# Statistical Inference and Multivariate Analysis (MA324)

## Lecture Slides
## Lecture 24

Interval Estimation

Indian Institute of Technology Guwahati

Jan-May 2023

# Two-sample Problems

Example 1: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu_1, \sigma^2)$ and $Y_1, Y_2, \ldots, Y_m \overset{i.i.d.}{\sim} N(\mu_2, \sigma^2)$. Also, assume that $X_i$'s and $Y_j$'s are independent. Here, $\mu_1$, $\mu_2$, and $\sigma$ are assumed to be unknown and we are interested to construct a $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$. Let us first try to construct a pivot based on minimal sufficient statistic $\left(\overline{X}, \overline{Y}, S^2\right)$, where the pooled sample variance $S^2$ is defined by

$$S^2 = \frac{1}{n+m-2}\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + \sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2\right].$$

Now, notice that

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right) \implies T = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$$

Of course, $T$ is a pivot, but we cannot use it to construct the required confidence interval due to the presence of unknown $\sigma$ in $T$.

**Example 2:** Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1, Y_2, \ldots, Y_m \overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2)$. Also, assume that $X_i$'s and $Y_j$'s are independent. Here, $\mu_1$, $\mu_2$, $\sigma_1$, and $\sigma_2$ are assumed to be unknown and we are interested to construct a $100(1-\alpha)\%$ CI for $\dfrac{\sigma_2^2}{\sigma_1^2}$. In this case minimal sufficient statistic is $\left(\overline{X}, \overline{Y}, S_1^2, S_2^2\right)$, where $S_1^2$ and $S_2^2$ are sample variances based on the samples $X_i$'s and $Y_j$'s, respectively. In this case,

$$T = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} = \frac{\frac{(n-1)S_1^2}{(n-1)\sigma_1^2}}{\frac{(m-1)S_2^2}{(m-1)\sigma_2^2}} \sim F_{n-1,\,m-1}.$$

Thus,

$$P\left(F_{n-1,\,m-1,\,1-\frac{\alpha}{2}} \leq \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \leq F_{n-1,\,m-1,\,\frac{\alpha}{2}}\right) = 1-\alpha$$

so that a $100(1-\alpha)\%$ CI for $\dfrac{\sigma_2^2}{\sigma_1^2}$ is

$$C(\boldsymbol{X}) = \left[\frac{S_2^2}{S_1^2} F_{n-1,\,m-1,\,1-\frac{\alpha}{2}},\; \frac{S_2^2}{S_1^2} F_{n-1,\,m-1,\,\frac{\alpha}{2}}\right].$$

# Asymptotic CI

- In many cases it is very **difficult** to **find pivot** for a **small sample**. For example, it is difficult to find a pivot to construct CI for successes probability of a Bernoulli distribution.

- However, we may able to find CI quite easily if the **sample size** is sufficiently **large**. This CI is called **asymptotic confidence interval.**

- For this purpose, **convergence in distribution** (mainly CLT or large sample distribution of MLE) and **convergence in probability** (consistent estimator) are handy tools.

# Distribution Free Population Mean

Let $X_1,\, X_2,\, \ldots$ be i.i.d. random variables with mean $\mu$ and finite variance $\sigma^2$. Then, using CLT

$$\frac{\sqrt{n}\,\left(\overline{X}_n - \mu\right)}{\sigma} \xrightarrow{\mathcal{L}} Z \sim N(0,\, 1).$$

Thus, if we have a RS with large sample size $n$, we can approximate the distribution of $\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma}$ using a standard normal distribution. Hence,

$$P\left(-z_{\alpha/2} \le \frac{\sqrt{n}\,\left(\bar{X}_n - \mu\right)}{\sigma} \le z_{\alpha/2}\right) \approx 1 - \alpha.$$

If $\sigma$ is **known** and $n$ is sufficiently large, we can use the last statement to find an asymptotic CI for $\mu$ and it is given by

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right].$$

If $\sigma$ is **unknown**, we can proceed as follows. Using WLLN, we have $\frac{S_n}{\sigma} \xrightarrow{P} 1$, and hence,

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} \xrightarrow{\mathcal{L}} Z \sim N(0,\, 1).$$

Hence, $P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} \leq z_{\alpha/2}\right) \approx 1 - \alpha$. An asymptotic CI for $\mu$ is given by

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} z_{\alpha/2},\, \bar{X}_n + \frac{S_n}{\sqrt{n}} z_{\alpha/2}\right].$$

Note that this **method can be used for any distribution** of $X_1,\, X_2,\, \ldots,\, X_n$, as long as the **conditions of CLT hold true**. Therefore, it is called **distribution free.**

# Using MLE

Let $\widehat{\theta}_n$ be a **consistent estimator** of $\theta$ and $\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} N(0,\, b^2(\theta))$, where $b(\theta) > 0$ for all $\theta \in \Theta$. Assume that $b(\cdot)$ is a **continuous function**.
Then, $\frac{b(\widehat{\theta}_n)}{b(\theta)} \xrightarrow{P} 1$, and hence, $\frac{\sqrt{n}\left(\widehat{\theta}_n - \theta\right)}{b(\widehat{\theta}_n)} \xrightarrow{\mathcal{L}} N(0,\, 1)$.
A $100(1-\alpha)\%$ asymptotic CI for $\theta$ is given by

$$\left[\widehat{\theta}_n - \frac{b(\widehat{\theta}_n)}{\sqrt{n}} z_{\alpha/2},\, \widehat{\theta}_n + \frac{b(\widehat{\theta}_n)}{\sqrt{n}} z_{\alpha/2}\right].$$

Under some regularity conditions, we may use MLE of $\theta$ in place of $\widehat{\theta}_n$.

**Example 3:** Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Bernaoulli(p)$, where $p \in (0, 1)$. We are interested to construct asymptotic CI for $p$. We know that $\widehat{p}_n = \overline{X}_n \xrightarrow{P} p$ and $\frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} N(0, 1)$. Here, $b(p) = \sqrt{p(1-p)}$, which is a continuous function in $p \in (0, 1)$. Hence, $\frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow{\mathcal{L}} N(0, 1)$. A $100(1-\alpha)\%$ asymptotic CI for $p$ is

$$\left[\overline{X}_n - \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} z_{\alpha/2}, \, \overline{X}_n + \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} z_{\alpha/2}\right].$$