evento SUCESSO e F o evento FRACASSO. $\mathbb{P}(S) = 2/6 = 1/3$ e $\mathbb{P}(F) = 4/6 = 2/3$.

A probabilidade de sucesso em 2 lançamentos é:
$\mathbb{P}(x = 2) = (1/3)^2$, somente se acontecer a sequência {SS};

O sucesso em 3 lançamentos pode vir de uma das sequências: {SFS, FSS}, logo:
$\mathbb{P}(x = 3) = 2(1/3)^2(2/3)$,

O sucesso em 4 lançamentos pode vir de uma das sequências {SFFS, FSFS, FFSS}, logo:
$\mathbb{P}(x = 4) = 3(1/3)^2(2/3)^2$

Portanto, o sucesso em n lançamentos pode vir de n-1 sequências diferentes:
$\mathbb{P}(x = n) = (n-1)(1/3)^2(2/3)^{n-2}$.

Avaliação: total de 3 pontos − 1 pontos para Ω e 2 pontos para $\mathbb{P}(x = n)$.

Comentário: Calcular a probabilidade $P(x = n)$ esquecendo do fator $(n-1)$ faz com que a soma infinita das probabilidades não totalize 1. Esse exercício não é exemplo de distribuição binomial, geométrica ou Bernoulli, mas sim binomial negativa.

---

5. A variável aleatória X assume valores no intervalo $[0, 1]$ e a sua função densidade de probabilidade é $f(x) = 2 - 2x$.

   - Encontre $\mathbb{E}(X)$ e $\mathbb{F}(x) = \mathbb{P}(X \leq x)$.
   - Gerei um número aleatório com distribuição UNIFORME em $(0, 1)$ e obtive $u = 0.135$. Use este valor gerado para obter um número aleatório em (0,1) gerado de acordo com a distribuição de X.

RESPOSTA: Temos

$$\mathbb{E}(X) = \int_0^1 x(2-2x)\ dx = 2\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{3}$$

Quanto à função de distribuição acumulada, temos

$$\mathbb{F}(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & \text{se } x \leq 0 \\ 2x - x^2 & \text{se } x \in (0,1) \\ 1, & \text{se } x \geq 1 \end{cases}$$

pois, para $x \in (0, 1)$,

$$\mathbb{F}(x) = \mathbb{P}(X \leq x) = \int_0^x (2-2t)\ dt = 2x - x^2$$

Quanto à simulação Monte Carlo, tomamos $X = \mathbb{F}^{-1}(U)$. Devemos encontrar a solução em x para a equação $2x - x^2 = u$ ou $x^2 - 2x + u$ onde $u \in (0, 1)$. Isto é um polinômio de segundo grau com soluções (ou raízes) $x_1 = 1 - \sqrt{4-4u}/2 \in (0,1)$ ou $x_1 = 1 + \sqrt{4-4u}/2 > 1$. Como X é uma v.a. com suporte em $(0, 1)$, só nos interessa a primeira solução. Isto é, $X = 1 - \sqrt{4-4U}/2$. Dessa forma, como o valor realizado de U é $u = 0.135$, temos $x = 1 - \sqrt{4-4(0.135)}/2 \approx 0.06995$.

Avaliação: total de 4 pontos.

Comentário: Observe a diferença entre $\mathbb{E}(x) = \int_{xmin}^{xmax} xf(x)dx$ e $\mathbb{F}(x) = \int_0^x f(x)dx$. O primeiro resulta em um número e o segundo em uma função de x.

---

6. Um sistema é formado por dois subsistemas A e B. Suponha que as seguintes probabilidades sejam conhecidas: $\mathbb{P}(A \text{ falhe}) = 0.20$, $\mathbb{P}(A \text{ e } B \text{ falhem}) = 0.15$, e $\mathbb{P}(B \text{ falhe SOZINHO}) = 0.15$.

   Usando estas três probabilidades obtenha $\mathbb{P}(B \text{ falhe})$. A seguir, obtenha $\mathbb{P}(A \text{ falhe } | B \text{ falhou})$ e $\mathbb{P}(A \text{ falhe SOZINHO})$.

RESPOSTA: Temos

$$\begin{aligned}
\mathbb{P}(B \text{ falhe}) &= \mathbb{P}(A \text{ e } B \text{ falhem}) + \mathbb{P}(A \text{ não falhe e } B \text{ falhe}) \\
&= 0.15 + \mathbb{P}(B \text{ falhe SOZINHO}) \\
&= 0.15 + 0.15 \\
&= 0.30
\end{aligned}$$

Agora

$$\mathbb{P}(A \text{ falhe } | B \text{ falhou}) = \frac{\mathbb{P}(A \text{ e } B \text{ falhem})}{\mathbb{P}(B \text{ falhe})} = \frac{0.15}{0.30} = 0.5$$

Finalmente,

$$\mathbb{P}(A \text{ falhe SOZINHO}) = \mathbb{P}(A \text{ falhe}) - \mathbb{P}(A \text{ e } B \text{ falhem}) = 0.20 - 0.15 = 0.05$$

Avaliação: total de 3 pontos – 1 ponto por probabilidade pedida.