Statistics, Fall 2025 - Practice Session 4
Interval Estimation

TA: Stylianos Daskalinas

Email: stylianosdaskalinas@gmail.com
Office hours by appointment. Office D3, 3rd floor , Bld. B

16/10/2025

Problem 1

Let $\{X_i\}_{i=1}^n$ be a random sample such that $X_i \sim N(\mu, \sigma^2)$, where σ^2 is an unknown parameter. Calculate the shortest confidence interval for σ given some significance level α.

Solution

If S^2 is the sample variance, then we know that

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$$

For confidence level a, the confidence interval will be given by

$$1-\alpha = \Pr\left(l \leq (n-1)\frac{S^2}{\sigma^2} \leq u\right) = \Pr\left(\sqrt{\frac{(n-1)S^2}{u}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{l}}\right)$$

Now, the length of the interval is given by

$$L(u,l) = \sqrt{\frac{(n-1)S^2}{l}} - \sqrt{\frac{(n-1)S^2}{u}} = \sqrt{(n-1)S^2}\left(\frac{1}{\sqrt{l}} - \frac{1}{\sqrt{u}}\right)$$

Now we can look for the values of u and l that minimize the length of the confidence interval. Notice that we still need the pivotal quantity to lie within the interval with probability $1-\alpha$, therefore the optimization problem is

$$\min_{u,l} L(u,l) \text{ s.t. } \int_l^u f(x)dx = 1-\alpha$$

Here, $f(x)$ stands for the PDF of $\chi^2(n-1)$, which is

$$f(x) = \frac{1}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}$$

Form a Lagrangian

$$\mathcal{L} = \sqrt{(n-1)S^2}\Big(\frac{1}{\sqrt{l}} - \frac{1}{\sqrt{u}}\Big) - \mu\Big[\int_l^u f(x)dx - (1-\alpha)\Big]$$

The first order conditions

$$\frac{\partial}{\partial u}\mathcal{L} = 0 \Rightarrow \frac{1}{2}\sqrt{(n-1)S^2}u^{-\frac{3}{2}} - \mu f(u) = 0 \Rightarrow \frac{1}{2}\sqrt{(n-1)S^2}u^{-\frac{3}{2}} = \mu f(u)$$

$$\frac{\partial}{\partial u}\mathcal{L} = 0 \Rightarrow -\frac{1}{2}\sqrt{(n-1)S^2}l^{-\frac{3}{2}} + \mu f(l) = 0 \Rightarrow \frac{1}{2}\sqrt{(n-1)S^2}l^{-\frac{3}{2}} = \mu f(l)$$

$$\frac{\partial}{\partial \mu}\mathcal{L} = 0 \Rightarrow \int_l^u f(x)dx = 1-\alpha$$

Dividing the first two conditions and using the shape of the χ^2 PDF we have that

$$\left(\frac{u}{l}\right)^{-\frac{3}{2}} = \left(\frac{u}{l}\right)^{\frac{n-1}{2}} e^{-\frac{u-l}{2}} \Rightarrow \left(\frac{u}{l}\right)^{\frac{n}{2}} = e^{\frac{u-l}{2}} \Rightarrow \log\left(\frac{u}{l}\right) = \frac{u-l}{n}$$

Utilizing this result along with the third condition, we reach a system of two nonlinear equations

$$\log\left(\frac{u}{l}\right) = \frac{u-l}{n}$$

$$\int_l^u f(x)dx = 1-\alpha$$

The solution gives the lower and upper bound of the shortest $(1-\alpha)$ confidence interval for σ. Since the system is non-linear, numerical methods such as the Newton-Raphson method must be employed.

Problem 2

Suppose that a random sample $\{X_i\}_{i=1}^n$ follows a density

$$f(x; b) = bx^{b-1}\mathbb{1}_{[0,1]}(x)$$

Find the symmetric $(1-\alpha)$ confidence interval for b.

Solution

First, note that the CDF of X_i is

$$F(x) = \int_0^x bt^{b-1}dt = x^b$$

Naturally, the CDF follows the Uniform distribution in the $[0, 1]$ interval, i.e.

$$\Pr(X^b \leq x) = x$$

therefore, we have that

$$\Pr(-\log(X^b) \leq x) = \Pr(X^b \geq e^{-x}) = 1 - \Pr(X^b \leq e^{-x}) = 1 - e^{-x}$$

Hence, the quantity $-\log(X^b) = -b\log(X)$ should follow the exponential distribution with parameter 1. Then, the quantity

$$-2b\sum_{i=1}^{n}\log(X_i)$$

should follow the χ^2 distribution with $2n$ degrees of freedom. Then, we can use the pivotal quantity $-2b\sum_{i=1}^{n}\log(X_i)$ to get the confidence interval of b. Since we're looking for a symmetric interval, we can choose the lower and upper bounds through the values that equate the area under the χ^2 curve over each of the remaining intervals with half of the confidence level, i.e. $\chi^2_{\frac{\alpha}{2},2n}$ and $\chi^2_{1-\frac{\alpha}{2},2n}$.

$$1-\alpha = \Pr\left(\chi^2_{\frac{\alpha}{2},2n} \leq -2b\sum_{i=1}^{n}\log(X_i) \leq \chi^2_{1-\frac{\alpha}{2},2n}\right) = \Pr\left(-\frac{\chi^2_{1-\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n}\log(X_i)} \leq b \leq -\frac{\chi^2_{\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n}\log(X_i)}\right)$$

which gives the confidence interval for b.

Problem 3

Suppose a random sample $\{X_i\}_{i=1}^n$ follows the Geometric distribution with parameter p. For $n \to \infty$, find the $(1-\alpha)$ confidence interval of p.

Solution

From Likelihood Theory we know that the Maximum Likelihood estimator is asymptotically normal, i.e.

$$\frac{\hat{p}_{\text{MLE}} - p}{\sqrt{\text{Var}(\hat{p}_{\text{MLE}})}} \to N(0,1)$$

Therefore, given that $n \to \infty$, we can use the normal distribution to construct the confidence intervals. Given the symmetry of the normal distribution, we can take the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the normal distribution as boundary values for the pivotal quantity mentioned above, i.e., $z_{\frac{\alpha}{2}}$ and $z_{1-\frac{\alpha}{2}} = -z_{\frac{\alpha}{2}}$.

First, consider the log-Likelihood function of the sample

$$\log\Big(L(p|x_1,\dots,x_n)\Big) = \log\Big((1-p)^n p^{\sum_{i=1}^n x_i}\Big) = \log\Big((1-p)^n p^{n\overline{x}}\Big) = n\log(1-p) + n\overline{x}\log(p)$$

Maximizing with respect to p we get

$$\frac{d}{dp}\log\Big(L(p|x_1,\dots,x_n)\Big) = 0 \Rightarrow -\frac{1}{1-p} + \frac{\overline{x}}{p} = 0 \Rightarrow \hat{p} = \frac{\overline{x}}{1+\overline{x}}$$

Since asymptotic variance is equal to the inverted Fisher information, we can estimate it as follows

$$\begin{aligned} I(p) &= -\mathbb{E}\Big[\frac{d^2}{dp^2}\log\Big(L(p|x_1,\dots,x_n)\Big)\Big] = -\mathbb{E}\Big[-\frac{n}{(1-p)^2} - \frac{n\overline{x}}{p^2}\Big] \\ &= \frac{n}{(1-p)^2} + \frac{n}{p^2}\mathbb{E}[\overline{x}] = \frac{n}{(1-p)^2} + \frac{n}{p^2}\frac{p}{1-p} = \frac{n}{p(1-p)^2} \end{aligned}$$

Then,

$$\text{Var}(\hat{p}) = \frac{p(1-p)^2}{n}$$

Substituting the estimated values in the variance, we get the pivotal quantity

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})^2}{n}}} = \sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}}(1-\hat{p})}$$

Therefore, the interval becomes

$$\left[z_{\frac{\alpha}{2}} \leq \sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}}(1-\hat{p})} \leq z_{1-\frac{\alpha}{2}}\right] = \left[\hat{p} - \frac{\sqrt{\hat{p}}(1-\hat{p})}{\sqrt{n}}z_{\frac{\alpha}{2}} \leq p \leq \hat{p} + \frac{\sqrt{\hat{p}}(1-\hat{p})}{\sqrt{n}}z_{\frac{\alpha}{2}}\right]$$

Remember that this result holds solely for large samples and is not too robust when $\hat{p}$ is close to 0 or 1.

Notes

- The CDF $F(X)$ of any random variable X is distributed as $F(X) \sim U(0,1)$
- If $X_i \sim Exp(c)$, then $\frac{2}{c}\sum_{i=1}^{n} X_i \sim \chi^2(2n)$
- The probability mass function of the Geometric distribution with parameter $1-p$ is

$$(1-p)p^x \mathbb{1}_{\mathbb{N}_0}(x)$$

- If $\{X_i\}_{i=1}^{n}$ and $X_i \sim N(\mu, \sigma^2)$, then

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi(n-1)$$

  where S^2 is the sample variance.