# FICHA 09 – resolução

DPC

1. Considere uma amostra de tamanho $n$ de uma variável aleatória $X$ com distribuição Poisson (fdp a seguir).

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, ...$$

Calcule o estimador de $\lambda$ pelo método da máxima verosimilhança.

**Resolução:**
Dada uma função de probabilidade $f_X$ com $k$ parâmetros desconhecidos $\theta_1, \theta_2, \cdots, \theta_k$ e dadas $n$ amostras $x_1, x_2, \cdots, x_n$, obtidas de forma independente e a partir da mesma distribuição $f_X$, o método da máxima verosimilhança procura os valores dos parâmetros que maximizam a probabilidade conjunta das amostras observadas:

$$(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k\ ) = \underset{\theta_1,\theta_2,\cdots,\theta_k}{\operatorname{argmax}}\ P(X_1 = x_1 \wedge X_2 = x_2 \wedge \cdots \wedge X_n = x_n).$$

Como (se assume que) as amostras são independentes, temos:

$$P(X_1 = x_1 \wedge X_2 = x_2 \wedge \cdots \wedge X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i).$$

Como todas as amostras seguem a distribuição $f_X$, temos $P(X_i = x_i) = f_X(x_i)$ para todo o $i$ e, assim:

$$\prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} f_X(x_i) = \mathcal{L}(\theta_1, \theta_2, \cdots, \theta_k).$$

Queremos então encontrar $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ que maximizam $\mathcal{L}(\theta_1, \theta_2, \cdots, \theta_k)$. Se $f_X$ for contínua e diferenciável em relação aos respetivos parâmetros, a solução está entre os pontos onde o gradiente de $\mathcal{L}$ é nulo:

$$\nabla\mathcal{L}(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) = 0 \Leftrightarrow \begin{cases} \frac{\partial\mathcal{L}}{\partial\theta_1}(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) = 0 \\ \frac{\partial\mathcal{L}}{\partial\theta_2}(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) = 0 \\ \vdots \\ \frac{\partial\mathcal{L}}{\partial\theta_k}(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) = 0 \end{cases}.$$

Note que o cálculo destas derivadas implica o cálculo da derivada de um produto, que pode ser bastante trabalhoso (e pouco estável numericamente). Para o evitar, em vez de maximizarmos $\mathcal{L}$ diretamente, podemos maximizar $\log\mathcal{L}$, visto que a função logaritmo não altera a posição do maximizante (porquê?). Temos, então:

$$\log\mathcal{L}(\theta_1, \theta_2, \cdots, \theta_k) = \log\left(\prod_{i=1}^{n} f_X(x_i)\right) = \sum_{i=1}^{n} \log f_X(x_i),$$

e assim transformámos o produto numa soma (de logaritmos). Como a posição do maximizante se mantém, restaria agora resolver $\nabla\log\mathcal{L}(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k) = 0$.

Neste exercício, temos:

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}.$$

Podemos resolver de imediato através de $\frac{d\mathcal{L}}{d\lambda}(\hat{\lambda}) = 0$ ou aplicar previamente o logaritmo, como sugerido anteriormente. Vamos resolver utilizando $\log \mathcal{L}$ e deixamos a outra resolução como exercício:

$$\begin{aligned}
\log \mathcal{L}(\lambda) &= \sum_{i=1}^{n} \log \left( \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) \\
&= \sum_{i=1}^{n} (-\lambda + x_i \log \lambda - \log x_i!) \\
&= -n\lambda + \log(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log x_i!
\end{aligned}$$

$$\frac{d \log \mathcal{L}}{d\lambda}(\hat{\lambda}) = 0 \Leftrightarrow -n + \frac{1}{\hat{\lambda}} \sum_{i=1}^{n} x_i = 0 \Leftrightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \text{(estimativa)}$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i. \quad \text{(estimador)}$$

Nota: Em rigor, só demonstrámos que $\hat{\lambda} = (1/n) \sum_{i=1}^{n} x_i$ é um ponto crítico de $\mathcal{L}$, faltando verificar que se trata de facto de um maximizante. Essa verificação pode ser feita de forma simples (por exemplo, pelo sinal da segunda derivada), pelo que é deixada como exercício. Para além disso, neste tipo de problemas, quando a solução obtida é única, sabemos que será necessariamente um maximizante e assim podemos ignorar esta verificação.

2. Considere uma amostra de tamanho $n$ de uma variável aleatória com distribuição

$$f(x) = \frac{1}{\theta^2} x e^{-x/\theta}, \quad 0 \leq x < \infty$$

Calcule o estimador de máxima verosimilhança de $\theta$.

**Resolução:**
Neste caso, temos uma distribuição contínua, mas o método de resolução é exatamente igual. A discussão elaborada no exercício anterior permanece válida, com a única diferença de que, agora, estamos a maximizar a densidade de probabilidade (e não a probabilidade) conjunta das $n$ amostras independentes.

$$\begin{aligned}
\log \mathcal{L}(\theta) &= \sum_{i=1}^{n} \log f(X_i) = \sum_{i=1}^{n} \left( -2 \log \theta + \log X_i - \frac{X_i}{\theta} \right) \\
&= -2n \log \theta + \sum_{i=1}^{n} \log X_i - \frac{1}{\theta} \sum_{i=1}^{n} X_i,
\end{aligned}$$

$$\frac{d \log \mathcal{L}}{d\theta}(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta} = \frac{1}{2n} \sum_{i=1}^{n} X_i.$$

Sugestão: Compare este estimador com o obtido pelo método dos momentos (ex. 6 da Ficha 8).

3. Uma variável aleatória $X$ tem fdp $f(x) = (\beta + 1)x^{\beta}, 0 < x < 1$.

   (a) Determine a estimativa de Máxima Verosimilhança de $\beta$ baseado numa amostra de tamanho $n$.

   (b) Calcule a estimativa quando os valores amostrais forem $0.3; 0.8; 0.27; 0.35; 0.62; 0.55$. Sabe-se que a soma destes valores é 2.89, o produto é 0.0077, o logaritmo da soma é 1.0613 e o logaritmo do produto é $-4.8621$.

**Resolução:**

(a)

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} \log f(x_i) = \sum_{i=1}^{n} (\log(\beta + 1) + \beta \log x_i) = n \log(\beta + 1) + \beta \sum_{i=1}^{n} \log x_i,$$

$$\frac{d \log \mathcal{L}}{d\beta}(\hat{\beta}) = 0 \Leftrightarrow \frac{n}{\hat{\beta} + 1} + \sum_{i=1}^{n} \log x_i \Leftrightarrow \hat{\beta} = -1 - \frac{n}{\sum_{i=1}^{n} \log x_i}.$$

(b) $n = 6$, $\sum_{i=1}^{6} \log x_i = \log\left(\prod_{i=1}^{6} x_i\right) = -4.8621 \implies \hat{\beta} \approx 0.234$.

Sugestão: Compare o estimador e a estimativa obtidos com os que se obtêm pelo método dos momentos (ex. 7 da Ficha 8).

4. Temos duas máquinas para produção de um dispositivo semiconductor com um comprimento médio de $\mu$. No entanto, a máquina 1 é mais nova do que a máquina 2. Portanto, a variância da máquina 1 é $\sigma_1^2$ e a variância da máquina 2 é $\sigma_2^2 = a\sigma_1^2$ com $a \geq 1$. Suponha que temos $n_1$ observações da máquina 1 e $n_2$ observações da máquina 2.

   (a) Demonstrar que $\hat{\mu} = \alpha \bar{X}_1 + (1-\alpha)\bar{X}_2$ é um estimador sem tendência para $\mu$ com $\alpha \in [0,1]$.

   (b) Qual é o erro padrão do estimador da alínea anterior?

   (c) Qual é o valor de $\alpha$ que minimiza o erro padrão do estimador de $\mu$.

   (d) Suponha que $a = 4$ e $n_1 = 2n_2$. Qual é o valor de $\alpha$ que minimiza o erro padrão? Quanto seria a diferença se o $\alpha$ fosse arbitrariamente escolhido como $\alpha = 1/2$?

**Resolução:**

(a) $E(X_1) = E(X_2) = \mu$, $V(X_1) = \sigma^2$, $V(X_2) = a\sigma^2$, $a \geq 1$.

$$\bar{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1,i} \implies E(\bar{X}_1) = \mu,$$

$$\bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2,i} \implies E(\bar{X}_2) = \mu,$$

$$E(\hat{\mu}) = E(\alpha\bar{X}_1 + (1-\alpha)\bar{X}_2) = \alpha E(\bar{X}_1) + (1-\alpha)E(\bar{X}_2) = \mu. \quad \square$$

(b) Assumindo que as amostras de cada máquina são independentes entre si e entre as duas máquinas, temos:

$$V(\bar{X}_1) = V\left(\frac{1}{n_1}\sum_{i=1}^{n_1} X_{1,i}\right) = \frac{1}{n_1^2}\sum_{i=1}^{n_1} V(X_{1,i}) = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\sigma^2 = \frac{\sigma^2}{n_1}, \quad V(\bar{X}_2) = \frac{a\sigma^2}{n_2},$$

$$\begin{aligned} V(\hat{\mu}) &= V(\alpha\bar{X}_1 + (1-\alpha)\bar{X}_2) = \alpha^2 V(\bar{X}_1) + (1-\alpha)^2 V(\bar{X}_2) \\ &= \frac{\alpha^2\sigma^2}{n_1} + \frac{(1-\alpha)^2 a\sigma^2}{n_2}, \end{aligned}$$

$$\varepsilon = \sqrt{E[(\hat{\mu}-\mu)^2]} = \sqrt{(E(\hat{\mu}-\mu))^2 + V(\hat{\mu})} = \sqrt{V(\hat{\mu})} = \sqrt{\frac{\alpha^2\sigma^2}{n_1} + \frac{(1-\alpha)^2 a\sigma^2}{n_2}}.$$

(c)

$$\frac{d\varepsilon}{d\alpha}(\alpha^*) = 0 \Leftrightarrow \alpha^* = \frac{an_1}{an_1 + n_2}.$$

(d) $a = 4$, $n_1 = 2n_2 \implies \alpha^* = 8/9$, $\quad \varepsilon(\alpha = 8/9)/\varepsilon(\alpha = 1/2) \approx 0.628$.

5. $X_1, X_2, ..., X_n$ é uma amostra aleatória de uma distribuição normal com média $\mu$ e variância $\sigma^2$. Sejam $X_{\min}$ e $X_{\max}$ o menor e o maior valor na amostra, respectivamente.

   (a) O estimador $\frac{X_{\min}+X_{\max}}{2}$ tem tendência?

   (b) Calcule o erro padrão deste estimador.

   (c) Que estimador utilizaria, a média aritmética ou este estimador?

**Resolução:**

(Este exercício é difícil!)

(a) Seja $X \sim N(\mu, \sigma^2)$. Pretende-se averiguar a tendência do estimador $\hat{\mu} = (X_{\min} + X_{\max})/2$ para a média $\mu$ da distribuição. Como habitualmente, temos um conjunto de $n$ observações independentes da v.a. $X$. Como $X_{\max}$ é, por definição, o valor máximo da amostra, se $X_{\max}$ é majorado um dado $x$, então todos os valores da amostra são majorados por esse $x$, ou seja:

$$X_{\max} \leq x \Leftrightarrow X_1 \leq x \wedge X_2 \leq x \wedge \cdots \wedge X_n \leq x.$$

Assim, a função de distribuição acumulada de $X_{\max}$ é dada por:

$$\begin{aligned} F_{X_{\max}}(x) &= P(X_{\max} \leq x) \\ &= P(X_1 \leq x \wedge X_2 \leq x \wedge \cdots \wedge X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x)\cdots P(X_n \leq x) && (X_i \text{ indep.}) \\ &= P(X \leq x)^n && (X_i \overset{d}{=} X) \\ &= F_X(x)^n, \end{aligned}$$

onde $F_X$ é a função de distribuição acumulada de $X$. A função densidade de probabilidade de $X_{\max}$ é, então:

$$f_{X_{\max}}(x) = \frac{d}{dx}F_{X_{\max}}(x) = n f_X(x) F_X(x)^{n-1},$$

onde $f_X = dF_X/dx$ é a função densidade de probabilidade de $X$. Uma vez encontrada $f_{X_{\max}}$, estamos em condições de escrever uma expressão para $E(X_{\max})$:

$$E(X_{\max}) = \int_{-\infty}^{\infty} x f_{X_{\max}}(x)\, dx = n \int_{-\infty}^{\infty} x f_X(x) F_X(x)^{n-1}\, dx.$$

Relativamente ao mínimo da amostra ($X_{\min}$) aplica-se o raciocínio recíproco: se $X_{\min}$ é minorado por um dado $x$, então todos os valores da amostra são minorados por esse $x$, ou seja:

$$X_{\min} > x \Leftrightarrow X_1 > x \wedge X_2 > x \wedge \cdots \wedge X_n > x.$$

Assim, a função de distribuição acumulada de $X_{\min}$ é:

$$\begin{aligned} F_{X_{\min}}(x) &= P(X_{\min} \leq x) = 1 - P(X_{\min} > x) \\ &= 1 - P(X_1 > x \wedge X_2 > x \wedge \cdots \wedge X_n > x) \\ &= 1 - P(X_1 > x)P(X_2 > x)\cdots P(X_n > x) && (X_i \text{ indep.}) \\ &= 1 - P(X > x)^n && (X_i \overset{d}{=} X) \\ &= 1 - (1 - P(X \leq x))^n = 1 - (1 - F_X(x))^n \end{aligned}$$

e a função densidade e valor esperado correspondentes ficam, respetivamente:

$$f_{X_{\min}}(x) = \frac{d}{dx}F_{X_{\min}}(x) = n f_X(x)(1 - F_X(x))^{n-1},$$

$$E(X_{\min}) = \int_{-\infty}^{\infty} x f_{X_{\min}}(x)\, dx = n \int_{-\infty}^{\infty} x f_X(x)(1 - F_X(x))^{n-1}\, dx.$$

A partir das expressões que obtivemos para $E(X_{\max})$ e $E(X_{\min})$, podemos escrever a expressão resultante para o valor esperado do estimador $\hat{\mu}$:

$$\begin{aligned} E(\hat{\mu}) &= E\left(\frac{X_{\min} + X_{\max}}{2}\right) = \frac{E(X_{\min}) + E(X_{\max})}{2} \\ &= \frac{n}{2}\int_{-\infty}^{\infty} x f_X(x)(1 - F_X(x))^{n-1}\, dx + \frac{n}{2}\int_{-\infty}^{\infty} x f_X(x) F_X(x)^{n-1}\, dx. \end{aligned}$$

Note que, até ao momento, não utilizámos a informação de que $X$ segue uma distribuição normal. Agora, iremos utilizá-la, tirando partido das propriedades dessa distribuição (em particular, da simetria) para prosseguir a resolução. Recorde que a distribuição $N(\mu, \sigma^2)$ é simétrica em torno de $\mu$, logo $f_X(\mu + x) = f_X(\mu - x)$ para todo o $x$, assim:

$$\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(x')\,dx' \\
&= \int_{-\infty}^{x} f_X(2\mu - x')\,dx' \qquad (f_X(x') = f_X(\mu + (x'-\mu)) \underset{\text{simetria}}{=} f_X(\mu - (x'-\mu)) = f_X(2\mu - x')) \\
&= \int_{2\mu - x}^{\infty} f_X(t)\,dt \qquad (t = 2\mu - x', \quad dt = -dx') \\
&= P(X > 2\mu - x) \\
&= 1 - F_X(2\mu - x)
\end{aligned}$$

Fazendo a substituição $1 - F_X(x) = F_X(2\mu - x)$ na expressão para $E(\hat{\mu})$, obtém-se:

$$\begin{aligned}
E(\hat{\mu}) &= \frac{n}{2}\int_{-\infty}^{\infty} x f_X(x) F_X(2\mu - x)^{n-1}\,dx + \frac{n}{2}\int_{-\infty}^{\infty} x f_X(x) F_X(x)^{n-1}\,dx \\
& \qquad (x \leftarrow 2\mu - x, \quad dx \leftarrow -dx) \\
&= \frac{n}{2}\int_{-\infty}^{\infty} (2\mu - x) f_X(2\mu - x) F_X(x)^{n-1}\,dx + \frac{n}{2}\int_{-\infty}^{\infty} x f_X(x) F_X(x)^{n-1}\,dx \\
& \qquad (f_X(2\mu - x) = f_X(x)) \\
&= \frac{n}{2}\int_{-\infty}^{\infty} (2\mu - x) f_X(x) F_X(x)^{n-1}\,dx + \frac{n}{2}\int_{-\infty}^{\infty} x f_X(x) F_X(x)^{n-1}\,dx \\
&= \mu \int_{-\infty}^{\infty} n f_X(x) F_X(x)^{n-1}\,dx \\
&= \mu \int_{-\infty}^{\infty} \frac{d}{dx} F_X(x)^n\,dx \\
&= \mu [F_X(x)^n]_{-\infty}^{\infty} = \mu(1^n - 0^n) = \mu,
\end{aligned}$$

logo $\hat{\mu}$ é um estimador centrado de $\mu$.

(b) –

(c) A média amostral (isto é, a média aritmética dos valores da amostra) é menos sensível do que $\hat{\mu}$ a valores extremos, logo apresenta menor variância. Assim, como ambos os estimadores são centrados, apresenta também menor erro padrão, pelo que escolheríamos a média amostral.

<u>Sugestão:</u> Através de simulação computacional, confirme empiricamente que: i) ambos os estimadores são centrados; ii) a variância da média amostral é inferior à variância de $\hat{\mu}$; iii) o erro padrão da média amostral é inferior ao erro padrão de $\hat{\mu}$.

6. Uma variável aleatória $X$ tem distribuição $N(\mu, 1)$. Tomam-se 20 observações de $X$, mas em vez de se registar o valor real, anotamos somente se $X$ era ou não negativo. Suponha que o evento $\{X < 0\}$ tenha ocorrido exactamente 14 vezes. Utilizando essa informação, determine a estimativa de MV de $\mu$.

**Resolução:**
$X \sim N(\mu, 1)$, vamos definir a v.a. $Y$ tal que:

$$Y = \begin{cases} 0, & \text{se } X < 0 \\ 1, & \text{se } X \geq 0 \end{cases}.$$

Temos então 20 amostras $y_1, y_2, \cdots, y_{20}$ de valores de $Y$ e sabemos que, entre estas, temos 14 amostras com o valor "0"(e, necessariamente, 6 amostras com o valor "1"). Assim,

$$\begin{aligned}
\log \mathcal{L}(\mu) &= \sum_{i=1}^{20} \log f_Y(y_i) = 14 \log f_Y(0) + 6 \log f_Y(1) \\
&= 14 \log P(Y = 0) + 6 \log P(Y = 1) = 14 \log \underbrace{P(X < 0)}_{1-p} + 6 \log \underbrace{P(X \geq 0)}_{p} \\
&= 6 \log p + 14 \log(1 - p).
\end{aligned}$$

Note que $p = P(X \geq 0)$ depende de $\mu$, pelo que $\log \mathcal{L}$ é função de $\mu$, como indicado. Pela regra da cadeia do cálculo de derivadas, temos:

$$\frac{d \log \mathcal{L}}{d\mu} = \frac{d \log \mathcal{L}}{dp} \frac{dp}{d\mu},$$

e, assim:

$$\begin{aligned} \frac{d \log \mathcal{L}}{d\mu}(\hat{\mu}) = 0 &\Leftrightarrow \frac{d \log \mathcal{L}}{dp}(p(\hat{\mu})) \frac{dp}{d\mu}(\hat{\mu}) = 0 \\ &\Leftrightarrow \frac{d \log \mathcal{L}}{dp}(p(\hat{\mu})) = 0 \ \vee \ \frac{dp}{d\mu}(\hat{\mu}) = 0. \end{aligned}$$

Observando que $p(\mu)$ é uma função estritamente crescente (confirme!), concluímos que a sua derivada é sempre estritamente positiva, pelo que a equação $\frac{dp}{d\mu}(\hat{\mu}) = 0$ não tem solução. Assim, resta-nos resolver:

$$\begin{aligned} \frac{d \log \mathcal{L}}{dp}(p(\hat{\mu})) = 0 &\Leftrightarrow \frac{d}{dp}(6 \log p + 14 \log(1-p)) \bigg|_{p=p(\hat{\mu})} = 0 \\ &\Leftrightarrow \frac{6}{p(\hat{\mu})} - \frac{14}{1 - p(\hat{\mu})} = 0 \Leftrightarrow p(\hat{\mu}) = \frac{3}{10}, \end{aligned}$$

$$\frac{3}{10} = p(\hat{\mu}) = P(X \geq 0) = P\Bigg( \underbrace{\frac{X - \hat{\mu}}{1}}_{Z \sim N(0,1)} \geq \frac{0 - \hat{\mu}}{1} \Bigg) = P(Z \geq -\hat{\mu}) \Leftrightarrow \hat{\mu} \approx -0.5244.$$