Introdução à Probabilidade e à Estatística (BCN0406-15)

Prova 2 (A) – 16/08/2018 – Correção

(1.5pt) 1. Dadas as seguintes probabilidades associadas à variável aleatória X:

x	-1	1	2
$p(x)$	1/2	1/3	1/6

a) Calcule a esperança e a variância de X.
b) Calcule a esperança e a variância de X^3.

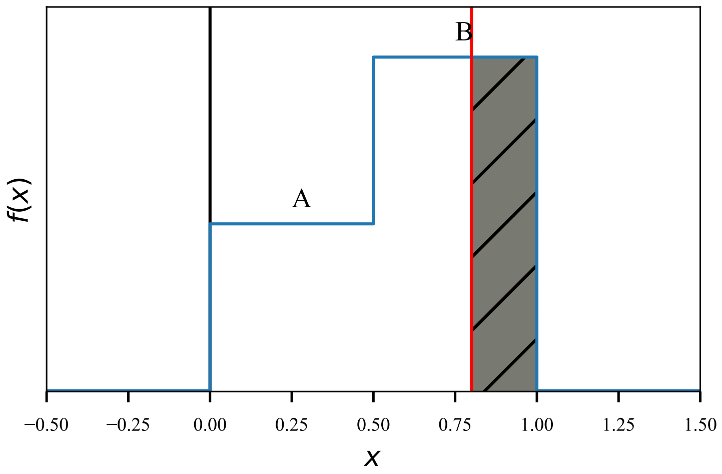
Solução
a) $E[X] = \sum_x xp(x) = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{6} = \frac{1}{6}$
$\text{Var}(X) = E[X^2] - (E[X])^2$, $\quad E[X^2] = \sum_x x^2 p(x) = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{6} = \frac{3}{2}$
$\text{Var}(X) = \frac{53}{36} - \left(\frac{1}{6}\right)^2 = \frac{53}{36}$
b) $E[X^3] = \sum_x x^3 p(x) = (-1)^3 \cdot \frac{1}{2} + 1^3 \cdot \frac{1}{3} + 2^3 \cdot \frac{1}{6} = \frac{7}{6}$
$\text{Var}(X^3) = E[(X^3)^2] - (E[X^3])^2$, $\quad E[X^6] = (-1)^6 \cdot \frac{1}{2} + 1^6 \cdot \frac{1}{3} + 2^6 \cdot \frac{1}{6} = \frac{23}{2}$
$\text{Var}(X) = \frac{23}{2} - \left(\frac{7}{6}\right)^2 = \frac{365}{36}$

(1.5pt) 2. Queremos fazer uma pesquisa com estudantes dos BI's da UFABC. Cerca de 70% deles são do BC&T (logo 30% são do BC&H). Ao selecionarmos 10 estudantes ao acaso, qual a probabilidade de encontrarmos exatamente 5 de cada curso? E qual a probabilidade de encontrarmos pelo menos 2 do BC&H?

Solução
Vamos chamar de X o número de estudantes do BC&H selecionados. A probabilidade de X é resultado de $n = 10$ escolhas independentes, cada uma com probabilidade de sucesso de $p = 0.3$, logo X é uma variável aleatória *binomial*. Queremos encontrar $P(X = 5)$, e também $P(X \geq 2)$, e para tanto basta usar a função de probabilidade da binomial:
$P(X = 5) = \binom{10}{5} \cdot 0.3^5 \cdot (1 - 0.3)^5 \approx 0.103$
$P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - \binom{10}{0} \cdot 0.3^0 \cdot (1 - 0.3)^{10} - \binom{10}{1} \cdot 0.3^1 \cdot (1 - 0.3)^9 \approx 0.851$

(2pt) 3. Escolhemos aleatoriamente um número real a entre 0 e 1/2, e um número real b entre 1/2 e 1. Finalmente, escolhemos X igual a a com probabilidade 1/3, ou igual a b com probabilidade 2/3.
a) Encontre a função densidade de probabilidade da variável aleatória X. Esboce o gráfico dessa função.
b) Calcule $P(X > 0.8)$. Para tanto, represente essa probabilidade graficamente no seu esboço acima.
c) Esboce o gráfico da função distribuição acumulada de X, $F(x)$.

Solução
a) X é uma variável aleatória contínua entre 0 e 1. Ela é aleatória (tem densidade de probabilidade constante) dentro do intervalo $[0; 1/2]$, e também dentro do intervalo $[1/2; 1]$, logo o gráfico da função densidade de probabilidade $f(x)$ é zero pra $x < 0$ ou $x > 1$, e é constante em cada um dos dois intervalos, com alturas dadas por A e B (ver gráfico). Agora precisamos calcular A e B.

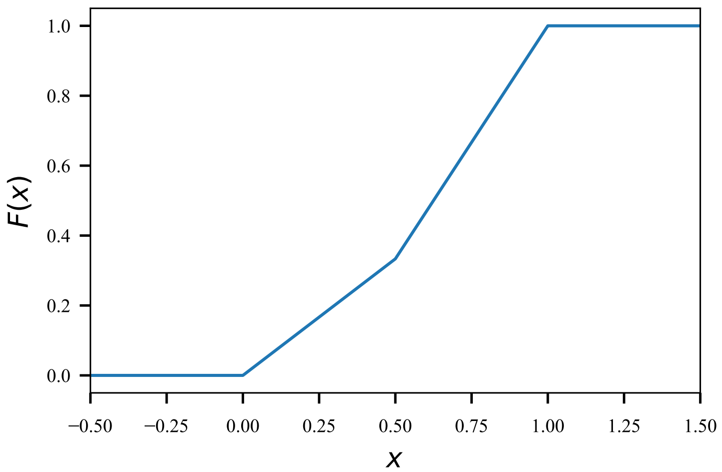


Para isso, notamos que a probabilidade de sair um número a entre 0 e 1/2 é 1/3, logo a área do primeiro retângulo deve ser 1/3. Essa área é dada por $0.5 \cdot A = \frac{1}{3}$, logo $A = \frac{2}{3}$. Da mesma forma para o segundo intervalo, teremos $0.5 \cdot B = \frac{2}{3}$, logo $B = \frac{4}{3}$. Com isso podemos escrever a $f(x)$ como:

$$f(x) = \begin{cases} \frac{2}{3}, \ 0 \leq x < \frac{1}{2} \\ \frac{4}{3}, \ \frac{1}{2} \leq x < 1 \\ 0, \ x < 0 \text{ ou } x > 1 \end{cases}$$

b) A probabilidade de X ser maior que 0.8 corresponde à área hachurada no gráfico acima, ou $\int_{0.8}^{\infty} f(x)dx$. Sendo um retângulo, calculamos essa área multiplicando os comprimentos dos lados: $P(X > 0.8) = (1 - 0.8) \cdot \frac{4}{3} = \frac{0.8}{3} \approx 0.267$.

c) A função de distribuição acumulada é $F(x) = P(X \leq x)$. Pra $x \leq 0$, ela é zero, e pra $x \geq 1$ ela vale 1 (já que X é no máximo 1). F deve crescer linearmente em cada um dos dois intervalos (da mesma maneira que uma variável uniforme), e, além disso, vale 1/3 quando $x = 1/2$. Com essas informações, chegamos ao gráfico esboçado abaixo.



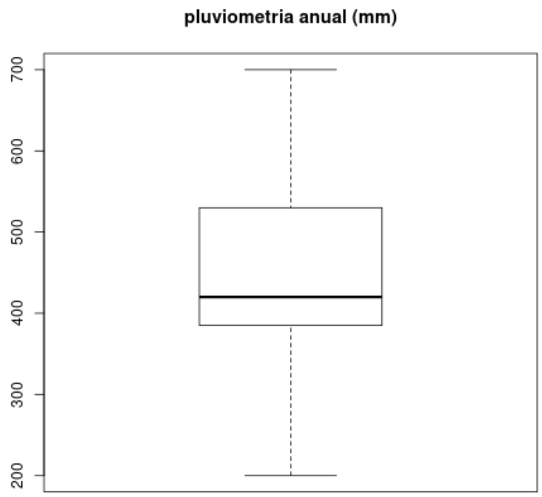
(2pt) 4. Uma escola de futebol mede o tempo que cada adolescente leva para correr 100m, que é modelado como uma variável aleatória normalmente distribuída com esperança 20s e desvio padrão de 4s. Qual é a probabilidade de encontrar alguém que corra 100m em menos de 12s?

Solução

O tempo de corrida é representado por uma variável aleatória contínua X normal com média $\mu = 20$ e desvio padrão $\sigma = 4$. Queremos encontrar a probabilidade de encontrar alguém com tempo de corrida menor que 12s, ou seja, $P(X < 12)$. Essa probabilidade é dada exatamente pela função distribuição acumulada nesse ponto, $F(12)$. Para obtê-la, calculamos a variável $Z = \frac{X-\mu}{\sigma}$ e usamos a tabela para encontrar esse valor. Neste caso, temos $F(12) = \Phi\left(\frac{12-20}{4}\right) = \Phi(-2)$. Na tabela temos apenas valores negativos de z, então usamos a igualdade $\Phi(-z) = 1 - \Phi(z)$, obtidade do fato de a densidade de probabilidade da distribuição normal padrão ser simétrica, chegando a $F(12) = \Phi(-2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228$.

(1.5pt) 5. Obtivemos os seguintes dados de pluviometria anual na região do Cantareira (em mm), num período de 12 anos: 380, 400, 480, 440, 390, 510, 550, 700, 580, 390, 380, 200.
a) Calcule a média e a mediana dos dados.
b) Encontre os quartis e esboce o *boxplot* (ou diagrama de caixa) que representa esses dados.

Solução
a) Média $= \frac{380+400+480+440+390+510+550+700+580+390+380+200}{12} = 450$.
A mediana é o valor central dos dados. Os dois pontos do meio são 400 e 440, logo a mediana é o valor médio desses 2 pontos, 420.
b) O segundo quartil, $Q2$, é a própria mediana. O primeiro e o terceiro quartis contêm, respectivamente, 25 e 75% dos dados – neste caso, 3 e 9 do total de 12 dados, ordenados em ordem crescente. Portanto podemos tomar o Q1 entre 380 e 390 e o Q3 entre 510 e 550: $Q1 = 385$ e $Q3 = 530$. O boxplot abaixo resume essa informação traçando uma caixa delimitada pelo Q1 e Q3, com a mediana no meio, e uma linha ligando a caixa aos valores mínimo e máximo encontrados (200 e 700, respectivamente).



(2pt) 6. O consumo de banda de internet mensal por pessoa em uma população foi modelado como uma variável uniforme de média 700M e desvio padrão de 500M. Para planejar o cabeamento de uma região com 10000 pessoas, uma empresa precisa estimar qual a banda total a ser alocada. Para tanto, ela precisa saber qual será a média de consumo dessa população com probabilidade de 95%. Calcule, aproximadamente, o intervalo de confiança, ou seja, o intervalo de valores ao redor da média, que deverá acontecer com probabilidade de 95%, usando o Teorema do Limite Central. Compare esse intervalo com os valores a mais ou menos um desvio padrão de consumo por pessoa e explique a discrepância entre eles.

Solução
De acordo com o Teorema do Limite Central, a média de n variáveis aleatórias independentes e identicamente distribuídas de esperança μ e desvio padrão σ tende a uma normal com média μ e desvio padrão $\sigma_n = \frac{\sigma}{\sqrt{n}}$ quando n tende a infinito. Neste caso, temos $\mu = 700$, $\sigma = 500$ e $n = 10000$, logo a média será uma normal de média 700 e desvio padrão $\frac{500}{\sqrt{10000}} = 5$. O intervalo de confiança de 95% é dado pela região $[\mu - s; \mu + s]$ tal que $P(\mu - s < X < \mu + s) = 0.95$. Essa probabilidade é calculada da mesma forma que fizemos na questão 4:

$$P(\mu-s < X < \mu+s) = F(\mu+s) - F(\mu-s) = \Phi\left(\frac{700 + s - 700}{5}\right) - \Phi\left(\frac{700 - s - 700}{5}\right) =$$
$$= \Phi\left(\frac{s}{5}\right) - \Phi\left(\frac{-s}{5}\right) = \Phi\left(\frac{s}{5}\right) - \left(1 - \Phi\left(\frac{s}{5}\right)\right) = 2\Phi\left(\frac{s}{5}\right) - 1 = 0.95\,. \quad (1)$$

Logo $\Phi\left(\frac{s}{5}\right) = (0.95 + 1) \cdot 2 = 0.975$. Pela tabela, vemos que o valor de z mais próximo de $\Phi(z) = 0.975$ é $z = 2$, portanto $\frac{s}{5} \approx 2$, e $s = 10$. Com isso, obtemos que o IC de 95% é $[690; 710]$.

(2pt) 7. Em cada item abaixo, leia atentamente e classifique a afirmação como verdadeira (V) ou falsa (F), justificando brevemente sua opção. Opções sem justificativa não serão aceitas.

(a) Uma variável aleatória X tem esperança maior que uma outra variável aleatória Y, logo podemos afirmar que X sempre assume um valor maior que Y.

(b) A probabilidade de que um evento raro ocorra k vezes em um certo intervalo de tempo é geralmente bem descrita por uma variável aleatória de Poisson.

(c) Podemos resumir dados qualitativos obtidos a partir de uma amostra calculando sua média e seu desvio padrão.

(d) A função de distribuição acumulada $F(x)$ de uma variável aleatória discreta X tende a 1 para valores de x tendendo a infinito, mas o mesmo não é verdade para uma variável aleatória contínua.

(e) A mediana de uma amostra não é afetada quando aumentamos o valor do maior dos dados obtidos; a média dessa amostra, sim, aumenta.

Solução
a) Falsa. A esperança é uma média ponderada dos valores possíveis, mas um valor esperado maior de X não garante que todos os valores possíveis sejam maiores. Por exemplo, se X pode ser 0 ou 1 com probabilidade de 1/2 pra cada, $E[X] = 0.5$, enquanto Y pode ser -2 ou 2 com probabilidade de 1/2 pra cada, $E[Y] = 0$, mas X pode assumir o valor 0 e Y o valor 2, sendo maior nesse caso.
b) Verdadeira. A descrição dada corresponde exatamente ao uso mais comum da variável de Poisson: o número de eventos que acontecem em um intervalo (de tempo ou espaço) fixo – X é discreta, acontece a uma taxa λ e é rara o suficiente para que seja improvável que 2 eventos aconteçam simultaneamente. c) Falsa. A média e o desvio padrão só podem ser calculados para dados quantitativos. Dados qualitativos (como cores, marcas etc.) não admitem operações aritméticas.
d) Falsa. Quando x tende a infinito, ele tende a ser maior que todos os valores possíveis de

uma variável aleatória – discreta ou contínua. Como $F(x) = P(X \leq x)$, essa probabilidade tende a 1, independente do tipo de variável.
e) Verdadeira. Quando aumentamos apenas o maior valor, o valor dos dados centrais da amostra não é alterado, logo a mediana não se altera. Já a média é calculada somando todos os valores, e portanto aumenta. *Nota:* se a amostra tem apenas 1 ou 2 valores, a mediana também aumenta.