Exercícios: Métodos de Estimação (Momentos, Máxima Verossimilhança)

Disciplina: Inferência Estatística (MSc)
Instrutor: Luiz Carvalho
Monitor: Isaque Pim

Novembro/2023

Motivação: Em aula vimos os métodos de momentos, de mínimos quadrados e de máxima verossimilhança para obter estimativas (e estimadores!) de parâmetros e quantidades de interesse. Nesta lista você vai praticar encontrar esses estimadores e entender suas propriedades.

Notação: Como convenção adotamos $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_+ = (0, \infty)$ e $\mathbb{N} = \{1, 2, \ldots\}$.

Dos livros-texto:

a) KN, Ch9: 17a, 21a, 21b.

b) CB, Ch7: 7.1, 7.2[1], 7.6, 7.11, 7.12, 7.19, 7.37, 7.38.

Extra:

1. **O equilíbrio de Hardy-Weinberg**[2] Considere uma amostra de uma população em equilíbrio genético com respeito a um único gene que tem dois alelos (A e a). Há três configurações possíveis, chamadas genótipos:

   - Homozigoto recessivo: aa;
   - Heterozigoto: Aa;
   - Homozigoto dominante: AA.

   Se assumirmos que os três genótipos são identificáveis, podemos supor que os três tipos de indivíduos occorem na população seguindo o *equilíbrio de Hardy-Weinberg*. Seja X a variável aleatória que guarda o genótipo do indivíduo. Então podemos escrever as probabilidades

$$\begin{aligned}
\Pr(X = aa) &:= p_1 = \theta^2, \\
\Pr(X = Aa) &:= p_2 = 2\theta(1-\theta), \\
\Pr(X = AA) &:= p_3 = (1-\theta)^2,
\end{aligned}$$

[1] Se precisar, peça ajuda ao monitor para a parte computacional.

[2] Godfrey Harold Hardy (1877-1947) foi um grande matemático Britânico famoso por dizer que seu trabalho não tinha a menor relevância para a vida real. Hardy estava errado, é claro. Wilhelm Weinberg (1862-1937) foi um médico alemão que descobriu o princípio do equilíbrio genético independentemente (e antes) de Hardy.

para $\theta \in (0,1)$. Tome uma amostra de tamanho n da população e seja N_i o número de indivíduos do genótipo i amostrados, para $i = 1, 2, 3$, de sorte que $n = \sum_{i=1}^{3} N_i$.

(a) Deduza a distribuição de $\boldsymbol{N} = (N_1, N_2, N_3)$;

(b) Mostre como estimar θ por máxima verossimilhança;

(c) Discuta como estimar θ pelo método dos momentos;

(d) Avalie $\delta_s = N_1/n + N_2/n$ como estimador de θ em relação aos anteriores.

2. Tome $n \geq 2$ e $X_1, X_2, \ldots, X_n$ amostra aleatória de uma distribuição com densidade comum com respeito a Lebesgue

$$p_\theta(x) = \frac{\theta}{x^2} \mathbb{I}\left(x \in [\theta, \infty)\right),$$

com $\theta \in \mathbb{R}_+$.

- Compute o EMV de θ. Este estimador é consistente? Podemos aplicar os teoremas discutidos em aula?
- Compute o viés do estimador de $g(\theta) = E_\theta[X^{-1}]$ como função do tamanho de amostra n e discuta o que acontece quando $n \to \infty$.

3. **PhD**: Tome $\mathcal{P}$ família dominada com densidade com respeito à medida de contagem em $\mathbb{N}$:

$$p_\theta(x) = \frac{\theta}{x^{\theta+1}} \mathbb{I}(x \geq 1),$$

para $x \in \mathbb{N}$ e $\theta > 1$. Encontre estimadores de (i) momentos e (ii) máxima verossimilhança para θ e compute (e compare) suas variâncias. Utilize métodos de Monte Carlo, se precisar.

4. **PhD**: Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma família dominada, cuja densidade comum (com respeito a Lebesgue) é

$$f_\theta(x) = \frac{\exp(-|x-\theta|)}{2} \mathbb{I}(x \in \mathbb{R}),\ \theta \in \mathbb{R}.$$

Suponha que n é par e encontre o estimator de máxima verossimilhança para $g(\theta) = \sin(\theta)/\theta$. Que peculiaridades tem esse estimador?

Dica: Considere escrever a função de verossimilhança em termos da função sinal:

$$\operatorname{sgn}(x) = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ 1, & x > 0. \end{cases}$$

para $x \in \mathbb{R}$.