Mathematical Statistics

Tutorial 5

1. Let $X_1, \dots, X_{n_1}$ be an independent random sample from $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \dots, Y_{n_2}$ be a second independent random sample from $\mathcal{N}(\mu_2, \sigma^2)$. The samples are mutually independent and share the same unknown variance σ^2. Find the $(1-\alpha)100\%$ confidence interval for the difference of the means, $\mu_1 - \mu_2$.

2. Let $X_1, \dots, X_n$ be a random sample whose PDF is $f_{X|\theta}(x) = \frac{2x}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x)$. Find the $(1-\alpha)100\%$ confidence interval based on the MLE and MoM.

Problem 1

Let $X_1, \ldots, X_{n_1}$ be an independent random sample from $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \ldots, Y_{n_2}$ be a second independent random sample from $\mathcal{N}(\mu_2, \sigma^2)$. The samples are mutually independent and share the same unknown variance σ^2. Find the $(1-\alpha)100\%$ confidence interval for the difference of the means, $\mu_1 - \mu_2$.

Solution

The difference of the sample means is $\bar{X} - \bar{Y}$. Since X_i and Y_i are from Normal distributions, the difference of their means is also Normal:

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

Standardizing this difference gives a **Standard Normal** variable Z:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}(0, 1)$$

If σ^2 were known, we would use Z. However, since σ^2 is unknown, we must estimate it.

The t-distribution is defined as the ratio of a Standard Normal variable (Z) and the square root of a Chi-Squared variable (W) divided by its degrees of freedom (ν):

$$T = \frac{Z}{\sqrt{W/\nu}}$$

We need an estimator for σ^2 that is independent of $\bar{X} - \bar{Y}$ and follows a χ^2 distribution when properly scaled.

The sum of the scaled sample variances follows a χ^2 distribution:

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2_{n_1 - 1} \quad \text{and} \quad \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_2 - 1}$$

Since the samples are independent, their sum is also χ^2:

$$W = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_1 + n_2 - 2}$$

The degrees of freedom are $\nu = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

The pooled variance S_p^2 is defined such that the numerator W is $\frac{(n_1+n_2-2)S_p^2}{\sigma^2}$:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

We construct the ratio $T = \frac{Z}{\sqrt{W/\nu}}$, substituting S_p^2 for σ^2 in the standard error:

$$T = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}{\sqrt{\frac{\frac{(n_1+n_2-2)S_p^2}{\sigma^2}}{n_1+n_2-2}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Because the numerator is $\mathcal{N}(0, 1)$ and the denominator is the square root of a χ^2 distribution (divided by its degrees of freedom, ν) and is independent of the numerator, the quantity T follows the **Student's t-distribution** with $\mathbf{df = n_1 + n_2 - 2}$.

The $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is found by solving the inequality $P(-t_{1-\alpha/2,df} < T < t_{1-\alpha/2,df}) = 1 - \alpha$:

$$\text{CI}(\mu_1 - \mu_2) = (\bar{X} - \bar{Y}) \pm t_{1-\alpha/2, n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $t_{1-\alpha/2,n_1+n_2-2}$ is the critical value leaving $\alpha/2$ in the upper tail of the t-distribution.

Problem 2

Let $X_1, \ldots, X_n$ be a random sample whose PDF is $f_{X|\theta}(x) = \frac{2x}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x)$. Find the $(1-\alpha)100\%$ confidence interval based on the MLE and MoM.

Solution

The **MLE** is the maximum order statistic:

$$\hat{\theta}_n = X_{(n)}$$

The population mean is $E[X] = \int_0^\theta x \frac{2x}{\theta^2} dx = \frac{2\theta}{3}$. Setting the sample mean $\bar{X}$ equal to the population mean gives the MME $\tilde{\theta}_n$:

$$\bar{X} = \frac{2\tilde{\theta}_n}{3} \implies \tilde{\theta}_n = \frac{3}{2}\bar{X}$$

Exact Confidence Interval

Since the MLE is a function of the sufficient statistic $X_{(n)}$, we use the exact distribution of a pivot quantity to construct the confidence interval (CI). The CDF of the maximum order statistic is $F_{X_{(n)}}(t) = \left(\frac{t^2}{\theta^2}\right)^n = \frac{t^{2n}}{\theta^{2n}}$ for $0 < t < \theta$. The pivot quantity $Y = \frac{X_{(n)}^2}{\theta^2}$ follows the distribution $F_Y(y) = y^n$ for $0 < y < 1$, which is the Beta$(n, 1)$ distribution. For a $(1-\alpha)100\%$ CI, we find the quantiles $y_{\alpha/2}$ and $y_{1-\alpha/2}$ of Y:

$$y_{\alpha/2} = \left(\frac{\alpha}{2}\right)^{1/n} \quad \text{and} \quad y_{1-\alpha/2} = \left(1 - \frac{\alpha}{2}\right)^{1/n}$$

Inverting the statement $P(y_{\alpha/2} < X_{(n)}^2/\theta^2 < y_{1-\alpha/2}) = 1-\alpha$ yields:

$$\text{CI}_{\text{EXACT}}(\theta) = \left[\frac{X_{(n)}}{\sqrt{y_{1-\alpha/2}}}, \frac{X_{(n)}}{\sqrt{y_{\alpha/2}}}\right] = \left[\frac{X_{(n)}}{\left(1-\frac{\alpha}{2}\right)^{1/(2n)}}, \frac{X_{(n)}}{\left(\frac{\alpha}{2}\right)^{1/(2n)}}\right]$$

Asymptotic Confidence Interval

The MME $\tilde{\theta}_n$ is based on the sample mean $\bar{X}$, which satisfies the conditions for the **Central Limit Theorem (CLT)**. The variance of X is $\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{\theta^2}{2} - \left(\frac{2\theta}{3}\right)^2 = \frac{\theta^2}{18}$. Using the **Delta Method** on $\tilde{\theta}_n = g(\bar{X}) = \frac{3}{2}\bar{X}$:

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, [g'(E[X])]^2 \,\text{Var}(X)\right)$$

Since $g'(\cdot) = 3/2$, the asymptotic variance is:

$$V_{\text{ASYM}}(\tilde{\theta}_n) = \left(\frac{3}{2}\right)^2 \frac{\theta^2}{18} = \frac{\theta^2}{8}$$

We use the approximate normality $\tilde{\theta}_n \sim \mathcal{N}(\theta, V_{\text{ASYM}}(\tilde{\theta}_n)/n)$ and replace θ with the estimator $\tilde{\theta}_n$ in the variance:

$$\text{CI}_{\text{MME}}(\theta) \approx \tilde{\theta}_n \pm Z_{1-\alpha/2}\sqrt{\frac{\tilde{\theta}_n^2/8}{n}} = \tilde{\theta}_n \left[1 \pm \frac{Z_{1-\alpha/2}}{\sqrt{8n}}\right]$$

Asymptotic Confidence Interval for MLE

Since the support of the PDF depends on θ, the MLE $\hat{\theta}_n = X_{(n)}$ does not follow the standard asymptotic normality theorem. Instead, the asymptotic distribution is based on the limit of the maximum order statistic. The specialized result for this distribution, where the density is non-zero at θ, states that the quantity $n(\theta - X_{(n)})$ converges to a scaled Exponential distribution:

$$n(\theta - X_{(n)}) \xrightarrow{d} W, \quad \text{where } W \sim \text{Exponential}(2/\theta)$$

The CDF of W is $F_W(w) = 1 - e^{-2w/\theta}$ for $w > 0$. For large n, we use $n(\theta - X_{(n)}) \approx W$. We find the quantiles $w_{\alpha/2}$ and $w_{1-\alpha/2}$ such that $P(W > w_\gamma) = 1 - \gamma$.

$$w_\gamma = -\frac{\theta}{2}\ln(1-\gamma)$$

Substituting this into the confidence interval derivation (and using $\hat{\theta}_n$ for θ in the quantile terms for the final interval):

$$\text{CI}_{\text{MLE-ASYM}}(\theta) \approx \left[\hat{\theta}_n\left(\frac{1}{1+\frac{\ln(1-\alpha/2)}{2n}}\right), \hat{\theta}_n\left(\frac{1}{1+\frac{\ln(\alpha/2)}{2n}}\right)\right]$$

This interval is often simpler to calculate asymptotically than the exact one, but is based on the non-regular limit distribution of the MLE.