PPGEST2220, Prova 1, 2/out/2025. Entregar a solução manuscrita até 13/out/2025 (em mãos ou depositar no escaninho do professor). Outros meios não serão aceitos.

Primeiro Exercício

Uma amostra aleatória $X_1, \ldots, X_n$ é retirada de uma população com média μ e desvio padrão 1, sendo que os elementos dessa amostra podem ser correlacionados como $\text{Corr}(X_i, X_j) = \rho_{ij}$, em que $-1 \leq \rho_{ij} \leq 1$. Com respeito a essa situação, determine as condições suficientes tais que a média amostral, definida como $\bar{X}_n = \sum_{i=1}^n X_i/n$, seja estimador consistente para μ.

Segundo Exercício

Considerando uma amostra aleatória simples $X_1, \ldots, X_n$ retirada de uma distribuição absolutamente contínua com função de densidade desconhecida $f(x)$, tal que $||f||_\infty = \sup_{x \in \mathbb{R}} f(x) \leq M < \infty$, tome seu estimador

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

para qualquer $x \in \mathbb{R}$, em que $h_n > 0$ tal que $h_n \downarrow 0$ e $nh_n \uparrow \infty$ à medida de n aumenta, e $K \in L^2(\mathbb{R})$ é uma função real não negativa que possui as mesmas propriedades de uma função de densidade simétrica em torno de zero. Com base nessas informações, mostre que, para todo ponto de continuidade x de f, $\hat{f}(x) \xrightarrow{P} f(x)$.

Terceiro Exercício

[Matsushita *et al.*, 2026, Assessing goodness-of-fit for sparse categories using Rényi divergence, Journal of Statistical Planning and Inference, 242, 106350] Seja $\hat{p}_1, \ldots, \hat{p}_m$ uma distribuição de probabilidade empírica obtida por meio de uma amostra aleatória simples de tamanho n, na qual $\sum_{k=1}^m \hat{p}_k = 1$. Uma forma de medir a discrepância entre a distribuição empírica e uma distribuição de probabilidade hipotética denotada como $p_1, \ldots, p_m$ seria estatística qui-quadrado, definida como

$$\chi^2 = n \sum_{k=1}^{m} \frac{(\hat{p}_k - p_k)^2}{p_k}.$$

Para n suficientemente grande, essa estatística segue aproximadamente uma distribuição qui-quadrado com $m - 1$ graus de liberdade. Embora essa estatística apresente boas propriedades estatísticas, ela não é recomendada se houver pelo menos uma frequência hipotética p_k muito baixa, digamos, menor que $5/n$. Como uma alternativa, considere a divergência de Rényi de ordem α da distribuição empírica relativamente à hipotética dada por

$$R_\alpha = \frac{1}{\alpha - 1} \log \sum_{k=1}^{m} p_k^\alpha \hat{p}_k^{1-\alpha},$$

na qual $\alpha > 0$, sendo que

$$\lim_{\alpha \to 1} R_\alpha = \sum_{k=1}^{m} p_k \log \frac{p_k}{\hat{p}_k}.$$

Com a ajuda do método delta, estabeleça uma relação assintótica entre R_α e a estatística χ^2.

Quarto Exercício

Seja $X_1, \ldots, X_n$ uma amostra aleatória simples retirada de uma população contínua com mediana $\tilde{\mu}$ e função de distribuição de probabilidade acumulada $F(x)$. Se n for ímpar, definindo a mediana amostral como $\tilde{X} = X_{([n+1]/2)}$, mostre que $\tilde{X} \xrightarrow{P} \tilde{\mu}$.

Quinto Exercício

Considere uma amostra aleatória simples $X_1, \ldots, X_n$ retirada de uma população Pareto simétrica na forma

$$f(x) = \frac{\alpha \mu^\alpha}{2(\mu + |x - \mu|)^{\alpha+1}},$$

em que $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ e $\alpha > 2$. Se $\tilde{X}$ e $\bar{X}$ representam, respectivamente, a mediana e a média amostrais, explique qual desses estimadores propicia menor erro quadrático médio.

Sexto Exercício

Considerando uma amostra aleatória simples $X_1, \ldots, X_n$ retirada de uma distribuição $\mathrm{N}(\mu, \sigma^2)$, compare os erros quadráticos médios (EQMs) de $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ e $V^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$. Qual desses estimadores de σ^2 seria mais vantajoso em termos desses EQMs?

Setimo Exercício

[Matsushita *et al.*, 2019, The Touchard distribution, Communications in Statistics - Theory and Methods, 48:8, 2049-2059] Considere a distribuição de Touchard definida como

$$\mathrm{P}(X = x; \lambda, \delta, \nu) = \frac{\lambda^x (x+\nu)^\delta}{x! \tau(\lambda, \delta, \nu)},$$

em que $\lambda > 0$ é a taxa, $\delta \in \mathbb{R}$ é o parâmetro de forma, ν é um parâmetro de deslocamento que assume valores $1, 2, 3, \ldots$, $x \in \mathbb{N}$ e $\tau(\lambda, \delta, \nu) = \sum_{k \in \mathbb{N}} \frac{\lambda^k (k+\nu)^\delta}{k!}$ é a função de normalização.

i) Se $\theta_1 = (\lambda, \delta)'$ representa o vetor de parâmetros desconhecidos (ou seja, considerando ν conhecido), reescreva essa distribuição na forma da família exponencial e determine a estatística suficiente minimal para o vetor θ_1.

ii) Se $\theta_2 = (\lambda, \delta, \nu)'$ representa o vetor de parâmetros desconhecidos, discuta se haveria estatística suficiente para o vetor θ_2. Em caso afirmativo, determine a forma dessa estatística.

iii) Se $\delta = 0$ e $\nu = 1$ e se $\theta_3 = \lambda$ representa o parâmetro desconhecido, discuta se haveria estatística suficiente minimal para θ_3. Em caso afirmativo, determine a forma dessa estatística.

Oitavo Exercício

Considere uma amostra aleatória simples $X_1, \ldots, X_n$ retirada de uma população com densidade $f_\theta(x) = \exp[-x + \theta]$, em que $x > \theta > 0$. Determine a estatística suficiente, completa e minimal, e mostre que ela é independente da variância amostral S^2.