# MATH 4752/6752 – Mathematical Statistics II

# Interval Estimation

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University

**Interval estimation** is to find (the values of) two statistics to bound the value of the parameter with certain probability.

**Definition.** Let $\alpha \in (0,1)$, and $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ be two statistics such that

$$\mathrm{P}(\widehat{\Theta}_1 < \theta < \widehat{\Theta}_2) = 1 - \alpha.$$

Suppose we obtain the values $\widehat{\Theta}_1 = \widehat{\theta}_1$ and $\widehat{\Theta}_2 = \widehat{\theta}_2$, then we call $(\widehat{\theta}_1, \widehat{\theta}_2)$ a $(1-\alpha)\cdot 100\%$ **confidence interval (CI)** of $\theta$. Here $1-\alpha$ is called the **degree of confidence**, and $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are called the lower and upper **confidence limits**.

## Interval estimation of means

Suppose $X_1, \ldots, X_n$ is a random sample of size $n$ from distribution $N(\mu, \sigma^2)$ with unknown $\mu$ and known $\sigma^2$. Then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. In other words,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Hence

$$\begin{aligned} 1 - \alpha &= \mathsf{P}\Big( - z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \Big) \\ &= \mathsf{P}\Big( \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Big) \end{aligned}$$

So the $(1 - \alpha)\cdot$100% confidence interval of $\mu$ is

$$\Big( \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Big)$$

Another interpretation is that the error of $\bar{x}$ to $\mu$ is bounded by $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ with probability $1 - \alpha$.

**Example.** Consider the normal population $N(\mu, \sigma^2)$ where $\sigma^2 = 225$. We obtain the value of a random sample of size $n = 20$ with sample mean $\bar{x} = 64.5$. Find the 95% confidence interval of $\mu$.

**Solution.** We have $\sigma = \sqrt{225} = 15$, $\alpha = 0.05$ and hence $z_{\alpha/2} = z_{0.025} = 1.96$ from the table of normal distribution. So the 95% confidence interval of $\mu$ is

$$\left(64.5 - 1.96 \cdot \frac{15}{\sqrt{20}}, 64.5 + 1.96 \cdot \frac{15}{\sqrt{20}}\right) = (57.7, 70.9).$$

**Remarks.**

- $(1-\alpha)\cdot 100\%$ confidence interval is not unique. For example,

$$\left(\bar{x} - z_{2\alpha/3}\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/3}\frac{\sigma}{\sqrt{n}}\right)$$

  is also a $(1-\alpha)\cdot 100\%$ confidence interval.

- We can also construct **one-sided** $(1-\alpha)\cdot 100\%$ confidence interval such as

$$\left(-\infty, \quad \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

What if we have **general distribution, unknown variance, large sample size** $n \geq 30$? In this case, we can invoke the central limit theorem to obtain approximate confidence interval.

**Example.** Suppose we obtain the following values of a random sample from a distribution:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 13 | 18 | 19 | 17 | 21 | 29 | 22 | 16 | 28 | 21 | 15 |
| 26 | 23 | 24 | 20 | 8 | 17 | 17 | 21 | 32 | 18 | 25 | 22 |
| 16 | 10 | 20 | 22 | 19 | 14 | 30 | 22 | 12 | 24 | 28 | 11 |

Construct a 95% confidence interval of the mean.

**Solution.** We have $n = 36$, $\bar{x} = 19.92$, $s = 5.73$, and $z_{0.025} = 1.96$. By CLT, we know $\bar{X}$ approximately follow $N(\mu, \frac{\sigma^2}{n})$. We approximate $\sigma$ using $s$, and construct the 95% CI as

$$\Big(19.92 - 1.96 \cdot \frac{5.73}{\sqrt{36}}, 19.92 + 1.96 \cdot \frac{5.73}{\sqrt{36}}\Big) = (18.05, 21.79).$$

What if we have **normal distribution** $N(\mu, \sigma^2)$ **with unknown** $\sigma^2$ **and small sample size**? We can use the $t$-distribution:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Therefore

$$\mathrm{P}\left(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Hence a $(1-\alpha)\cdot$100% confidence interval of $\mu$ is

$$\left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \quad \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right).$$

**Example.** Suppose we have a random sample of size $n = 12$ from a normal population $N(\mu, \sigma^2)$ where $\sigma^2$ is unknown. We obtain the values of the random sample which yields $\bar{x} = 66.3$ and $s^2 = 8.4^2$. Find a 95% confidence interval of $\mu$.

**Solution.** We have $t_{0.025,11} = 2.21$. Hence the 95% confidence interval is

$$\left(66.3 - 2.21 \cdot \frac{8.4}{\sqrt{12}}, \quad 66.3 + 2.21 \cdot \frac{8.4}{\sqrt{12}}\right) = (61.0, 71.6).$$

**Interval estimation of the difference between two means from two normal populations with known variances.**

Suppose $\bar{X}_i$ is the sample mean of a random sample of size $n_i$ from the normal population $N(\mu_i, \sigma_i^2)$ where $\sigma_i^2$ is known for $i = 1, 2$. Then

$$\bar{X}_1 - \bar{X}_2 \sim N\Big(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\Big).$$

Then a $(1-\alpha)\cdot$100% confidence interval of $\mu_1 - \mu_2$ is

$$\left( (\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \quad (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right).$$

**Example.** Suppose we have random samples from two normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with

$$\bar{x}_1 = 418, \quad \sigma_1^2 = 26^2, \quad n_1 = 40,$$
$$\bar{x}_2 = 406, \quad \sigma_2^2 = 22^2, \quad n_2 = 50.$$

Find a 94% confidence interval of $\mu_1 - \mu_2$.

**Solution.** We have $\bar{x}_1 - \bar{x}_2 = 12$ and $z_{0.03} = 1.88$. Hence the 94% confidence interval of $\mu_1 - \mu_2$ is

$$\left(12 - 1.88 \cdot \sqrt{\frac{26^2}{40} + \frac{22^2}{50}},\ 12 + 1.88 \cdot \sqrt{\frac{26^2}{40} + \frac{22^2}{50}}\right) = (6.3, 25.7).$$

**Remark.** If we have two general distributions with unknown variances but large sample sizes ($n_1, n_2 \geq 30$), then we can apply central limit theorem and approximate $\sigma_i$ with $s_i$ to obtain approximate $(1-\alpha)\cdot 100\%$ confidence interval of $\mu_1 - \mu_2$ is

$$\left( (\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right).$$

If we have two independent **normal populations** with **unknown variances** and **small sample sizes** ($n_1, n_2 < 30$), then we know

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1),$$

$$Y = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} + \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_1+n_2-2}.$$

are independent, and thus

$$T = \frac{Z}{\sqrt{Y/(n_1 + n_2 - 2)}} \sim t_{n_1+n_2-2}.$$

However, the unknown $\sigma_1^2$ and $\sigma_2^2$ cannot be canceled in this ratio, and therefore we cannot construct confidence intervals based on $t$-distribution.

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (assuming the two normal populations have the same variance), then they can be canceled!

To see this, we notice that

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1),$$

$$Y = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} + \frac{(n_2 - 1)S_2^2}{\sigma_2^2} = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2},$$

where $S_p^2$ is called the **pooled sample variance** defined by

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

In this case, we have

$$T = \frac{Z}{\sqrt{Y/(n_1+n_2-2)}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

Therefore a $(1-\alpha)$·100% confidence interval of $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, n_1+n_2-2} \cdot s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

**Example.** Suppose we have random samples from two normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unknown but equal variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and

$$\bar{x}_1 = 3.1, \quad s_1 = 0.5, \quad n_1 = 10,$$
$$\bar{x}_2 = 2,7, \quad s_2 = 0.7, \quad n_2 = 8.$$

Find a 95% confidence interval of $\mu_1 - \mu_2$.

**Solution.** We have $\bar{x}_1 - \bar{x}_2 = 0.4$, $n_1 + n_2 - 2 = 16$, $t_{0.025,16} = 2.212$ and

$$s_p^2 = \frac{9 \cdot 0.5^2 + 7 \cdot 0.7^2}{16} = 0.596^2, \qquad \sqrt{\frac{1}{10} + \frac{1}{8}} = 0.474.$$

Hence the 95% confidence interval of $\mu_1 - \mu_2$ is

$$\Big(0.4 - 2.212 \cdot 0.596 \cdot 0.474,\ 0.4 + 2.212 \cdot 0.596 \cdot 0.474\Big) = (-0.22, 1.02).$$

## Interval estimation of proportions

Suppose $X$ follows Binomial$(n, \theta)$ with known large $n$, then by central limit theorem we know

$$\frac{\frac{X}{n} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \sim N(0, 1).$$

Denoting $\widehat{\theta} = \frac{x}{n}$ and approximate the variance $\theta(1-\theta)$ with $\widehat{\theta}(1-\widehat{\theta})$, we obtain an approximate $(1-\alpha)$·100% confidence interval of $\theta$:

$$\Big(\widehat{\theta} - z_{\alpha/2} \cdot \sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}},\ \widehat{\theta} + z_{\alpha/2} \cdot \sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}}\ \Big).$$

**Remark.** Let $Y_1, \dots, Y_n$ be a random sample of Bernoulli($\theta$), then $X = Y_1 + \cdots + Y_n$ follows Binomial$(n, \theta)$. Then $\hat{\Theta} := \frac{X}{n}$ is the sample mean. Let $S^2$ denote the sample variance, then

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{\Theta})^2 = \frac{1}{n-1} \Big( \sum_{i=1}^{n} Y_i^2 - n\hat{\Theta}^2 \Big) = \frac{n}{n-1} \hat{\Theta}(1-\hat{\Theta})$$

**Example.** Suppose 136 of 400 persons received flu shot experienced discomfort. Find a 95% confidence interval of the proportion of persons would experience discomfort after the flu shot.

**Solution.** We have $x = 136$, $n = 400$, and hence $\widehat{\theta} = \frac{136}{400} = 0.34$. We find $z_{0.025} = 1.96$. Hence the 95% confidence interval of $\theta$ is

$$\Big(0.34 - 1.96 \cdot \sqrt{\frac{0.34 \cdot 0.66}{400}},\ 0.34 + 1.96 \cdot \sqrt{\frac{0.34 \cdot 0.66}{400}}\Big) = (0.294, 0.386).$$

**Interval estimation of the difference between two proportions**

Suppose $X_i$ follows Binomial($n_i, \theta_i$) with known large $n_i$ for $i = 1, 2$. Then we have approximately

$$\frac{\frac{X_i}{n_i} - \theta_i}{\sqrt{\frac{\theta_i(1-\theta_i)}{n_i}}} \sim N(0, 1), \quad \text{for } i = 1, 2,$$

which are independent. Denote $\widehat{\Theta}_i = \frac{X_i}{n_i}$. Then there is approximately

$$\frac{(\widehat{\Theta}_1 - \widehat{\Theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \sim N(0, 1).$$

Hence we obtain an approximate $(1 - \alpha) \cdot 100\%$ confidence interval of $\theta_1 - \theta_2$:

$$(\widehat{\theta}_1 - \widehat{\theta}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\widehat{\theta}_1(1 - \widehat{\theta}_1)}{n_1} + \frac{\widehat{\theta}_2(1 - \widehat{\theta}_2)}{n_2}}$$

**Example.** Suppose 132 of 200 male voters and 90 of 150 female voters favor a candidate running for governor. Find a 99% confidence interval of the difference between the proportions of male and female voters favor this candidate.

**Solution.** We have $x_1 = 132$, $n_1 = 200$, and hence $\widehat{\theta}_1 = \frac{132}{200} = 0.66$. Similarly $x_2 = 90$, $n_2 = 150$, and hence $\widehat{\theta}_2 = \frac{90}{150} = 0.60$. We find $z_{0.005} = 2.575$. Hence the 99% confidence interval of $\theta$ is

$$(0.66 - 0.60) \pm 2.575 \cdot \sqrt{\frac{0.66 \cdot 0.34}{200} + \frac{0.60 \cdot 0.40}{150}},$$

which is $(-0.074, 0.194)$.

**Interval estimation of variances**

Suppose $S^2$ is the sample variance of a random sample of size $n$ from a normal population $N(\mu, \sigma^2)$. Then we know

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

Therefore,

$$\mathsf{P}\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) = 1-\alpha$$

or equivalently

$$\mathsf{P}\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

Hence, given the value $s^2$ of $S^2$, we can obtain a $(1-\alpha)\cdot$100% CI of $\sigma^2$ as

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}},\ \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

**Example.** Suppose we obtain sample variance $s^2 = 2.2^2$ for a random sample of size $n = 16$ from a normal population $N(\mu, \sigma^2)$. Find a 99% confidence interval of $\sigma^2$.

**Solution.** We have $n = 16$, $s = 2.2$, $\chi^2_{0.005,15} = 32.801$ and $\chi^2_{0.995,15} = 4.601$. Hence a 99% confidence interval of $\sigma^2$ is

$$\frac{15(2.2)^2}{32.801} < \sigma^2 < \frac{15(2.2)^2}{4.601}$$

which is $(2.21, 15.78)$.

**Interval estimation of the ratio of two variances**

Suppose $S_i^2$ is the sample variance of a random sample of size $n_i$ from the normal population $N(\mu_i, \sigma_i^2)$ for $i = 1, 2$ and the two populations are independent. Then we know

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}.$$

Therefore,

$$\mathrm{P}\left(f_{1-\alpha/2,n_1-1,n_2-1} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2,n_1-1,n_2-1}\right) = 1-\alpha.$$

Furthermore, we know that $f_{1-\alpha/2,n_1-1,n_2-1} = \frac{1}{f_{\alpha/2,n_2-1,n_1-1}}$. Hence, given the value $s^2$ of $S^2$, we can obtain a $(1-\alpha)$·100% CI of $\sigma^2$ as

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2,n_2-1,n_1-1}$$

**Example.** Suppose we obtain sample variances $s_1^2 = 0.5^2$ and $s_2^2 = 0.7^2$ from two normal random samples where $n_1 = 10$ and $n_2 = 8$. Find a 98% confidence interval of $\frac{\sigma_1^2}{\sigma_2^2}$.

**Solution.** We find that $f_{0.01,9,7} = 6.72$ and $f_{0.01,7,9} = 5.61$. Hence a 99% confidence interval of $\frac{\sigma_1^2}{\sigma_2^2}$ is

$$\frac{0.25}{0.49} \cdot \frac{1}{6.72} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0.25}{0.49} \cdot 5.61$$

which is $(0.076, 2.862)$.