FICHA 05 – resolução

DPC

1. Seja X uma variável aleatória contínua, com função densidade de probabilidade dada por

$$f(x) = \begin{cases} ax, & 0 \le x < 1 \\ a, & 1 \le x < 2 \\ -ax + 3a, & 2 \le x < 3 \\ 0, & \text{outros valores de } x \end{cases}$$

(a) Determine a constante a.

(b) Determine a função de distribuição acumulada e esboce o seu gráfico.

(c) Se X_1, X_2 e X_3 forem três observações independentes de X, qual será a probabilidade de exactamente um desses três números ser maior do que 1.5?

Resolução:

(a) Se f é uma densidade de probabilidade, então tem de satisfazer $f(x) \ge 0,\ \forall x \in \mathbf{R}$, e $\int_{-\infty}^{\infty} f(x)\,dx = 1$. A segunda condição permite-nos determinar o valor de a:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Leftrightarrow \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (-ax+3a)\,dx = 1 \Leftrightarrow a = \frac{1}{2}.$$

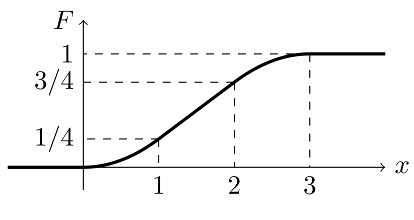
(b) A função de distribuição F é dada por:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x')\,dx'$$

$$= \begin{cases} 0, & x \le 0 \\ \int_0^x x'/2\,dx', & 0 \le x \le 1 \\ \int_0^1 x'/2\,dx' + \int_1^x 1/2\,dx', & 1 \le x \le 2 \\ \int_0^1 x'/2\,dx' + \int_1^2 1/2\,dx' + \int_2^x (-x'/2 + 3/2)\,dx'\ , & 2 \le x \le 3 \\ 1, & x \ge 3 \end{cases}$$

$$= \begin{cases} 0, & x \le 0 \\ x^2/4, & 0 \le x \le 1 \\ x/2 - 1/4, & 1 \le x \le 2 \\ -x^2/4 + 3x/2 - 5/4, & 2 \le x \le 3 \\ 1, & x \ge 3 \end{cases}$$

Note que f é descontínua, mas, ainda assim, F é contínua. A continuidade de F é óbvia observando o respetivo gráfico:



(c) A probabilidade pedida é:

$$\begin{aligned}\text{Prob(uma observ.} > 1.5) &= P(X_1 > 1.5 \wedge X_2 \leq 1.5 \wedge X_3 \leq 1.5) \\ &\quad + P(X_1 \leq 1.5 \wedge X_2 > 1.5 \wedge X_3 \leq 1.5) \\ &\quad + P(X_1 \leq 1.5 \wedge X_2 \leq 1.5 \wedge X_3 > 1.5) \\ &= P(X_1 > 1.5)P(X_2 \leq 1.5)P(X_3 \leq 1.5) \\ &\quad + P(X_1 \leq 1.5)P(X_2 > 1.5)P(X_3 \leq 1.5) \qquad \text{(v.a. indep.)} \\ &\quad + P(X_1 \leq 1.5)P(X_2 \leq 1.5)P(X_3 > 1.5).\end{aligned}$$

Como X_1, X_2 e X_3 são observações de X, a sua distribuição é igual à de X, pelo que $P(X_i \leq x) = F(x)$ e $P(X_i > x) = 1 - F(x)$, logo:

$$\text{Prob(uma observ.} > 1.5) = 3F(1.5)^2(1 - F(1.5)) = \frac{3}{8}.$$

Nota: Observe que, dadas n observações independentes de uma v.a. X com função de distribuição acumulada F_X e dado um x fixo, a probabilidade de que exatamente k dessas n observações sejam maiores do que x é dada por uma distribuição binomial (na variável k):

$$\text{Prob}(k \text{ observ.} > x) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \cdots, n\},$$

onde $p = P(X > x) = 1 - F_X(x)$. O exercício que acabámos de resolver é apenas um caso particular (onde $n = 3$, $k = 1$, $x = 1.5$ e $F_X = F$, definida acima).

2. Considere a variável aleatória X com função densidade de probabilidade

$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ x, & 0 < x \leq 1 \\ -x + 2, & 1 < x \leq 2 \\ 0, & x > 2 \end{cases}$$

(a) Calcule a média, a moda e a mediana de X.

(b) A distribuição será simétrica? Justifique.

Resolução:

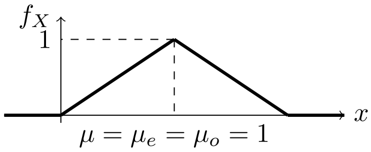
(a) Vamos calcular a média (μ), a mediana (μ_e) e a moda (μ_o) de X:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x^2\,dx + \int_1^2 x(-x + 2)\,dx = 1,$$

$$\begin{aligned}\mu_e : \quad & P(X \leq \mu_e) = \frac{1}{2} \Leftrightarrow F_X(\mu_e) = \frac{1}{2} \\ & \Leftrightarrow \left(\int_0^{\mu_e} x\,dx = \frac{1}{2} \wedge 0 < \mu_e \leq 1\right) \vee \left(\int_0^1 x\,dx + \int_1^{\mu_e} (-x+2)\,dx = \frac{1}{2} \wedge 1 < \mu_e \leq 2\right) \\ & \Leftrightarrow \left(\frac{\mu_e^2}{2} = \frac{1}{2} \wedge 0 < \mu_e \leq 1\right) \vee \left(\frac{1}{2} - \frac{\mu_e^2}{2} + 2\mu_e + \frac{1}{2} - 2 = \frac{1}{2} \wedge 1 < \mu_e \leq 2\right) \\ & \Leftrightarrow (\mu_e = \pm 1 \wedge 0 < \mu_e \leq 1) \vee \left(-\frac{\mu_e^2}{2} + 2\mu_e - \frac{3}{2} = 0 \wedge 1 < \mu_e \leq 2\right) \\ & \Leftrightarrow (\mu_e = \pm 1 \wedge 0 < \mu_e \leq 1) \vee ((\mu_e = 1 \vee \mu_e = 3) \wedge 1 < \mu_e \leq 2) \\ & \Leftrightarrow \mu_e = 1,\end{aligned}$$

$$\mu_o = \operatorname*{argmax}_x f_X(x) = 1.$$

(b) Qualquer distribuição onde a mediana e a média (se existir) são iguais é simétrica em torno da média. Podemos desenhar o gráfico de f_X para confirmar visualmente a simetria em torno do eixo $x = 1$:



3. Um dado alarme contra ladrões dispara se a tensão na sua entrada exceder $5v$ em 3 instantes consecutivos de amostragem. Se as amostras de tensão forem independentes e uniformemente distribuidas em $[0,\ 7]$, qual a probabilidade do alarme disparar?

   Resolução:
   $X-$ tensão na entrada do alarme, num dado instante. $X \sim U(0,7) \Leftrightarrow f_X(x) = \frac{1}{7-0} = 1/7,\ x \in [0,7]$. Temos três observações de X independentes (X_1, X_2 e X_3). O alarme dispara se todas elas excederem 5V, logo:

$$\begin{aligned} \text{Prob(alarme disparar)} &= P(X_1 > 5 \wedge X_2 > 5 \wedge X_3 > 5) = P(X > 5)^3 \\ &= \left(\int_5^7 \frac{1}{7}\, dx \right)^3 = \left(\frac{2}{7} \right)^3. \end{aligned}$$

4. O tempo de vida de uma máquina tem uma distribuição contínua no intervalo $[0, 40]$ com uma função densidade de probabilidade f, onde $f(x)$ é proporcional a $(x + 10)^{-2}$. Qual a probabilidade que o tempo de vida da máquina seja inferior a 5?

   Resolução:
   Para $x \in [0, 40]$, temos $f(x) \propto (x + 10)^{-2} \Leftrightarrow f(x) = \frac{a}{(x+10)^2}$, onde a é uma constante que temos de determinar. Sabemos que f é uma densidade de probabilidade, logo:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \Leftrightarrow \int_0^{40} \frac{a}{(x+10)^2}\, dx = 1 \Leftrightarrow a = \frac{25}{2}.$$

   Conhecido o valor de a, podemos então calcular a probabilidade pedida:

$$\text{Prob(tempo de vida} < 5) = \int_{-\infty}^{5} f(x)\, dx = \int_0^5 \frac{25}{2(x+10)^2}\, dx = \frac{5}{12}.$$

5. Suponha que a v.a. contínua X tem a função densidade de probabilidade $f_X(x) = 2xe^{-x^2}$, $x \geq 0$. Seja $Y = X^2$. Calcule $E(Y)$ sem obter a função densidade de probabilidade de Y.

   Resolução:
   Vamos utilizar um resultado que já tínhamos visto anteriormente (ex. 9 da Ficha 3) para o caso de v.a. discretas e que se mantém válido no caso contínuo:

$$Y = g(X) \implies E_{f_Y}(Y) = E_{f_X}\left[g(X)\right].$$

   No presente exercício, temos $Y = X^2$, ou seja, $g(x) = x^2$. Assim,

$$\begin{aligned} E_{f_Y}(Y) = E_{f_X}\left[g(X)\right] &= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \\ &= \int_0^{\infty} x^2 2xe^{-x^2}\, dx = 1. \qquad \text{(int. p/ partes: } u' = 2xe^{-x^2},\ v = x^2) \end{aligned}$$

6. O diâmetro de um cabo eléctrico é normalmente distribuído com média 0.8 e variância 0.0004.

   (a) Qual a probabilidade de que o diâmetro ultrapasse 0.81?

   (b) O cabo é considerado defeituoso se o diâmetro diferir de sua média em mais de 0.025. Qual a probabilidade de se encontrar um cabo defeituoso?

Resolução:
Se X é uma v.a. que segue uma distribuição normal com média μ e variância σ^2, então a sua função densidade de probabilidade é:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \ x \in \mathbf{R}.$$

Infelizmente, não é possível calcular analiticamente o integral da função f_X, ou seja, não conseguimos escrever uma expressão utilizando funções "comuns"(polinomiais, exponenciais, logarítmicas, trigonométricas, etc.) que seja uma primitiva de f_X. Assim, para calcular probabilidades do tipo $P(X \leq x)$ (ou $P(X \geq x)$) para um dado x, temos de recorrer a software de cálculo numérico (máquina calculadora, Python, R, Octave, Matlab, etc.) ou utilizar uma tabela da distribuição normal.

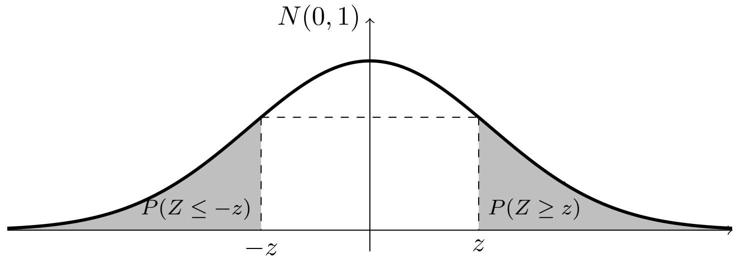
A tabela considera uma v.a. Z com distribuição normal padrão (isto é, com $\mu = 0$ e $\sigma^2 = 1$) e, geralmente, apresenta os valores de $P(Z \leq z)$ para vários $z < 0$. Assim, quando queremos calcular $P(X \leq x)$, temos de, em primeiro lugar, transformar X numa v.a. com distribuição normal padrão e, em seguida, podemos ter de utilizar as propriedades de simetria da distribuição.

Padronização:
É fácil verificar (tente!) que:

$$X \sim N(\mu, \sigma^2) \implies Z = \frac{X-\mu}{\sigma} \sim N(0,1),$$

e assim $P(X \leq x) = P(Z \leq (x-\mu)/\sigma)$ pode ser obtida a partir da tabela da distribuição normal padrão.

Simetria:
Uma distribuição $N(\mu, \sigma^2)$ é simétrica em torno de μ, pelo que $N(0,1)$ é simétrica em torno de zero. Assim, $P(Z \leq -z) = P(Z \geq z)$ para todo o $z \in \mathbf{R}$. O gráfico seguinte ilustra esta propriedade:



(a) $X \sim N(0.8, 0.0004)-$ diâmetro do cabo elétrico.

$$P(X > 0.81) \underset{\text{padronização}}{=} P\Big(\underset{N(0,1)}{\underset{\wr}{Z}} > \frac{0.81-0.8}{\sqrt{0.0004}}\Big) = P(Z > 0.5) \underset{\text{simetria}}{=} P(Z < -0.5) \underset{\text{tabela}}{\approx} 0.309.$$

(b)

$$\begin{aligned}
\text{Prob(cabo defeituoso)} &= P(|X-0.8| > 0.025) = P(X-0.8 > 0.025) + P(X-0.8 < -0.025) \\
&= P\left(\frac{X-0.8}{\sqrt{0.0004}} > \frac{0.025}{\sqrt{0.0004}}\right) + P\left(\frac{X-0.8}{\sqrt{0.0004}} < -\frac{0.025}{\sqrt{0.0004}}\right) \\
&= P(Z > 1.25) + P(Z < -1.25) = 2P(Z < -1.25) \approx 0.211.
\end{aligned}$$

7. Suponha-se que a duração da vida de dois dispositivos electrónicos D_1 e D_2 tenham distribuições $N(49, 36)$ e $N(45, 9)$, respectivamente. Se o dispositivo electrónico tiver de ser usado por um período de 45 horas, qual dos dois dispositivos deve ser preferido? Se tiver de ser usado por um período de 40 horas, qual deles deve ser preferido?

Resolução:
$P(D_1 \geq 45) \approx 0.748$, $P(D_2 \geq 45) = 0.5 < P(D_1 \geq 45)$, logo, se o período de utilização for de 45 horas, preferimos o dispositivo 1. $P(D_1 \geq 40) \approx 0.933$, $P(D_2 \geq 40) \approx 0.952 > P(D_1 \geq 40)$, logo, se o período de utilização for de 40 horas, preferimos o dispositivo 2.

8. Suponha que X tenha distribuição $N(\mu, \sigma^2)$. Determine c (como função de μ e σ) tal que $P(X \leq c) = 2P(X > c)$.

Resolução:

$$\begin{aligned} c: \quad P(X \leq c) = 2P(X > c) &\Leftrightarrow P\left(Z \leq \frac{c-\mu}{\sigma}\right) = 2P\left(Z > \frac{c-\mu}{\sigma}\right) \\ &\Leftrightarrow P\left(Z \leq \frac{c-\mu}{\sigma}\right) = 2\left(1 - P\left(Z \leq \frac{c-\mu}{\sigma}\right)\right) \\ &\Leftrightarrow P\left(Z \leq \frac{c-\mu}{\sigma}\right) = \frac{2}{3} \\ &\Leftrightarrow \frac{c-\mu}{\sigma} \approx 0.431 \Leftrightarrow c \approx 0.431\sigma + \mu. \end{aligned}$$

9. Suponha que X, a carga de ruptura de um cabo (em kg), tenha distribuição $N(100, 16)$. Cada rolo de 100 metros de cabo dá um lucro de 25 Eur desde que $X \geq 95$. Se $X < 95$, o cabo será vendido para uma finalidade diferente e com um lucro de 10 Euro. Determine o lucro esperado por rolo.

Resolução:
Seja L a v.a. que representa o lucro obtido, em euros. Sabemos que L é dado, em função de X, por:

$$L = g(X) = \begin{cases} 10, & \text{se } X < 95, \\ 25, & \text{se } X \geq 95. \end{cases}$$

Note que, apesar de X ser uma v.a. contínua, L é uma v.a. discreta. Assim,

$$E(L) = \sum_{l \in \{10,25\}} l f_L(l) = 10 f_L(10) + 25 f_L(25) = 10P(X < 95) + 25P(X \geq 95) \approx 23.4 \text{ Eur.}$$

10. Seja X_t o número de partículas emitidas em t horas por uma fonte radioactiva e suponha-se que X_t tenha uma distribuição de Poisson, com parâmetro βt. Faça-se igual a T o número de horas entre emissões sucessivas. Se $\beta = 30$, qual a probabilidade de que o tempo entre duas emissões seja maior do que 5 min? E maior do que 10 min? Menor do que 30 segundos?

Resolução:
Note que o número de partículas emitidas é uma v.a. discreta ($X_t \sim \text{Pois}(\beta t)$), mas o tempo T entre duas emissões sucessivas é uma v.a. contínua. Vamos determinar a distribuição de T a partir da função de probabilidade de X_t, que conhecemos. Se em t horas não houve qualquer partícula emitida, então temos necessariamente $X_t = 0$ e $T > t$, assim:

$$P(T > t) = P(X_t = 0) = \frac{e^{-\beta t}(\beta t)^0}{0!} = e^{-\beta t} \Leftrightarrow P(T \leq t) = 1 - e^{-\beta t}, \quad t \geq 0.$$

$$\beta = 30 \implies P(T > 5/60) = e^{-2.5},\ P(T > 10/60) = e^{-5},\ P(T \leq 30/60^2) = 1 - e^{-0.25}.$$

<u>Nota:</u> A distribuição de T a que chegámos é conhecida por *distribuição exponencial* e, como vimos, modela o tempo entre duas ocorrências sucessivas de dado fenómeno quando o número de ocorrências deste segue uma distribuição de Poisson. A função densidade de probabilidade da distribuição exponencial é:

$$f_T(t) = \frac{d}{dt}[P(T \leq t)] = \beta e^{-\beta t}, \quad t \geq 0.$$

11. O Professor Joaquim ajusta sempre as notas dos seus alunos. Ele adiciona um número c a cada teste de forma a que 50% dos alunos tenham entre 70 e 80. Se as notas seguem uma distribuição $N(65, 38)$, qual o valor de c a usar? Nota: há duas soluções mas os alunos preferem uma delas.

Resolução:

$X \sim N(65, 38)$

$$\begin{aligned} c: \quad & P(70 \leq X + c \leq 80) = 0.50 \\ \Leftrightarrow & P\left(\frac{5-c}{\sqrt{38}} \leq Z \leq \frac{15-c}{\sqrt{38}}\right) = 0.50, \quad Z \sim N(0,1) \\ \Leftrightarrow & P\left(k \leq Z \leq k + \frac{10}{\sqrt{38}}\right) = 0.50, \quad k = \frac{5-c}{\sqrt{38}} \\ \Leftrightarrow & F_Z\left(k + \frac{10}{\sqrt{38}}\right) - F_Z(k) = 0.50, \end{aligned}$$

onde F_Z é a função de distribuição acumulada da distribuição normal padrão. Esta equação pode ser resolvida numericamente recorrendo a software de cálculo, obtendo-se duas soluções:

$$k \approx -1.429 \vee k \approx -0.193 \implies c \approx 13.81 \vee c \approx 6.19.$$

Obviamente, os alunos preferem a solução $c \approx 13.81$.

12. Compare o limite superior da probabilidade $P[|X - E[X]| \geq 2\sqrt{V(X)}]$, obtida pela desigualdade de Chebyshev, com a probabilidade exacta se X for uniformemente distribuida em $[-1, 3]$.

    Resolução:
    Desigualdade de Chebyshev: Seja X uma v.a. com valor esperado $E(X)$ finito e variância $V(X)$ finita e não nula. Então, para qualquer número real $k > 0$,

    $$P\left(|X - E(X)| \geq k\sqrt{V(X)}\right) \leq \frac{1}{k^2}.$$

    (Note que este majorante é trivial para $k \leq 1$.)

    Sabemos que $X \sim U(-1, 3)$, logo $f_X(x) = 1/4$, $x \in [-1, 3]$, e, assim:

    $$E(X) = \int_{-1}^{3} \frac{x}{4}\, dx = 1, \quad V(X) = \int_{-1}^{3} \frac{(x-1)^2}{4}\, dx = \frac{4}{3}.$$

    Utilizando a desigualdade de Chebyshev (com $k = 2$), temos $P(|X-1| \geq 4/\sqrt{3}) \leq 1/4$. Como, neste caso, a distribuição de X é conhecida, podemos calcular o valor exato da probabilidade (que terá de ser menor ou igual ao majorante obtido):

    $$\begin{aligned} P\left(|X-1| \geq \frac{4}{\sqrt{3}}\right) &= P\left(X - 1 \geq \frac{4}{\sqrt{3}} \ \vee\ X - 1 \leq -\frac{4}{\sqrt{3}}\right) \\ &= P\Big(X \geq \underbrace{1 + \frac{4}{\sqrt{3}}}_{\approx 3.309}\Big) + P\Big(X \leq \underbrace{1 - \frac{4}{\sqrt{3}}}_{\approx -1.309}\Big) \\ &= 0 < \frac{1}{4}. \end{aligned}$$

13. Seja X uma v.a. tal que $E[X] = 0$ e $P(-3 < X < 2) = 0.5$. Determine um limite inferior para a variância de X.

    Resolução:
    A desigualdade de Chebyshev é verdadeira para todo o $k > 0$, pelo que podemos reescrevê-la fazendo a substituição $k = \delta/\sqrt{V(X)}$, para qualquer $\delta > 0$, obtendo:

    $$P\left(|X - E(X)| \geq \delta\right) \leq \frac{V(X)}{\delta^2}.$$

    Esta é uma forma equivalente da desigualdade de Chebyshev que nos será mais conveniente neste exercício.

    Temos:

    $$0.5 = P(-3 < X < 2) = P(-3 < X \leq -2) + P(-2 < X < 2) = P(-3 < X \leq -2) + P(|X| < 2),$$

    logo $P(|X| < 2) \leq 0.5$ e, portanto, $P(|X| \geq 2) \geq 1 - 0.5 = 0.5$. Como $E(X) = 0$, escrevendo a desigualdade de Chebyshev (na forma dada acima) para $\delta = 2$, obtém-se $P(|X| \geq 2) \leq V(X)/4$. Assim, finalmente,

    $$0.5 \leq P(|X| \geq 2) \leq \frac{V(X)}{4} \implies 0.5 \leq \frac{V(X)}{4} \Leftrightarrow V(X) \geq 2.$$