Universidade de São Paulo
Instituto de Matemática e Estatística
Aluna: Nara Avila Moraes
Nro. USP: 5716734

2o. semestre, 2025
Exercício 02 - 15/10

MAE5911/IME: Fundamentos de Estatística e Machine Learning.
Prof.: Alexandre Galvão Patriota

Questão 01: Seja $(Z_1, \dots, Z_n)$ uma amostra aleatória de $Z \sim \text{Ber}(\theta)$, $\theta \in (0, 1)$.

(a) Encontre o EMV para $g(\theta) = P_\theta(Z = 0)$.

(b) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.

(c) Considere que os dados foram observados (0, 0, 1, 0, 0, 1). Encontre as estimativas de MV nos itens acima.

(d) Construa o valor-p para a hipótese H: "$\theta = 0.1$" usando os dados do item anterior.

Podemos calcular os ítens (a) e (b) utilizando a propriedade de invariância do estimador de máxima verossimilhança (EMV).

Teorema (Invariância do EMV). Seja $\hat{\theta}_{MV}$ o estimador de máxima verossimilhança de um parâmetro $\theta \in \Theta$ e seja $g : \Theta \to \mathcal{G}$ uma função. Então o estimador de máxima verossimilhança de

$$\tau = g(\theta)$$

é dado por

$$\hat{\tau}_{MV} = g(\hat{\theta}_{MV}).$$

Isto é, para obter o EMV de qualquer função de θ, basta aplicar essa função ao EMV de θ.

Cálculo do EMV de θ ($\hat{\theta}_{MV}$)

Função de verossimilhança

Seja $z_1, \dots, z_n$ uma amostra aleatória de $Z \sim \text{Bernoulli}(\theta), \quad \theta \in (0, 1)$. Com:

$$P_\theta(z_i = 1) = \theta, \qquad P_\theta(z_i = 0) = 1 - \theta.$$

A função de probabilidade para Z $= z_i$ é então:

$$f(z_i; \theta) = \theta^{z_i}(1 - \theta)^{1 - z_i}.$$

Como $z_1, \dots, z_n$ são independentes, a probabilidade conjunta, definida como a função de verossimilhança é o produto das densidades individuais:

$$\ell(\theta, z_1, \dots, z_n) = \prod_{i=1}^{n} \theta^{z_i}(1 - \theta)^{1 - z_i}.$$

Utilizando a propriedade de produto de exponenciais, temos:

$$\ell(\theta, z_1, \dots, z_n) = \theta^{\sum_{i=1}^n z_i} (1 - \theta)^{\sum_{i=1}^n (1 - z_i)} = \theta^{\sum_{i=1}^n z_i} (1 - \theta)^{n - \sum_{i=1}^n z_i}.$$

Seja

$$S = \sum_{i=1}^{n} z_i,$$

Podemos escrever:

$$\ell(\theta, S) = \theta^S (1 - \theta)^{n - S}.$$

Função Log-verossimilhança

Podemos simplificar os cálculos trabalhando com o logaritmo desta função, mantendo o estimador de máxima verossimilhança inalterado, uma vez que este trata-se do ponto crítico desta função. Maximizar $\ell(\theta)$ ou $\mathcal{L}(\theta) = \log \ell(\theta)$ é equivalente, devido ao fato do logaritmo ser estritamente crescente. Portanto, aplicando o logaritmo, simplificamos para:

$$\mathcal{L}(\theta, S) = \log \ell(\theta, S) = \log\left(\theta^S (1-\theta)^{n-S}\right) = S \log \theta + (n-S)\log(1-\theta).$$

Ponto crítico da log-verossimilhança

Calculando a derivada de $\mathcal{L}(\theta, S)$ em relação a θ:

$$\frac{\partial \mathcal{L}(\theta, S)}{\partial \theta} = \frac{\partial}{\partial \theta}\big[S \log \theta + (n-S)\log(1-\theta)\big] = S \cdot \frac{1}{\theta} + (n-S) \cdot \left(-\frac{1}{1-\theta}\right) = \frac{S}{\theta} - \frac{n-S}{1-\theta}.$$

Seja a condição de máximo onde está definido $\hat{\theta}_{MV}$:

$$\frac{\partial \mathcal{L}(\theta, S)}{\partial \theta} = 0 \iff \frac{S}{\hat{\theta}_{MV}} - \frac{n-S}{1-\hat{\theta}_{MV}} = 0.$$

$$\frac{S}{\hat{\theta}_{MV}} = \frac{n-S}{1-\hat{\theta}_{MV}}.$$

$$S(1-\hat{\theta}_{MV}) = \hat{\theta}_{MV}(n-S).$$

$$S - S\hat{\theta}_{MV} = n\hat{\theta}_{MV} - S\hat{\theta}_{MV}.$$

Os termos $-S\hat{\theta}_{MV}$ cancelam em ambos os lados, restando:

$$S = n\hat{\theta}_{MV}.$$

Isolando $\hat{\theta}_{MV}$:

$$\boxed{\hat{\theta}_{MV} = \frac{S}{n}.}$$

Estimador de máxima verossimilhança

Substituindo de volta em funcão da variável aleatória z_i, obtemos:

$$\hat{\theta}_{MV} = \frac{1}{n}\sum_{i=1}^{n} z_i = \bar{Z}.$$

Teste de hipótese

O teste de hipótese é fomulado como um framework de decisão. No processo de decisão existem dois possíveis cenários de enganos, sendo que ambos não podem ser reduzidos simultaneamente, a menos do aumento da amostra, e se um cenário de engano diminui o seu complementar aumenta. Escolhe-se por arbitrariedade priorizar a mitigação do erro **falso positivo**, em detrimento do erro **falso negativo**. Desta forma a hipótese nula é definida como a hipótese que se deseja refutar. Entretanto, o teste é construído para que haja apenas uma baixa probabilidade, comumente 5%, de que H0 seja refutada, ou seja, a sua rejeição é definida como um evento raro. Consequentemente, quando ela é rejeitada, considera-se que o processo de decisão está com uma margem de erro suficientemente reduzida, de acordo com o nível de significância escolhido, que reflete a magnitude do erro tolerado.

Deste framework, o valor-p é a métrica que caracteriza a amostra observada no cenário em que a hipótese nula é verdadeira. O valor-p então calculado é comparado com o nível de significância escolhido para que a decisão de rejeitar ou não a hipótese nula seja tomada.

(a) $g(\theta) = P_\theta(Z = 0)$

$$g(\theta) = P_\theta(Z = 0) = 1 - \theta.$$

Pelo teorema da **invariância do EMV**:

$$\widehat{g}_{MV} = g(\widehat{\theta}_{MV}) = 1 - \widehat{\theta}_{MV}.$$

Já encontramos que $\widehat{\theta}_{MV} = \overline{Z}$, então:

$$\boxed{\widehat{g}_{MV} = 1 - \overline{Z}}.$$

(b) $g(\theta) = \mathrm{Var}_\theta(Z)$

A variância é encontrada por:

$$\mathrm{Var}(Z) = E(Z^2) - E(Z)^2.$$

Pela definição de esperança para variáveis discretas,

$$E(Z) = \sum_{z \in \{0,1\}} z\,P(Z = z).$$

Mas $P(Z = 1) = \theta$ e $P(Z = 0) = 1 - \theta$, então:

$$E(Z) = 0 \cdot (1 - \theta) + 1 \cdot \theta = \theta.$$

$E(Z^2) = E(Z) = \theta$ (pois $Z \in \{0, 1\}$), então:

$$\mathrm{Var}_\theta(Z) = \theta - \theta^2 = \theta(1 - \theta).$$

Portanto:

$$g(\theta) = \mathrm{Var}_\theta(Z) = \theta(1 - \theta).$$

Pela **invariância do EMV**:

$$\widehat{g}_{MV} = g(\widehat{\theta}_{MV}) = \widehat{\theta}_{MV}(1 - \widehat{\theta}_{MV}).$$

Para $\widehat{\theta}_{MV} = \overline{Z}$, encontramos:

$$\boxed{\widehat{g}_{MV} = \overline{Z}(1 - \overline{Z}).}$$

(c) Sejam os dados observados: $(0, 0, 1, 0, 0, 1)$

Para o cálculo da média: $n = 6$ e $\sum z_i = 2$, portanto

$$\hat{\theta}_{MV} = \bar{Z} = \frac{2}{6} = \frac{1}{3}.$$

Assim:

$$(a) \quad g\hat{_{MV}} = \widehat{P_\theta(Z=0)} = 1 - \bar{Z} = \frac{2}{3},$$

$$(b) \quad g\hat{_{MV}} = \widehat{\mathrm{Var}_\theta}(Z) = \bar{Z}(1 - \bar{Z}) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}.$$

(d) Teste de hipótese $H_0 : \theta = 0.1$

Na ditribuição de Bernoulli, o parâmetro θ representa a probabilidade de sucesso, equivalente à média. Se supormos que a hipótese nula é verdadeira, o valor-p dos dados observados em relação a H0 é:

Dados amostrados:

$$(0, 0, 1, 0, 0, 1)$$

Sob a hipótese nula $\theta = 0.1$, contruimos X:

$$X \sim \text{Binomial}(n = 6,\ \theta = 0.1).$$

A estatística de interesse é a média amostral ou probabilidade de sucessos. Em relação à amostra observada, é:

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{6}\sum_{i=1}^{6} z_i = \frac{1}{6} * 2 = \frac{1}{3}.$$

Vamos analisar a média da amostra contra a média testada (a média da hipótese nula). O valor-p então é definido a partir de: no cenário em que $H0$ é válida, ou seja, temos uma distribuicao de Bernoulli com média = 0.1, qual a probabilidade de observar média $\bar{Z} = 1/3$ ou maior? O posicionamento da média observada em relação à média testada indica a direção do teste.

$$p = P_{H_0}\left(\bar{Z} \geq \frac{1}{3}\right).$$

Como $\quad X = \sum_{i=1}^{6} z_i, \quad \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ e $\quad n = 6$:

$$\bar{Z} = \frac{X}{6}.$$

Então,

$$\bar{Z} \geq \frac{1}{3} \iff \frac{X}{6} \geq \frac{1}{3} \iff X \geq 2.$$

Logo, o valor-p é:

$$p = P_{H_0}(X \geq 2),$$

$$p = P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6).$$

Cada termo é dado por:

$$P(X=k) = \binom{6}{k}(0.1)^k(0.9)^{6-k}.$$

Portanto:

$$P(X=2) = \binom{6}{2}(0.1)^2(0.9)^4 = 15 \cdot 0.01 \cdot 0.6561 = 0.098415.$$

$$P(X=3) = \binom{6}{3}(0.1)^3(0.9)^3 = 20 \cdot 0.001 \cdot 0.729 = 0.01458.$$

$$P(X=4) = \binom{6}{4}(0.1)^4(0.9)^2 = 15 \cdot 0.0001 \cdot 0.81 = 0.001215.$$

$$P(X=5) = \binom{6}{5}(0.1)^5(0.9) = 6 \cdot 0.00001 \cdot 0.9 = 0.000054.$$

$$P(X=6) = \binom{6}{6}(0.1)^6 = 1 \cdot 0.000001 = 0.000001.$$

Assim, o valor-p é:

$$p = 0.098415 + 0.01458 + 0.001215 + 0.000054 + 0.000001 = 0.114265.$$

Pela convervação da probabilidade, seria mais simples calcular o valor-p pelo seu valor complementar:

$$\bar{p} = P_{H0}(X < 2) = (1-p)$$

$$P(X=0) = (0.9)^6 = 0.531441, \qquad P(X=1) = \binom{6}{1}(0.1)(0.9)^5 = 0.354294.$$

Logo:

$$\bar{p} = 1 - p = P(X=0) + P(X=1) = 0.531441 + 0.354294 = 0.885735.$$

$$\boxed{\text{valor-p} = 1 - 0.885735 = 0.114265}$$

—

Considerando um nível de significância $\alpha = 0.05$ não há evidências suficientes para rejeitar H0, isto é, para supor que a média seja diferente de 0.1.

Questão 02: Seja $(Z_1, \ldots, Z_n)$ uma amostra aleatória de $Z \sim \text{Exp}(\theta)$, $\theta \in (0, \infty)$.

(a) Encontre o EMV para $g(\theta) = P_\theta(Z > 1)$.

(b) Encontre o EMV para $g(\theta) = P_\theta(0.1 < Z < 1)$.

(c) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.

(d) Considere que os dados foram observados (0.2, 0.6, 0.3, 0.2, 0.8, 0.12). Encontre um IC aproximado de 95% de confiança para $g(\theta)$ nos itens acima.

(e) Faça uma simulação de Monte Carlo para verificar se os IC's aproximados obtidos no passo anterior têm cobertura próxima do nível de confiança estabelecido. Caso não tenham, proponha um tamanho amostral que produza IC's mais confiáveis para cada caso.

Podemos calcular os ítens (a), (b) e (c) utilizando a propriedade de invariância do estimador de máxima verossimilhança (EMV), calculando uma única vez $\hat{\theta}_{MV}$.

Cálculo do EMV de θ ($\hat{\theta}_{MV}$) para a distribuição Exponencial

Função de verossimilhança

Seja $(z_1, \ldots, z_n)$ uma amostra aleatória de $Z \sim \text{Exp}(\theta)$, com $\theta \in (0, \infty)$, a função densidade de probabilidade para z_i é então:

$$f(z_i; \theta) = \theta e^{-\theta z_i}, \qquad z_i > 0.$$

Como $z_1, \ldots, z_n$ são independentes, a probabilidade conjunta, definida como a função de verossimilhança é o produto das densidades individuais:

$$\ell(\theta; z_1, \ldots, z_n) = \prod_{i=1}^{n} f(z_i \mid \theta) = \prod_{i=1}^{n} \theta e^{-\theta z_i} \mathbf{1}_{\{z_i > 0\}}.$$

O termo indicador $\prod_{i=1}^{n} \mathbf{1}_{\{z_i > 0\}}$ não depende de θ, logo não influencia a maximização e pode ser ignorado. Assim,

$$\ell(\theta; z_1, \ldots, z_n) = \prod_{i=1}^{n} \theta e^{-\theta z_i}.$$

Aplicando a propriedade distributiva do produto e de produto de exponenciais, obtemos:

$$\ell(\theta; z_1, \ldots, z_n) = \left(\prod_{i=1}^{n} \theta\right)\left(\prod_{i=1}^{n} e^{-\theta z_i}\right) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} z_i\right).$$

Seja

$$S = \sum_{i=1}^{n} z_i,$$

Podemos escrever:

$$\ell(\theta, S) = \theta^n \exp(-\theta S).$$

Função log-verossimilhança

Para simplificar os cálculos, trabalhamos com a função logaritmo da verossimilhança $\mathcal{L}(\theta, S)$:

$$\mathcal{L}(\theta, S) = \log \ell(\theta, S) = \log(\theta^n) + \log(\exp(-\theta S)).$$

Usando as propriedades $\log(a^b) = b \log a$ e $\log(e^x) = x$, obtemos

$$\mathcal{L}(\theta, S) = n \log \theta - \theta S.$$

Ponto crítico da log-verossimilhança

Calculando a derivada de $\mathcal{L}(\theta, S)$ em relação a θ:

$$\frac{\partial \mathcal{L}(\theta, S)}{\partial \theta} = \frac{\partial}{\partial \theta}\big[n \log \theta - \theta S\big] = n \cdot \frac{1}{\theta} - S \cdot 1 = \frac{n}{\theta} - S.$$

Seja a condição de máximo onde está definido $\hat{\theta}_{MV}$:

$$\frac{\partial \mathcal{L}(\theta, S)}{\partial \theta} = 0 \iff \frac{n}{\hat{\theta}_{MV}} - S = 0 \iff \frac{n}{\hat{\theta}_{MV}} = S.$$

Isolando $\hat{\theta}_{MV}$:

$$\boxed{\hat{\theta}_{MV} = \frac{n}{S},}$$

Estimador de máxima verossimilhança

Substituindo de volta em funcão da variável aleatória z_i, obtemos:

$$\hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^{n} z_i} = \frac{1}{\overline{Z}},$$

Composição da distribuição Exponencial contínua acumulada. Separando a distribuição exponencial em função de distribuição acumulada até a observação de interesse e o restante da distribuição acumulada denominada a cauda da distribuição, temos:

Expressão para densidade total:

$$f(z_i \mid \theta) = \theta e^{-\theta z_i} \mathbf{1}_{\{z_i \geq 0\}}.$$

Função de distribuição acumulada:

$$F(z) = P_\theta(Z \leq z) = \int_0^z \theta e^{-\theta t}\, dt = \big[- e^{-\theta t} \big]_0^z = 1 - e^{-\theta z}.$$

Cauda:

$$P_\theta(Z > z) = 1 - F(z) = e^{-\theta z}.$$

(a) $g(\theta) = P_\theta(Z > 1)$

Integrando a função densidade de probabilidade $f(z_i, \theta)$ para encontrar $P_\theta(Z > 1)$:

$$g(\theta) = P_\theta(Z > 1) = \int_1^\infty f(z, \theta)\, dz = \int_1^\infty \theta e^{-\theta z}\, dz.$$

Calculando a integral:

$$\int_1^\infty \theta e^{-\theta z}\, dz = \Big[- e^{-\theta z} \Big]_{z=1}^{\infty} = 0 - (-e^{-\theta}) = e^{-\theta}.$$

Portanto,

$$g(\theta) = e^{-\theta}.$$

Pela **invariância do EMV**, o estimador de máxima verossimilhança de $g(\theta)$ é

$$\widehat{g}_{MV} = g(\widehat{\theta}_{MV}).$$

Já encontramos que $\widehat{\theta}_{MV} = \dfrac{1}{\overline{Z}}$ para distribuição Exponencial, então:

$$\boxed{\widehat{g}_{MV} = e^{-\widehat{\theta}_{MV}} = \exp\left(-\frac{1}{\overline{Z}}\right)}.$$

(b) $g(\theta) = P_\theta(0.1 < Z < 1)$

Integrando a função densidade de probabilidade $f(z_i, \theta)$ para encontrar $P_\theta(0.1 < Z < 1)$:

$$g(\theta) = P_\theta(0.1 < Z < 1) = \int_{0.1}^{1} f(z, \theta)\, dz = \int_{0.1}^{1} \theta e^{-\theta z}\, dz.$$

Calculando a integral:

$$\int_{0.1}^{1} \theta e^{-\theta z}\, dz = \Big[- e^{-\theta z} \Big]_{z=0.1}^{1} = -e^{-\theta} \; - \big(- e^{-0.1\theta} \big) = e^{-0.1\theta} - e^{-\theta}.$$

Logo,

$$g(\theta) = e^{-0.1\theta} - e^{-\theta}.$$

Pela **invariância do EMV**,

$$\widehat{g}_{MV} = g(\widehat{\theta}_{MV}) = e^{-0.1\widehat{\theta}_{MV}} - e^{-\widehat{\theta}_{MV}}.$$

Substituindo $\widehat{\theta}_{MV} = \dfrac{1}{\overline{Z}}$, obtemos:

$$\boxed{\widehat{g}_{MV} = \exp\Big(-\frac{0.1}{\overline{Z}} \Big) - \exp\Big(-\frac{1}{\overline{Z}} \Big)}.$$

(c) $g(\theta) = \mathrm{Var}_\theta(Z)$

A variância é encontrada por:

$$\mathrm{Var}(Z) = E_\theta(Z^2) - \big(E_\theta(Z) \big)^2.$$

Pela definição de esperança para variáveis contínuas, sendo $h(z)$ a função de interesse:

$$E_\theta(h(Z)) = \int_{-\infty}^{\infty} h(z)\, f(z;\theta)\, dz.$$

Em particular, a esperança do primeiro e segundo momentos de Z, são:

$$E_\theta(Z) = \int_0^\infty z\, \theta e^{-\theta z}\, dz, = \big[-z e^{-\theta z}\big]_0^\infty + \int_0^\infty e^{-\theta z}\, dz = 0 + \Big[-\frac{1}{\theta} e^{-\theta z} \Big]_0^\infty = 0 + \Big(0 - \Big(-\frac{1}{\theta} \Big) \Big) = \frac{1}{\theta}.$$

e

$$E_\theta(Z^2) = \int_0^\infty z^2\, \theta e^{-\theta z}\, dz = \int_0^\infty \Big(\frac{y}{\theta} \Big)^2 \theta e^{-y}\, \frac{dy}{\theta} = \frac{1}{\theta^2} \int_0^\infty y^2 e^{-y}\, dy = \frac{1}{\theta^2} \cdot 2! = \frac{2}{\theta^2}.$$

Então:

$$\mathrm{Var}_\theta(Z) = E(Z^2) - [E(Z)]^2 = \frac{2}{\theta^2} - \Big(\frac{1}{\theta} \Big)^2 = \frac{1}{\theta^2}.$$

Portanto:

$$g(\theta) = \mathrm{Var}_\theta(Z) = \frac{1}{\theta^2}.$$

Pela **invariância do EMV**,

$$\widehat{g}_{MV} = g(\widehat{\theta}_{MV}) = \frac{1}{\hat{\theta}^2_{MV}}.$$

Substituindo $\widehat{\theta}_{MV} = \dfrac{1}{\overline{Z}}$, obtemos:

$$\boxed{\widehat{g}_{MV} = \bar{Z}^2}.$$

Questão 03: Seja $(Z_1, \dots, Z_n)$ uma amostra aleatória de $Z \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$.

(a) Encontre o EMV para $g(\theta) = E_\theta(Z)$.

(b) Encontre o EMV para $g(\theta) = P_\theta(Z < 2)$.

(c) Encontre o EMV para $g(\theta) = P_\theta(2.6 < Z < 4)$.

(d) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.

(e) Considere que os dados foram observados (2.4, 2.7, 2.3, 2, 2.5, 2.6). Encontre as estimativas de MV nos itens acima.

Podemos calcular os ítens (a), (b), (c) e (d) utilizando a propriedade de invariância do estimador de máxima verossimilhança (EMV), calculando uma única vez $\hat{\theta}_{MV}$.

Cálculo do EMV de θ ($\hat{\theta}_{MV}$) para a distribuição Normal

Função de verossimilhança

Seja $(z_1, \dots, z_n)$ uma amostra aleatória de $Z \sim N(\mu, \sigma^2)$, com parâmetro $\theta = (\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$.

A função densidade de probabilidade para z_i é

$$f(z_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(z_i - \mu)^2}{2\sigma^2}\right).$$

Como $z_1, \dots, z_n$ são independentes, a função de verossimilhança é o produto das densidades individuais:

$$\ell(\mu, \sigma^2; z_1, \dots, z_n) = \prod_{i=1}^{n} f(z_i; \mu, \sigma^2).$$

Aplicando a propriedade distributiva do produto e de produto de exponenciais, obtemos:

$$\ell(\mu, \sigma^2; z_1, \dots, z_n) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \mu)^2\right).$$

Seja

$$S = \sum_{i=1}^{n} z_i \qquad e \qquad Q = \sum_{i=1}^{n} z_i^2, \qquad \iff \qquad \sum_{i=1}^{n} (z_i - \mu)^2 = \sum_{i=1}^{n} (z_i^2 - 2\mu z_i + \mu^2) = Q - 2\mu S + n\mu^2.$$

Podemos escrever:

$$\ell(\mu, \sigma^2, S, Q) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(Q - 2\mu S + n\mu^2\right).$$

Para simplificar os cálculos, trabalhamos com a função logaritmo da verossimilhança $\mathcal{L}(\mu, \sigma^2, S, Q)$:

$$\mathcal{L}(\mu, \sigma^2, S, Q) = \log \ell(\mu, \sigma^2, S, Q) = \log \left[\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\right] + \log \left[\exp\left(-\frac{1}{2\sigma^2}(Q - 2\mu S + n\mu^2)\right)\right],$$

$$\mathcal{L}(\mu, \sigma^2, S, Q) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2}\left(Q - 2\mu S + n\mu^2\right).$$

Calculando a derivada de $\mathcal{L}(\mu, \sigma^2, S, Q)$ em relação a μ:

$$\frac{\partial \mathcal{L}(\mu, \sigma^2, S, Q)}{\partial \mu} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \mu} \left[Q - 2\mu S + n\mu^2\right] = -\frac{1}{2\sigma^2} \left(-2S + 2n\mu\right) = \frac{S - n\mu}{\sigma^2}.$$

Seja a condição de máximo onde está definido $\hat{\mu}_{MV}$:

$$\frac{\partial \mathcal{L}(\mu, \sigma^2, S, Q)}{\partial \mu} = 0 \iff \frac{S - n\hat{\mu}_{MV}}{\sigma^2} = 0 \iff S - n\hat{\mu}_{MV} = 0 \iff \hat{\mu}_{MV} = \frac{S}{n}.$$

Isolando $\hat{\theta}_{MV}$:

Agora, igualando a zero para encontrar o ponto crítico:

Como $S = \sum_{i=1}^n z_i$, isso é exatamente

$$\widehat{\mu}_{MV} = \bar{Z} = \frac{1}{n}\sum_{i=1}^n z_i.$$

$$\frac{\partial \mathcal{L}(\mu, \sigma^2, S, Q)}{\partial \mu} = 0 \iff \sum_{i=1}^n (z_i - \mu) = 0 \iff \sum_{i=1}^n z_i = n\mu \iff \widehat{\mu}_{MV} = \overline{Z} = \frac{1}{n}\sum_{i=1}^n z_i.$$

Agora derivamos em relação a σ^2. É conveniente reescrever ℓ como função de σ^2:

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (z_i - \mu)^2.$$

Logo,

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{\sum_{i=1}^n (z_i - \mu)^2}{(\sigma^2)^2}.$$

Igualando a zero:

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \iff -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{\sum_{i=1}^n (z_i - \mu)^2}{(\sigma^2)^2} = 0.$$

Multiplicando ambos os lados por $2(\sigma^2)^2$:

$$-n\sigma^2 + \sum_{i=1}^n (z_i - \mu)^2 = 0 \iff \sigma^2 = \frac{1}{n}\sum_{i=1}^n (z_i - \mu)^2.$$

Substituindo μ pelo seu EMV $\widehat{\mu}_{MV} = \overline{Z}$, obtemos o EMV de σ^2:

$$\widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^n (z_i - \overline{Z})^2.$$

Portanto, o **vetor de estimadores de máxima verossimilhança** é

$$\boxed{\widehat{\mu}_{MV} = \overline{Z}, \qquad \widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^n (z_i - \overline{Z})^2.}$$

Pela **invariância do EMV**, para qualquer função $g(\mu, \sigma^2)$ o EMV de g é dado por

$$\widehat{g}_{MV} = g(\widehat{\mu}_{MV}, \widehat{\sigma}^2_{MV}).$$

A distribuição Normal surge como limite assintótico da soma de variáveis aleatórias independentes com variância finita, conforme estabelecido pelo **Teorema Central do Limite (TCL)**.

TCL: Seja $\{X_1, X_2, \ldots, X_n\}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média $\mu = \mathbb{E}[X_i]$ e variância $\sigma^2 = \mathrm{Var}(X_i) < \infty$. Definimos a média amostral

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Então, a variável associada à média amostral centralizada e reescalada para ter média 0 e variância 1

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

converge em distribuição para uma Normal padrão, isto é,

$$Z_n \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{quando } n \to \infty.$$

Em outras palavras, independentemente da distribuição original das X_i, a média amostral tende, para amostras grandes, a seguir aproximadamente uma **distribuição Normal** com média μ e variância σ^2/n.

—

Para modelar matematicamente a função densidade da distribuição Normal, buscamos uma função contínua, que satisfaça as condições de simetria e normalização, da forma:

$$f(z) = A\, e^{-k(z-\mu)^2},$$

onde $A > 0$ e $k > 0$ são constantes a determinar. Essa escolha se justifica porque o termo $e^{-k(z-\mu)^2}$ é simétrico e decai rapidamente conforme $|z - \mu|$ aumenta, representando a concentração de probabilidade em torno de μ.

Impondo a condição de normalização $\int_{-\infty}^{+\infty} f(z)\, dz = 1$, temos:

$$A \int_{-\infty}^{+\infty} e^{-k(z-\mu)^2}\, dz = 1.$$

Usando a mudança de variável $x = \sqrt{k}\,(z - \mu)$, obtemos $dz = dx/\sqrt{k}$, e assim:

$$A \int_{-\infty}^{+\infty} e^{-x^2} \frac{dx}{\sqrt{k}} = 1.$$

Sabendo que $\displaystyle\int_{-\infty}^{+\infty} e^{-x^2}\, dx = \sqrt{\pi}$, segue que

$$A = \sqrt{\frac{k}{\pi}}.$$

Chamamos μ de **média** (ou valor esperado) e associamos k à **variância** σ^2 por meio da relação

$$k = \frac{1}{2\sigma^2},$$

que garante que $\mathrm{Var}(Z) = \sigma^2$. Substituindo A e k, obtemos:

$$\boxed{f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\}.}$$

Essa é a **função densidade de probabilidade** da distribuição Normal, denotada por $Z \sim \mathcal{N}(\mu, \sigma^2)$.

—

Agora, considerando uma amostra aleatória $(Z_1, \ldots, Z_n)$ de $Z \sim \mathcal{N}(\mu, \sigma^2)$, temos a função de densidade conjunta, a **função de verossimilhança**, como produtório das densidades individuais:

$$L(\mu, \sigma^2; z_1, \ldots, z_n) = \prod_{i=1}^{n} f(z_i \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \mu)^2\right].$$

Calculando a função **log verossimilhança**, para simplificar a maximização:

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \mu)^2.$$

Sabendo que $\frac{d}{d\mu}(z_i - \mu)^2 = -2(z_i - \mu)$ e que a derivada é linear em relação à soma, derivamos em relação a μ:

$$\frac{\partial \ell}{\partial \mu} = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [-2(z_i - \mu)] = \frac{1}{\sigma^2} \sum_{i=1}^{n} (z_i - \mu).$$

Igualando a zero para encontrar o ponto crítico:

$$\frac{\partial \ell}{\partial \mu} = 0 \quad \iff \quad \sum_{i=1}^{n} (z_i - \mu) = 0 \quad \iff \quad n\mu = \sum_{i=1}^{n} z_i.$$

Assim, o **estimador de máxima verossimilhança** para μ é

$$\boxed{\widehat{\mu}_{MV} = \bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i.}$$

—

Agora derivamos a **log-verossimilhança** em relação a σ^2:

$$\frac{\partial \ell}{\partial (\sigma^2)} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (z_i - \mu)^2.$$

Igualando a zero e substituindo $\mu = \widehat{\mu}_{MV}$:

$$0 = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (z_i - \bar{Z})^2.$$

Multiplicando ambos os lados por $2(\sigma^2)^2$ e rearranjando os termos:

$$n\sigma^2 = \sum_{i=1}^{n} (z_i - \bar{Z})^2 \quad \Longrightarrow \quad \boxed{\widehat{\sigma^2}_{MV} = \frac{1}{n} \sum_{i=1}^{n} (Z_i - \bar{Z})^2.}$$

—

Observação importante: O estimador de máxima verossimilhança da variância é *viciado* para amostras finitas, pois subestima a verdadeira dispersão. Calculando a esperança para o estimador obtido, tem-se que:

$$E[\widehat{\sigma^2}_{MV}] = \frac{n-1}{n}\sigma^2.$$

Esse viés ocorre porque a média amostral $\bar{Z}$ é calculada a partir dos próprios dados, introduzindo uma restrição linear entre os desvios $(Z_i - \bar{Z})$. Assim, apenas $n-1$ deles são linearmente independentes, o que caracteriza a perda de um grau de liberdade.

$$\sum_{i=1}^{n}(Z_i - \bar{Z}) = 0 \quad \Longrightarrow \quad (Z_n - \bar{Z}) = -\sum_{i=1}^{n-1}(Z_i - \bar{Z}).$$

Para eliminar o viés, multiplicamos o estimador por $\frac{n}{n-1}$, restaurando a estimativa não viciada da variância populacional:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2, \qquad \text{para o qual} \quad E[S^2] = \sigma^2.$$

Essa correção garante que o estimador seja não viciado. Quando n é grande, a diferença entre os estimadores é desprezível, e $\widehat{\sigma^2}_{MV}$ é assintoticamente não viciado e consistente.

—

Derivando novamente para confirmar que o ponto crítico é de máximo:

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n}{\sigma^2} < 0, \qquad \frac{\partial^2 \ell}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(z_i - \mu)^2.$$

Substituindo $\sum(z_i - \widehat{\mu}_{MV})^2 = n\widehat{\sigma}^2_{MV}$, obtemos valores negativos, confirmando o máximo.

—

Os estimadores de máxima verossimilhança para os parâmetros da distribuição Normal são:

$$\boxed{\widehat{\mu}_{MV} = \bar{Z}, \qquad \widehat{\sigma^2}_{MV} = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})^2.}$$

E pela **propriedade de invariância do EMV**, para qualquer função $g(\mu, \sigma^2)$, temos:

$$\boxed{\widehat{g}_{MV} = g(\widehat{\mu}_{MV}, \widehat{\sigma}^2_{MV}).}$$

(a) $g(\theta) = E_\theta(Z)$ Para a Normal $N(\mu, \sigma^2)$, a esperança é

$$E_\theta(Z) = \mu.$$

Logo,

$$g(\mu, \sigma^2) = \mu.$$

Pela invariância do EMV,

$$\widehat{g}_{MV} = g(\widehat{\mu}_{MV}, \widehat{\sigma}^2_{MV}) = \widehat{\mu}_{MV} = \overline{Z}.$$

Portanto,

$$\boxed{\widehat{g}_{MV} = \overline{Z}.}$$

(b) $g(\theta) = P_\theta(Z < 2)$ Para uma Normal $N(\mu, \sigma^2)$, podemos padronizar:

$$P_\theta(Z < 2) = P\left(\frac{Z-\mu}{\sigma} < \frac{2-\mu}{\sigma}\right) = \Phi\left(\frac{2-\mu}{\sigma}\right),$$

onde $\Phi(\cdot)$ é a função de distribuição da Normal padrão $N(0,1)$ e $\sigma = \sqrt{\sigma^2}$.

Assim,

$$g(\mu, \sigma^2) = \Phi\left(\frac{2-\mu}{\sqrt{\sigma^2}}\right).$$

Pela invariância do EMV, o estimador de máxima verossimilhança de $g(\theta)$ é

$$\widehat{g}_{MV} = g(\widehat{\mu}_{MV}, \widehat{\sigma}^2_{MV}) = \Phi\left(\frac{2-\widehat{\mu}_{MV}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right) = \Phi\left(\frac{2-\overline{Z}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right),$$

com

$$\widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{Z})^2.$$

Portanto,

$$\boxed{\widehat{g}_{MV} = \Phi\left(\frac{2-\overline{Z}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right).}$$

(c) $g(\theta) = P_\theta(2.6 < Z < 4)$ De forma análoga, escrevemos a probabilidade em termos da Normal padrão:

$$P_\theta(2.6 < Z < 4) = P_\theta(Z < 4) - P_\theta(Z \leq 2.6) = \Phi\left(\frac{4-\mu}{\sigma}\right) - \Phi\left(\frac{2.6-\mu}{\sigma}\right).$$

Logo,

$$g(\mu, \sigma^2) = \Phi\left(\frac{4-\mu}{\sqrt{\sigma^2}}\right) - \Phi\left(\frac{2.6-\mu}{\sqrt{\sigma^2}}\right).$$

Pela invariância do EMV,

$$\widehat{g}_{MV} = g(\widehat{\mu}_{MV}, \widehat{\sigma}^2_{MV}) = \Phi\left(\frac{4-\overline{Z}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right) - \Phi\left(\frac{2.6-\overline{Z}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right),$$

onde novamente

$$\widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{Z})^2.$$

Portanto,

$$\boxed{\widehat{g}_{MV} = \Phi\left(\frac{4-\overline{Z}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right) - \Phi\left(\frac{2.6-\overline{Z}}{\sqrt{\widehat{\sigma}^2_{MV}}}\right).}$$

(d) $g(\theta) = \mathrm{Var}_\theta(Z)$ Para a Normal $N(\mu, \sigma^2)$,

$$\mathrm{Var}_\theta(Z) = \sigma^2.$$

Portanto,

$$g(\mu, \sigma^2) = \sigma^2.$$

Pela invariância do EMV,

$$\widehat{g}_{MV} = g(\widehat{\mu}_{MV}, \widehat{\sigma}^2_{MV}) = \widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{Z})^2.$$

Assim,

$$\boxed{\widehat{g}_{MV} = \widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{Z})^2.}$$

(a) $g(\theta) = E_\theta(Z)$

Esperança matemática. A esperança (ou valor esperado) de uma variável aleatória é o análogo contínuo da média ponderada. Ela representa o valor médio que esperaríamos observar após infinitas repetições do experimento.

- **Caso discreto:** se Z assume valores z_i com probabilidades p_i, então

$$E(Z) = \sum_i z_i p_i.$$

- **Caso contínuo:** se Z tem densidade $f(z)$, então

$$E(Z) = \int_{-\infty}^{+\infty} z\, f(z)\, dz.$$

Exemplo: Normal $N(\mu, \sigma^2)$. A densidade é

$$f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right].$$

Então:

$$E(Z) = \int_{-\infty}^{+\infty} z\, f(z \mid \mu, \sigma^2)\, dz.$$

Fazendo a mudança de variável $x = \frac{z-\mu}{\sigma}$, obtemos

$$E(Z) = \int_{-\infty}^{+\infty} (\mu + \sigma x)\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \mu \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx}_{=1} + \sigma \underbrace{\int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}} e^{-x^2/2}\, dx}_{=0}.$$

Logo,

$$\boxed{E(Z) = \mu.}$$

Fato conhecido da Normal: se $Z \sim N(\mu, \sigma^2)$ então $E(Z) = \mu$.

▷ **Como justificar rapidamente:** pela *linearidade da esperança* e pela *padronização* $Z = \mu + \sigma T$ com $T \sim N(0, 1)$,

$$E(Z) = E(\mu + \sigma T) = \mu + \sigma E(T) = \mu + \sigma \cdot 0 = \mu.$$

Logo, $g(\theta) = \mu$ e, pela **invariância do EMV**,

$$\boxed{\widehat{g}_{(a)} = \widehat{\mu}_{MV} = \overline{Z}}$$

(b) $g(\theta) = P_\theta(Z < 2)$

Passo 1 — Padronização (transformação linear): Se $Z \sim N(\mu, \sigma^2)$ então

$$T = \frac{Z - \mu}{\sigma} \sim N(0, 1).$$

Passo 2 — Troca de variável na probabilidade (monotonicidade):

$$P_\theta(Z < 2) = P\left(\frac{Z-\mu}{\sigma} < \frac{2-\mu}{\sigma}\right) = \Phi\left(\frac{2-\mu}{\sigma}\right).$$

(Usamos que a função $z \mapsto (z-\mu)/\sigma$ é estritamente crescente quando $\sigma > 0$.)

Passo 3 — Invariância do EMV (plug-in):

$$\boxed{\widehat{g}_{(b)} = \Phi\left(\frac{2-\widehat{\mu}_{MV}}{\widehat{\sigma}_{MV}}\right)}$$

(c) $g(\theta) = P_\theta(2.6 < Z < 4)$

Passo 1 — Padronização:

$$T = \frac{Z-\mu}{\sigma} \sim N(0,1).$$

Passo 2 — Regra da janela + CDF da Normal padrão:

$$P_\theta(2.6 < Z < 4) = P\left(\frac{2.6-\mu}{\sigma} < T < \frac{4-\mu}{\sigma}\right) = \Phi\left(\frac{4-\mu}{\sigma}\right) - \Phi\left(\frac{2.6-\mu}{\sigma}\right).$$

Passo 3 — Invariância do EMV (plug-in):

$$\boxed{\widehat{g}_{(c)} = \Phi\left(\frac{4-\widehat{\mu}_{MV}}{\widehat{\sigma}_{MV}}\right) - \Phi\left(\frac{2.6-\widehat{\mu}_{MV}}{\widehat{\sigma}_{MV}}\right)}$$

(d) $g(\theta) = \mathrm{Var}_\theta(Z)$

Variância. A **variância** de uma variável aleatória Z mede a dispersão dos valores em torno da média $E(Z)$. Formalmente, ela é definida por

$$\mathrm{Var}(Z) = E\big[(Z - E(Z))^2\big].$$

Definição expandida. Usando a propriedade de linearidade da esperança e a expansão do quadrado,

$$(Z - E(Z))^2 = Z^2 - 2Z\,E(Z) + E(Z)^2,$$

temos

$$\mathrm{Var}(Z) = E(Z^2) - 2E(Z)E(Z) + [E(Z)]^2 = E(Z^2) - [E(Z)]^2.$$

Portanto,

$$\boxed{\mathrm{Var}(Z) = E(Z^2) - [E(Z)]^2.}$$

Cálculo para a Normal. Se $Z \sim \mathcal{N}(\mu, \sigma^2)$, já sabemos que $E(Z) = \mu$. Logo precisamos calcular $E(Z^2)$.

Por definição:

$$E(Z^2) = \int_{-\infty}^{+\infty} z^2\, f(z \mid \mu, \sigma^2)\, dz, \quad \text{onde} \quad f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right].$$

Mudança de variável (padronização). Definimos $x = \frac{z-\mu}{\sigma}$, de modo que $z = \mu + \sigma x$ e $dz = \sigma dx$. Substituímos na integral:

$$E(Z^2) = \int_{-\infty}^{+\infty} (\mu + \sigma x)^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Expandindo o quadrado:

$$(\mu + \sigma x)^2 = \mu^2 + 2\mu\sigma x + \sigma^2 x^2.$$

Substituindo e separando termos:

$$E(Z^2) = \mu^2 \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx}_{=1} + 2\mu\sigma \underbrace{\int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}} e^{-x^2/2}\, dx}_{=0} + \sigma^2 \underbrace{\int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2}\, dx}_{=1}.$$

Usamos propriedades da Normal padrão $X \sim N(0,1)$:

$$E(X) = 0, \quad E(X^2) = 1.$$

Logo:

$$E(Z^2) = \mu^2 + \sigma^2.$$

Substituindo na definição da variância:

$$\mathrm{Var}(Z) = E(Z^2) - [E(Z)]^2 = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2.$$

Conclusão.

$$\boxed{\mathrm{Var}(Z) = \sigma^2.}$$

Assim, o parâmetro σ^2 da distribuição normal é, por definição, a variância populacional — ele controla a dispersão dos valores de Z em torno da média μ. **Fato conhecido da Normal:** se $Z \sim N(\mu, \sigma^2)$, então $\mathrm{Var}(Z) = \sigma^2$.

▷ **Justificativa curta:** pela padronização $Z = \mu + \sigma T$ com $T \sim N(0,1)$, usando *homogeneidade da variância* para fatores constantes,

$$\mathrm{Var}(Z) = \mathrm{Var}(\mu + \sigma T) = \sigma^2\, \mathrm{Var}(T) = \sigma^2 \cdot 1 = \sigma^2.$$

Logo, $g(\theta) = \sigma^2$ e, pela **invariância do EMV**, com o EMV já obtido para σ^2,

$$\boxed{\widehat{g}_{(d)} = \widehat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n} (Z_i - \overline{Z})^2}$$

(Recordando: para a Normal com μ desconhecido, o EMV de σ^2 usa divisor n, não $n-1$.)

Questão 04: Seja $(Z_1, \ldots, Z_n)$ uma amostra aleatória de $Z \sim f_\theta$, $\theta \in (0, \infty)$, tal que a função densidade de probabilidade é dada por

$$f_\theta(x) = \theta\, x^{\theta - 1}, \quad x \in (0,1),$$

e $f_\theta(x) = 0$, caso contrário.

(a) Encontre o EMV para $g(\theta) = E_\theta(Z)$.

(b) Encontre o EMV para $g(\theta) = P_\theta(Z > 0.3)$.

(c) Encontre o EMV para $g(\theta) = P_\theta(0 < Z < 0.1)$.

(d) Encontre o EMV para $g(\theta) = \mathrm{Var}_\theta(Z)$.

(e) Considere que os dados foram observados (0.12, 0.50, 0.20, 0.23, 0.30, 0.11). Encontre as estimativas de MV para os itens acima.

Objetivo. Procuramos uma densidade $f(x)$ em $(0,1)$, flexível o bastante para modelar proporções, com dois parâmetros de forma que controlem o comportamento perto de 0 e 1.

1) Escolha do núcleo. Uma família natural em $(0,1)$ é

$$f(x) \;\propto\; x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0<x<1,$$

com $\alpha>0,\ \beta>0$. *Justificativas:* (i) $x^{\alpha-1}$ controla a massa perto de 0; (ii) $(1-x)^{\beta-1}$ controla a massa perto de 1; (iii) quando $\alpha=\beta$ a forma é simétrica; (iv) a família é conjugada à Binomial.

Logo, tomamos

$$f(x) = C\,x^{\alpha-1}(1-x)^{\beta-1}\,\mathbf{1}_{(0,1)}(x),$$

e determinamos a constante de normalização C impondo $\int_0^1 f(x)\,dx = 1$.

2) Normalização (Integral Beta). Defina a *função Beta* de Euler

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx, \qquad \alpha,\beta>0.$$

Então

$$1 = \int_0^1 f(x)\,dx = C\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = C\,B(\alpha,\beta) \implies C = \frac{1}{B(\alpha,\beta)}.$$

Assim, a densidade fica

$$\boxed{f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}, \quad 0<x<1.}$$

3) Ligação com Γ. Recorde a função Gama: $\Gamma(t) = \displaystyle\int_0^\infty u^{t-1}e^{-u}\,du$. Vale a identidade clássica

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Logo,

$$\boxed{f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}, \quad 0<x<1.}$$

4) Construção alternativa (razão de Gammas). Se $U\sim\mathrm{Gamma}(\alpha,1)$ e $V\sim\mathrm{Gamma}(\beta,1)$ são independentes e definimos

$$X = \frac{U}{U+V}\in(0,1), \qquad T = U+V\in(0,\infty),$$

então $X\sim\mathrm{Beta}(\alpha,\beta)$.

Esboço da prova: O jacobiano da transformação $(u,v)\mapsto(x,t)$ é $\left|\frac{\partial(u,v)}{\partial(x,t)}\right| = t$. O conjunto transformado é $\{0<x<1,\ t>0\}$. A densidade conjunta de (X,T) é

$$f_{X,T}(x,t) = \frac{t^{\alpha+\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}e^{-t}.$$

Integrando t em $(0,\infty)$,

$$f_X(x) = \int_0^\infty f_{X,T}(x,t)\,dt = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}.$$

5) Momentos (fórmulas conhecidas). Para $X\sim\mathrm{Beta}(\alpha,\beta)$,

$$E[X] = \frac{\alpha}{\alpha+\beta}, \qquad \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

No caso particular Beta$(\theta, 1)$, obtém-se

$$E[X] = \frac{\theta}{\theta+1}, \qquad \mathrm{Var}(X) = \frac{\theta}{(\theta+1)^2(\theta+2)}.$$

Distribuição Beta. A distribuição Beta é contínua no intervalo $(0, 1)$, com densidade

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

onde $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ é a função Beta. No caso particular Beta$(\theta, 1)$,

$$f_\theta(x) = \theta\, x^{\theta-1}, \quad 0 < x < 1,$$

e a normalização é verificada por

$$\int_0^1 \theta x^{\theta-1}\, dx = 1.$$

Modelo. Se $X \sim$ Beta$(\theta, 1)$, então a densidade é

$$f_\theta(x) = \theta\, x^{\theta-1}\, \mathbf{1}_{(0,1)}(x), \qquad \theta > 0.$$

Esperança $E[X]$. Usamos a definição (caso contínuo) e regras elementares de integração:

$$E[X] = \int_0^1 x\, f_\theta(x)\, dx = \int_0^1 x\, \theta\, x^{\theta-1}\, dx = \theta \int_0^1 x^\theta\, dx.$$

Propriedade usada: $\displaystyle\int_0^1 x^a\, dx = \frac{1}{a+1}$ para $a > -1$.

Como $\theta > 0$, segue que

$$E[X] = \theta \cdot \frac{1}{\theta+1} = \boxed{\frac{\theta}{\theta+1}}.$$

Segundo momento $E[X^2]$. De modo análogo,

$$E[X^2] = \int_0^1 x^2 f_\theta(x)\, dx = \int_0^1 x^2\, \theta\, x^{\theta-1}\, dx = \theta \int_0^1 x^{\theta+1}\, dx = \theta \cdot \frac{1}{\theta+2} = \boxed{\frac{\theta}{\theta+2}}.$$

Variância $\mathrm{Var}(X)$. Pela definição, $\mathrm{Var}(X) = E[X^2] - (E[X])^2$:

$$\mathrm{Var}(X) = \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 = \frac{\theta(\theta+1)^2 - \theta^2(\theta+2)}{(\theta+2)(\theta+1)^2}.$$

Álgebra no numerador:

$$\theta(\theta+1)^2 - \theta^2(\theta+2) = \theta\big(\theta^2 + 2\theta + 1\big) - \theta^2(\theta+2) = \theta^3 + 2\theta^2 + \theta - \theta^3 - 2\theta^2 = \theta.$$

Logo,

$$\mathrm{Var}(X) = \boxed{\frac{\theta}{(\theta+1)^2(\theta+2)}}.$$

O parâmetro θ controla a concentração: valores $\theta < 1$ favorecem x próximos de 0, $\theta > 1$ favorecem x próximos de 1, e $\theta = 1$ dá a distribuição uniforme.

Função de verossimilhança. Como as observações são independentes e identicamente distribuídas, a densidade conjunta é o produto das individuais:

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \theta x_i^{\theta-1}.$$

Aplicando a propriedade distributiva do produto:

$$L(\theta; x) = \theta^n \prod_{i=1}^{n} x_i^{\theta-1} = \theta^n \exp\left\{(\theta-1)\sum_{i=1}^{n} \log x_i\right\}.$$

—

Função log-verossimilhança. Usando as propriedades dos logaritmos:

$$\begin{aligned} &\text{(i)}\ \log(ab) = \log a + \log b, \\ &\text{(ii)}\ \log(a^b) = b \log a, \\ &\text{(iii)}\ \log\Big(\prod_i a_i\Big) = \sum_i \log a_i, \end{aligned}$$

temos:

$$\ell(\theta) = \log L(\theta; x) = n \log \theta + (\theta - 1)\sum_{i=1}^{n} \log x_i.$$

—

Derivada primeira. Aplicando as propriedades diferenciais:

$$\frac{d}{d\theta} \log \theta = \frac{1}{\theta}, \qquad \frac{d}{d\theta}\big[(\theta-1)c\big] = c,$$

obtemos:

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i.$$

—

Ponto crítico (máxima verossimilhança). Igualando a derivada a zero:

$$\ell'(\theta) = 0 \iff \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0 \iff \boxed{\widehat{\theta}_{MV} = -\frac{n}{\sum_{i=1}^{n} \log x_i}}.$$

Como $0 < x_i < 1$, temos $\log x_i < 0$, logo o estimador é positivo.

—

Segunda derivada.

$$\ell''(\theta) = \frac{d}{d\theta}\Big(\frac{n}{\theta}\Big) = -\frac{n}{\theta^2} < 0,$$

garantindo que o ponto crítico é um *máximo.*

—

Resultado final (EMV).

$$\boxed{\widehat{\theta}_{MV} = -\frac{n}{\sum_{i=1}^{n} \log X_i}}, \qquad X_i \in (0,1).$$

—

Observações.

- O estimador de máxima verossimilhança é sempre positivo, pois $\log X_i < 0$.
- A estatística suficiente é $T = -\sum_{i=1}^n \log X_i$, que segue uma distribuição Gamma$(n, \text{taxa} = \theta)$.
- Pela *invariância do EMV*, para qualquer função $g(\theta)$,

$$\widehat{g}_{MV} = g(\widehat{\theta}_{MV}).$$

Por exemplo, para $g(\theta) = P_\theta(X > c) = 1 - c^\theta$, tem-se

$$\widehat{g}_{MV} = 1 - c^{\widehat{\theta}_{MV}}.$$

Questão 04. Seja $(Z_1, \ldots, Z_n)$ uma amostra i.i.d. de densidade

$$f_\theta(x) = \theta\, x^{\theta-1}, \qquad x \in (0,1), \quad \theta > 0,$$

e $f_\theta(x) = 0$ caso contrário. (Trata-se de uma Beta$(\theta, 1)$.)

EMV de θ. A verossimilhança é

$$L(\theta; \mathbf{z}) = \prod_{i=1}^n \theta\, z_i^{\theta-1} = \theta^n \prod_{i=1}^n z_i^{\theta-1} = \theta^n \Big(\prod_{i=1}^n z_i\Big)^{\theta-1}.$$

A log-verossimilhança é

$$\ell(\theta) = \log L(\theta; \mathbf{z}) = n \log \theta + (\theta - 1) \sum_{i=1}^n \log z_i.$$

Derivando e igualando a zero,

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^n \log z_i = 0 \quad \Longrightarrow \quad \widehat{\theta}_{MV} = -\frac{n}{\sum_{i=1}^n \log Z_i}.$$

A segunda derivada é

$$\ell''(\theta) = -\frac{n}{\theta^2} < 0 \quad (\theta > 0),$$

logo o ponto crítico é máximo global. Assim,

$$\boxed{\widehat{\theta}_{MV} = -\frac{n}{\sum_{i=1}^n \log Z_i}}.$$

Momentos e probabilidades. Para $X \sim f_\theta$, para $a > -\theta$,

$$\mathbb{E}_\theta(X^a) = \int_0^1 x^a\, \theta x^{\theta-1} dx = \theta \int_0^1 x^{a+\theta-1} dx = \frac{\theta}{a+\theta}.$$

Em particular,

$$\mathbb{E}_\theta(X) = \frac{\theta}{\theta+1}, \qquad \mathbb{E}_\theta(X^2) = \frac{\theta}{\theta+2}, \qquad \mathrm{Var}_\theta(X) = \frac{\theta}{(\theta+1)^2(\theta+2)}.$$

Para $0 < a < b \leq 1$,

$$P_\theta(a < X < b) = \int_a^b \theta x^{\theta-1} dx = \Big[x^\theta\Big]_a^b = b^\theta - a^\theta.$$

(a) $g(\theta) = \mathbb{E}_\theta(Z)$.

$$\mathbb{E}_\theta(Z) = \frac{\theta}{\theta+1} \quad \Longrightarrow \quad \boxed{\widehat{g}_{(a)} = \frac{\widehat{\theta}_{MV}}{\widehat{\theta}_{MV} + 1}}.$$

(b) $g(\theta) = P_\theta(Z > 0.3)$**.**

$$P_\theta(Z > 0.3) = 1 - P_\theta(0 < Z \le 0.3) = 1 - (0.3)^\theta \quad \Longrightarrow \quad \boxed{\widehat{g}_{(b)} = 1 - (0.3)^{\widehat{\theta}_{MV}}}.$$

(c) $g(\theta) = P_\theta(0 < Z < 0.1)$**.**

$$P_\theta(0 < Z < 0.1) = (0.1)^\theta \quad \Longrightarrow \quad \boxed{\widehat{g}_{(c)} = (0.1)^{\widehat{\theta}_{MV}}}.$$

(d) $g(\theta) = \mathrm{Var}_\theta(Z)$**.**

$$\mathrm{Var}_\theta(Z) = \frac{\theta}{(\theta+1)^2(\theta+2)} \quad \Longrightarrow \quad \boxed{\widehat{g}_{(d)} = \frac{\widehat{\theta}_{MV}}{(\widehat{\theta}_{MV}+1)^2(\widehat{\theta}_{MV}+2)}}.$$

(e) Estimativas numéricas. Dados: (0.12, 0.50, 0.20, 0.23, 0.30, 0.11). Com $n = 6$,

$$\sum_{i=1}^{n} \log z_i = \log(0.12) + \log(0.50) + \log(0.20) + \log(0.23) + \log(0.30) + \log(0.11) \approx -9.3038.$$

Logo,

$$\widehat{\theta}_{MV} = -\frac{6}{-9.3038} \approx 0.6449.$$

Então,

$$\widehat{g}_{(a)} = \frac{\widehat{\theta}}{1+\widehat{\theta}} \approx \frac{0.6449}{1.6449} \approx 0.3921,$$

$$\widehat{g}_{(b)} = 1 - (0.3)^{\widehat{\theta}} \approx 1 - (0.3)^{0.6449} \approx 0.5400,$$

$$\widehat{g}_{(c)} = (0.1)^{\widehat{\theta}} \approx (0.1)^{0.6449} \approx 0.2265,$$

$$\widehat{g}_{(d)} = \frac{\widehat{\theta}}{(\widehat{\theta}+1)^2(\widehat{\theta}+2)} \approx \frac{0.6449}{(1.6449)^2(2.6449)} \approx 0.0901.$$

Resumo: $\widehat{\theta}_{MV} = -n/\sum \log Z_i \approx 0.6449$ e, por invariância, as estimativas de (a)–(d) são os valores de $g(\theta)$ avaliados em $\widehat{\theta}$ como mostrado acima.