Voluntary Exercises 4

Gabriel Pons Fontoira

27/09/2024

1 Voluntary exercises from chapter 4

1.1. Let $\{\mathbf{X_1}, ..., \mathbf{X_n}\}$ a s.r.s. of a r.v. $\mathbf{X} \sim U(0, \theta)$. From estimator $\hat{\theta} = \mathbf{X}_{(n)}$, compute a confidence interval for θ with confidence level 0.95.

Disclaimer: I'm not sure if this is the approach you were looking for to solve the problem. I got the idea to use this method because when I was studying for the first midterm, I saw it in my old notes from my bachelor. I hope you like it and that this isn't a problem to obtain the extra point :)

To solve this problem, using Neymann's method to obtain IC's seems a possible outcome. This method is use to obtain IC's when we lack a pivot function.

First of all, the Neymann's method requires the use of the MLE of the targeted parameter and it's density function $f_{\hat{\theta}}(t)$.

$$\hat{\theta}_{MLE} = X_{(n)}$$

Proof:

Given a sample $x_1, x_2, \ldots, x_n$ from a uniform distribution $U(0, \theta)$, the probability density function (PDF) is:

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n},$$

However, for the likelihood to be non-zero, the parameter θ must be greater than or equal to the maximum of the sample values. Thus, we need to impose the condition $\theta \geq \max(x_1, x_2, \ldots, x_n)$.

$$L(\theta|x_1, x_2, \ldots, x_n) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \geq \max(x_1, x_2, \ldots, x_n) \\ 0 & \text{if } \theta < \max(x_1, x_2, \ldots, x_n) \end{cases}$$

The log-likelihood function is:

$$\ell(\theta) = \log L(\theta) = -n \log \theta.$$

Again is valid only if $\theta \geq \max(x_1, x_2, \ldots, x_n)$.

To maximize the log-likelihood, we note that since $\ell(\theta) = \log L(\theta) = -n \log \theta.$ is a decreasing function of θ, the maximum value of θ that satisfies the condition $\theta \geq \max(x_1, x_2, \ldots, x_n)$ occurs when $\theta = \max(x_1, x_2, \ldots, x_n)$.

$$\hat{\theta}_{MLE} = \max(x_1, x_2, \ldots, x_n).$$

Continuing with Neymann's method, lets define two values, α_1 y $\alpha_2 \in [0, \alpha]$ such that $\alpha_1 + \alpha_2 = \alpha$. Existirán $g_1(\theta)$ y $g_2(\theta)$ such that:

$$P(\hat{\theta} \leq g_1(\theta)) = \alpha_1 \quad and \quad P(\hat{\theta} \geq g_2(\theta)) = \alpha_2$$

Thus, $P(g_1(\theta) \leq \hat{\theta} \leq g_2(\theta)) = 1 - \alpha$. Solving the inequations $g_1(\theta) \leq \hat{\theta} \leq g_2(\theta)$, we obtain an interval for the targeted parameter θ. infinite intervals will be obtained for every α_1 and α_2 relation. We will choose the shorter interval because with the same confidence level, the shorter one will be the most precise.

We already have proven that $\hat{\theta}_{MLE} = X_{(n)}$. Now lets compute its PDF.

The Cumulative Distribution Function (CDF) of $\hat{\theta}$, $F_{\hat{\theta}}(t)$, is:

$$F_{\hat{\theta}}(t) = P(\hat{\theta} \leq t) = \left(\frac{t}{\theta}\right)^n, \quad 0 \leq t \leq \theta; \quad \text{For } (t > \theta), (F_{\hat{\theta}}(t) = 1).$$

The Probability Density Function (PDF) of $\hat{\theta}$: it's obtained differentiating the CDF with respect to t.

$$f_{\hat{\theta}}(t) = \frac{d}{dt} F_{\hat{\theta}}(t) = n \frac{t^{n-1}}{\theta^n}, \quad 0 \leq t \leq \theta.$$

Thus, the PDF is:

$$f_{\hat{\theta}}(t) = \begin{cases} n \frac{t^{n-1}}{\theta^n}, & 0 \leq t \leq \theta \\ 0 & t > \theta. \end{cases}$$

Let $g_1(\theta)$ y $g_2(\theta)$ such that $P(\hat{\theta} \le g_1(\theta)) = \alpha_1$ y $P(g_2(\theta) \ge \hat{\theta}) = \alpha_2$.

Then:

$$\alpha_1 = P(\hat{\theta} \le g_1(\theta)) = \int_0^{g_1(\theta)} \frac{nt^{n-1}}{\theta^n} dt = \frac{g_1^n(\theta)}{\theta^n} \implies g_1(\theta)^n = \alpha_1 \theta^n \implies g_1(\theta) = \theta \sqrt[n]{\alpha_1}$$

.

$$\alpha_2 = P(g_2(\theta) \le \hat{\theta}) = \int_{g_2(\theta)}^{\theta} \frac{nt^{n-1}}{\theta^n} dt = \frac{1}{\theta^n}[\theta^n - g_2^n(\theta)] = 1 - \frac{g_2^n(\theta)}{\theta^n} \implies \frac{g_2^n(\theta)}{\theta^n} = 1 - \alpha_2$$

$$g_2^n(\theta) = (1 - \alpha_2)\theta^n \implies g_2(\theta) = \theta \sqrt[n]{1 - \alpha_2}$$

Then:

$$1 - \alpha = P(g_1(\theta) \le \hat{\theta} \le g_2(\theta)) = P(\theta \sqrt[n]{\alpha_1} \le \hat{\theta} \le \sqrt[n]{1 - \alpha_2}).$$

So the Confidence Interval is:

$$IC_{(1-\alpha)}(\theta) = \left[\frac{\hat{\theta}}{\sqrt[n]{1 - \alpha_2}}, \frac{\hat{\theta}}{\sqrt[n]{\alpha_1}} \right]$$

For any selection of α_1 y α_2, we have an interval. We are looking for the shortest one minimizing its length.

The length as a function of α_1 is:

$$L(\alpha_1) = \hat{\theta} \left(\frac{1}{\sqrt[n]{\alpha_1}} - \frac{1}{\sqrt[n]{1 - \alpha_2}} \right) = \hat{\theta} \left(\frac{1}{\sqrt[n]{\alpha_1}} - \frac{1}{\sqrt[n]{1 - (\alpha - \alpha_1)}} \right)$$

The derivative is:

$$L'(\alpha_1) = \hat{\theta} \left(-\frac{1}{n} \alpha_1^{-\frac{1}{n}-1} + \frac{1}{n}(1 - \alpha + \alpha_1)^{-\frac{1}{n}-1} \right) = 0 \implies \frac{1}{n}(1 - \alpha + \alpha_1)^{\frac{-\frac{1}{n}}{-1}} = \frac{1}{n}\alpha_1^{-\frac{1}{n}-1} \implies 1 - \alpha = 0.$$

It's a monotonous function, is increasing or decreasing. If we take $\alpha_1 = \alpha$:

$$L'(\alpha) = \hat{\theta}\left(-\frac{1}{n}\alpha^{-\frac{1}{n}-1} + \frac{1}{n}\left(\frac{1}{\sqrt[n]{\alpha}} + 1\right)\right) < 0.$$

The larger is α_1,the shorter the interval will be. The shorter interval is obtained by the largest possible value for α_1, which is α. If $\alpha_2 = 0$,the shortest interval with confidence level $1 - \alpha$ is:

$$\left[\hat{\theta}, \frac{\hat{\theta}}{\sqrt[n]{\alpha}}\right]$$

In our particular case with $\alpha = 0.05$:

$$IC_{(0.95)}(\theta) = \left[\hat{\theta}, \frac{\hat{\theta}}{\sqrt[n]{0.05}}\right]$$

1.2. A manufacturer has developed a new steel wire that is subjected to a weight of 1000kg to check if it breaks or not. Analyzing 100 threads it has been found that 2 of them have broken. Find a confidence interval with coefficient 0.9 for the proportion of threads that break when they are subjected to a weight of 1000kg.

To solve this problem, we will use the following steps:

1. Identify the problem as one of estimating a population proportion.

2. Estimate the sample proportion of threads that break.

3. Use the normal approximation to the binomial distribution to compute the confidence interval for the population proportion.

Step 1: Identify the Population Proportion

We define p as the proportion of steel threads that break under a weight of 1000 kg. We want to find a confidence interval for p, based on the sample of 100 steel wires, where 2 wires broke.

The sample size is $n = 100$ and the number of threads that broke is $x = 2$. The sample proportion $\hat{p}$, which serves as an estimate of p, is calculated by:

$$\hat{p} = \frac{x}{n} = \frac{2}{100} = 0.02$$

Thus, the sample proportion of threads that broke is 0.02.

Step 2: Confidence Interval

The confidence interval for a population proportion can be computed as:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where:

- $\hat{p}$ is the sample proportion.
- n is the sample size.
- $z_{\alpha/2}$ is the critical value from the standard normal distribution for the desired confidence level.
- $\alpha = 1 -$ Confidence Level.
- $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is the standard error of the sample proportion.

In this case, we are asked to find a 90% confidence interval, so the confidence level is 0.90. Therefore, $\alpha = 1 - 0.90 = 0.10$, and $\alpha/2 = 0.05$. From standard normal distribution tables, the critical value for a 90% confidence interval is:

$$z_{\alpha/2} = 1.645$$

Step 3: Compute the Standard Error

Now, we calculate the standard error of the sample proportion:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.02 \times (1-0.02)}{100}} = \sqrt{\frac{0.02 \times 0.98}{100}} = \sqrt{0.000196} = 0.014$$

Thus, the standard error of the sample proportion is 0.014.

$$z_{\alpha/2} \times SE = 1.645 \times 0.014 = 0.02303$$

Step 4: Compute the Confidence Interval

Finally, we can compute the confidence interval by adding and subtracting the margin of error from the sample proportion $\hat{p}$:

$$\hat{p} \pm \text{Margin of Error} = 0.02 \pm 0.02303$$

This gives the confidence interval:

$$[-0.00303, 0.04303]$$

However, since the proportion cannot be negative, we adjust the lower bound to 0.

$$[0, 0.04303]$$