Point estimation — Exercises

Statistics II

Szabolcs Sebestyén
szsebe@fcee.ucp.pt

Faculdade de Ciências Económicas e Empresariais
Universidade Católica Portuguesa

Example

Considere uma variável aleatória X com média μ e variância σ^2. Considere os seguintes estimadores para μ obtidos a partir de uma amostra aleatória de dimensão N:

$$\hat{\mu}_1 = \overline{X} + \frac{80}{N} \qquad \hat{\mu}_2 = \overline{X}\,\frac{N}{N-1}.$$

a. Mostre que ambos os estimadores são enviesados para μ.

Solution

$$\begin{aligned} \mathbb{E}\left(\hat{\mu}_1\right) &= \mathbb{E}\left(\overline{X}\right) + \frac{80}{N} = \mu + \frac{80}{N} \neq \mu \\ \mathbb{E}\left(\hat{\mu}_2\right) &= \mathbb{E}\left(\overline{X}\right)\frac{N}{N-1} = \mu\,\frac{N}{N-1} \neq \mu \end{aligned}$$

Example

- **b** Qual o estimador que tem um maior enviesamento (bias) se $N = 20$ e $\mu = 70$?

Solution

$$\begin{aligned} \text{bias}\left(\hat{\mu}_1\right) &= \mathbb{E}\left(\hat{\mu}_1\right) - \mu = \frac{80}{N} = 4 \\ \text{bias}\left(\hat{\mu}_2\right) &= \mathbb{E}\left(\hat{\mu}_2\right) - \mu = \mu\left(\frac{N}{N-1} - 1\right) = \frac{\mu}{N-1} = 3.6842 \end{aligned}$$

Assim, $\text{bias}\left(\hat{\mu}_2\right) < \text{bias}\left(\hat{\mu}_1\right)$.

Example

- Obtenha o erro quadrático médio para ambos os estimadores. Se $N = 20$, $\mu = 70$ e $\sigma^2 = 100$, por qual dos dois estimadores optaria?

Solution

$$\text{MSE}\left(\hat{\mu}_1\right) = \mathbb{V}\text{ar}\left(\hat{\mu}_1\right) + \left[\text{bias}\left(\hat{\mu}_1\right)\right]^2$$

$$\mathbb{V}\text{ar}\left(\hat{\mu}_1\right) = \mathbb{V}\text{ar}\left(\overline{X}\right) = \frac{\sigma^2}{N} = \frac{100}{20} = 5$$

Então, $\text{MSE}\left(\hat{\mu}_1\right) = 5 + 4^2 = 21$.

$$\mathbb{V}\text{ar}\left(\hat{\mu}_2\right) = \frac{N^2}{\left(N-1\right)^2}\mathbb{V}\text{ar}\left(\overline{X}\right) = \frac{20^2}{19^2} \cdot 5 = 5.54$$

Então, $\text{MSE}\left(\hat{\mu}_2\right) = 5.54 + 3.6842^2 = 19.11$
Optaria pelo estimador $\hat{\mu}_2$ *pois,* $\text{MSE}\left(\hat{\mu}_2\right) < \text{MSE}\left(\hat{\mu}_1\right)$.

Example

Considere-se uma determinada variável aleatória, X, com distribuição de Bernoulli, isto é, $X \sim \mathrm{b}(1; p)$. Foi proposto o seguinte estimador para p a partir de uma amostra aleatória de dimensão N:

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_{N-1}}{N-1}.$$

a. Verifique se este é um estimador centrado para p.

Solution

Como X é Bernoulli, $\mathbb{E}(X) = p$ e $\mathbb{V}\mathrm{ar}(X) = p(1-p)$

$$\begin{aligned}\mathbb{E}(\hat{p}) &= \mathbb{E}\left(\frac{X_1 + X_2 + \cdots + X_{N-1}}{N-1}\right) = \frac{1}{N-1}\mathbb{E}(X_1 + \cdots + X_{N-1}) = \\ &= \frac{1}{N-1}(N-1)\,p = p\end{aligned}$$

Conclusão: O estimador proposto é um estimador centrado para p

Example

(b) Será um estimador consistente para p?

Solution

Condições suficientes de consistência

1. $\lim_{N\to\infty} \mathbb{E}\left(\hat{p}\right) = \lim_{N\to\infty} p = p$
2. $$\begin{aligned}\lim_{N\to\infty} \mathbb{V}\mathrm{ar}\left(\hat{p}\right) &= \lim_{N\to\infty} \mathbb{V}\mathrm{ar}\left(\frac{X_1 + X_2 + \cdots + X_{N-1}}{N-1}\right) = \\ &= \lim_{N\to\infty} \frac{1}{\left(N-1\right)^2}\left(N-1\right)p\left(1-p\right) = \\ &= \lim_{N\to\infty} \frac{p\left(1-p\right)}{N-1} = 0\end{aligned}$$

Conclusão: O estimador proposto é um estimador consistente para p

Example

- c. Compare-o quanto à eficiência com o seguinte estimador alternativo para p:

$$\tilde{p} = \frac{X_1 + X_2 + \cdots + X_{N-2}}{N-2}.$$

Solution

$$\begin{aligned}
\mathbb{E}(\tilde{p}) &= \mathbb{E}\left(\frac{X_1 + \cdots + X_{N-2}}{N-2}\right) = \frac{1}{N-2}\mathbb{E}(X_1 + \cdots + X_{N-2}) = \\
&= \frac{1}{N-2}(N-2)\,p = p
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(\tilde{p}) &= \mathbb{V}\mathrm{ar}\left(\frac{X_1 + \cdots + X_{N-2}}{N-2}\right) = \frac{1}{(N-2)^2}(N-2)\,p\,(1-p) = \\
&= \frac{p\,(1-p)}{N-2}
\end{aligned}$$

Solution (cont.)

Ambos são estimadores centrados para p e $\mathbb{V}\text{ar}(\tilde{p}) > \mathbb{V}\text{ar}(\hat{p}) \Longrightarrow \hat{p}$ é relativamente mais eficiente que $\tilde{p}$

Example

Num determinado local da costa portuguesa e sob certas condições climatéricas a altura das ondas do mar é uma variável aleatória com função densidade de probabilidade dada por

$$f(x) = \begin{cases} \dfrac{x}{\alpha^2}\, e^{-\frac{1}{2}\left(\frac{x}{\alpha}\right)^2} & x \geq 0 \text{ com } \alpha > 0 \\ 0 & x < 0 \end{cases}$$

sendo $\mathbb{E}(X) = \alpha\sqrt{\frac{\pi}{2}}$ e $\mathbb{V}\text{ar}(X) = \alpha^2\left(2 - \frac{\pi}{2}\right)$.

Foram registadas as seguintes alturas (em metros) de 12 ondas:

1.4, 3.5, 2.4, 1.9, 3.1, 2.7, 2.5, 3.1, 4.1, 2.8, 2.5, 3.3.

Obtenha a estimativa da máxima verosimilhança para o valor esperado de X e para a variância de X.

(*Nota: não necessita resolver a condição de segunda ordem, sendo suficiente enunciá-la*).

Solution

$$\mathcal{L}(\alpha) = \prod_{i=1}^{N} f(x_i;\alpha) = \prod_{i=1}^{N} \frac{x_i}{\alpha^2} \exp\left\{-\frac{1}{2}\frac{x_i^2}{\alpha^2}\right\} =$$
$$= \alpha^{-2N}\left(\prod_{i=1}^{N} x_i\right) \exp\left\{-\frac{1}{2\alpha^2}\sum_{i=1}^{N} x_i^2\right\}$$

$$\ln \mathcal{L}(\alpha) = \sum_{i=1}^{N} \ln(x_i) - 2N\ln(\alpha) - \frac{1}{2\alpha^2}\sum_{i=1}^{N} x_i^2$$

$$\frac{\mathrm{d}\ln \mathcal{L}(\alpha)}{\mathrm{d}\alpha} = \frac{-2N}{\alpha} - \frac{-2}{2\alpha^3}\sum_{i=1}^{N} x_i^2 = \frac{1}{\alpha^3}\left(-2N\alpha^2 + \sum_{i=1}^{N} x_i^2\right) = 0$$

$$\hat{\alpha} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{2N}}$$

Solution (cont.)

Assim, pela propriedade da invariância dos estimadores de máxima verosimilhança,

$$\widehat{\mathbb{E}(X)} = \hat{\alpha}\sqrt{\frac{\pi}{2}} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{2N} \times \frac{\pi}{2}}$$

e

$$\widehat{\mathbb{V}\mathrm{ar}(X)} = \hat{\alpha}^2\left(2 - \frac{\pi}{2}\right) = \frac{\sum_{i=1}^{N} x_i^2}{2N}\left(2 - \frac{\pi}{2}\right)$$

Como $N = 12$ *e* $\sum_{i=1}^{N} x_i^2 = 98.13$, *tem-se:*

$$\hat{\alpha} = 2.022 \qquad \widehat{\mathbb{E}(X)} = 2.5343 \qquad \widehat{\mathbb{V}\mathrm{ar}(X)} = 1.7549$$