Méthodes d’analyse biostatistique projet 1

Wen, Zehai

2023-10-07 16:49:29-04:00

TEXT

0.1 Exercise 1

Fix $\alpha > 0$. For a $\theta \in (0, \alpha)$, let $\{X_i\}_{i=1}^n$ be a sequence of real independent identically distributed random variable defined on some probability space (Ω, F, P) with common probability density function with respect to the Lebesgue measure:

$$f_\theta(x) = \begin{cases} \frac{2x}{\alpha\theta} & x \in [0, \theta] \\ \frac{2(\alpha - x)}{\alpha(\alpha-\theta)} & x \in [\theta, \alpha] \\ 0 & \text{otherwise} \end{cases}$$

Prove that the maximum likelihood estimation of θ must be one of the given observation but not necessarily any particular observation. In case $\alpha = 5$ and $n = 3$, compute the maximum likelihood estimate of θ when the observations are $(1, 2, 4)$ or $(2, 3, 4)$.

Solution Let $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ be given. Write $g_\theta(x)$ for the joint probability density function of X_i's and $x_{(i)}$ for the ith smallest coordiante in x. If $x_i < 0$ or $x_i > \alpha$ for some i, then $g_\theta(x) = 0$ for any θ so that any estimate would be a maximum likelihood estimate. We exclude this pathology and prove that:

Théorème 1. *If $0 \leqslant x_{(1)} \leqslant \cdots \leqslant x_{(n)} \leqslant \alpha$, then $\theta_0 = x_{(i)}$ for some i.*

Proof. It never happens that $\theta_0 < x_{(1)}$ because $g_{\theta_1}(x) > g_{\theta_0}(x)$ whenever $\theta_1 \in (\theta_0, x_{(1)})$. Similarly, it never happens that $\theta_0 > x_{(n)}$ because $g_{\theta_1}(x) > g_{\theta_0}(x)$ whenever $\theta_1 \in (x_{(n)}, \theta_0)$. We assume from now on that $\theta_0 \in [x_{(i)}, x_{(i+1)}]$ for some $i \in \{1, \cdots, n-1\}$. Suppose, for the sake of contradiction, that $x_{(i)} < \theta_0 < x_{(i+1)}$. We have:

$$g_{\theta_{\theta_0}}(x) = \left(\frac{2}{\alpha}\right)^n \frac{x_{(1)}}{\theta_0} \cdots \frac{x_{(i)}}{\theta_0} \frac{\alpha - x_{(i+1)}}{\alpha - \theta_0} \cdots \frac{\alpha - x_{(n)}}{\theta_0}$$

The numerator does not depend on θ_0. This motivates us to define function $h : [x_{(i)}, x_{(i+1)}] \to \mathbb{R}$ by:

$$h(\theta) = \frac{1}{\theta^i} \frac{1}{(\alpha - \theta)^{n-i}}$$

Then the second derivative is:

$$h''(\theta) = i(i+1)\theta^{-i-2}(\alpha - \theta)^{i-n} + (n-i)(n-i+1)\theta^{-i}(\alpha - \theta)^{i-n-2} > 0$$

Therefore, h is strictly convex and the maximum can only be at the boundary points. $\square$

We now demonstrate that the choice of i is not unique in the above theorem. The simplest case will be $x_i = x_j$ for any i and any j. For a nontrivial example, let $\alpha = 5$, $n = 3$. If $x = (2, 3, 4)$, then the maximum likelihood estimate is one of $\{2, 3, 4\}$. An estimate of 3 or 4 yields maximum likelihood $\frac{8}{375}$ while an estimate of 2 yields likelihood $\frac{16}{1125}$. Therefore, the maximum likelihood estimate can be 3 or 4 and is not unique.

Finally, the additional example $x = (1, 2, 4)$, estimate $\theta = 1, 2, 4$ gives likelihood $\frac{3}{250}, \frac{4}{375}, \frac{1}{125}$ repsectively. We conclude that $\theta = 1$ is the maximum likelihood estimate in this case.////

TEXT