Large-sample approximate confidence interval; example 2, Rayleigh distribution

Spatial Organization of Chromosome (Rice Problem 8.45)

In a previous example, we analyzed data from experiments that were conducted to learn about the spatial organization of chromosomes. We had n measurements of the distances between pairs of small DNA sequences, which we modeled as $X_i \overset{\text{i.i.d.}}{\sim} \text{Rayleigh}(\theta)$.

In that example, we found the maximum likelihood estimator and estimate of θ; now, we will find an approximate confidence interval for θ.

Here are some results we had previously:

If $X \sim \text{Rayleigh}(\theta)$ (with parameter $\theta > 0$), then the probability density function is given by

$$f(x|\theta) = \frac{x}{\theta} \exp\left(\frac{-x^2}{2\theta}\right)$$

for positive values of x (and the probability density function is 0 for non-positive values of x).

We have the following results about the moments of a Rayleigh-distributed random variable:

$$\begin{aligned} E(X) &= \left(\frac{\theta\pi}{2}\right)^{1/2} \\ E(X^2) &= 2\theta \\ Var(X) &= 2\theta - \frac{\theta\pi}{2} \end{aligned}$$

The likelihood is given by

$$\begin{aligned} \mathcal{L}(\theta|x_1, \ldots, x_n) &= f(x_1, \ldots, x_n|\theta) \\ &= \prod_{i=1}^{n} f(x_i|\theta) \\ &= \prod_{i=1}^{n} \frac{x_i}{\theta} \exp\left(\frac{-x_i^2}{2\theta}\right) \end{aligned}$$

The log-likelihood is therefore

$$\begin{aligned} \ell(\theta|x_1, \ldots, x_n) &= \log\left\{\prod_{i=1}^{n} \frac{x_i}{\theta} \exp\left(\frac{-x_i^2}{2\theta}\right)\right\} \\ &= \sum_{i=1}^{n} \left\{\log(x_i) - \log(\theta) - \frac{x_i^2}{2\theta}\right\} \\ &= \sum_{i=1}^{n} \log(x_i) - n\log(\theta) - \frac{1}{2\theta}\sum_{i=1}^{n} x_i^2 \end{aligned}$$

The first derivative of the log-likelihood is

$$\begin{aligned}\frac{d}{d\theta}\ell(\theta|x_1,\ldots,x_n) &= \frac{d}{d\theta}\left\{\sum_{i=1}^n \log(x_i) - n\log(\theta) - \frac{1}{2\theta}\sum_{i=1}^n x_i^2\right\} \\ &= \frac{-n}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2\end{aligned}$$

Setting this equal to 0 and solving for theta, we obtain a maximum likelihood estimator of $\hat{\theta}_{MLE} = \frac{1}{2n}\sum_{i=1}^n X_i^2$.

The second derivative of the log-likelihood is:

$$\begin{aligned}\frac{d^2}{d\theta^2}\ell(\theta|x_1,\ldots,x_n) &= \frac{d}{d\theta}\left\{\frac{-n}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2\right\} \\ &= \frac{-n}{\theta^2}(-1) + \frac{1}{2\theta^3}(-2)\sum_{i=1}^n x_i^2 \\ &= \frac{n}{\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^n x_i^2\end{aligned}$$

1. Find an expression for the observed Fisher information as a function of θ^* (not much work to do here...).

$J(\theta^*) = -\frac{n}{(\theta^*)^2} + \frac{1}{(\theta^*)^3}\sum_{i=1}^n x_i^2$

2. Find an expression for the Fisher information as a function of θ^*. When you evaluate at θ^*, it will be most convenient if you first evaluate the expected value and then plug in θ^* for θ. You can do this because the Rayleigh distribution pdf satisfies the regularity conditions necessary interchange differentiation and integration.

$$\begin{aligned}I(\theta^*) &= E\left[-\frac{n}{(\theta)^2} + \frac{1}{(\theta)^3}\sum_{i=1}^n X_i^2\right]|_{\theta=\theta^*} \\ &= \left[-\frac{n}{(\theta)^2} + \frac{1}{(\theta)^3}\sum_{i=1}^n E(X_i^2)\right]|_{\theta=\theta^*} \\ &= \left[-\frac{n}{(\theta)^2} + \frac{1}{(\theta)^3}n2\theta\right]|_{\theta=\theta^*} \\ &= -\frac{n}{(\theta^*)^2} + \frac{2n}{(\theta^*)^2} \\ &= \frac{n}{(\theta^*)^2}\end{aligned}$$