Probability and Mathematical Statistics

Miskolc, 2025.

Dr. Tamás Glavosits

Seminar 11

Statistics II. Interval Estimates

1. $1 - \alpha$ Confidence Interval

$1-\alpha$ Confidence Interval

Definition

Let $\xi_1, \xi_2, \ldots, \xi_n$ be an independent sample with unknown parameter ϑ. We look for statistics $\underline{g}$ and $\overline{g}$ such that

$$\mathbb{P}\left(\underline{g}(\xi_1, \xi_2, \ldots, \xi_n) \leqslant \vartheta \leqslant \overline{g}(\xi_1, \xi_2, \ldots, \xi_n)\right) \geqslant 1-\alpha.$$

Then the interval $(\underline{g}(\xi_1, \xi_2, \ldots, \xi_n), \overline{g}(\xi_1, \xi_2, \ldots, \xi_n))$ is called the $1-\alpha$ **confidence interval with a given confidence level** for the unknown parameter ϑ.

Important Confidence Intervals

In the sequence the sample $\xi_1, \xi_2, \ldots, \xi_n$ will always be an independent sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. We will deal with the following cases:

- Estimating μ when
  1. σ^2 is known: the applied statistic is: $\frac{\bar{\xi}-\mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)$;
  2. σ^2 is unknown: the applied statistic is: $\frac{\bar{\xi}-\mu}{s_n^*}\sqrt{n} \sim t_{n-1}$.
- Estimating σ^2
  3. the applied statistic is: $\frac{ns_n^2}{\sigma^2} \sim \chi^2_{n-1}$

The procedure for constructing the confidence interval

1. Compute the value of $\overline{\xi}$ from the sample, with n and σ^2 are given.
2. Compute the value of $u_{\frac{\alpha}{2}}$. Since $\frac{\overline{\xi}-\mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)$ we have

$$1-\alpha = \mathbb{P}\left(-u_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi}-\mu}{\sigma}\sqrt{n} \leqslant u_{\frac{\alpha}{2}}\right) = \\ = \Phi\left(u_{\frac{\alpha}{2}}\right) - \left(1-\Phi\left(u_{\frac{\alpha}{2}}\right)\right) = 2\Phi\left(u_{\frac{\alpha}{2}}\right) - 1.$$

3. Since the value of α is given, the corresponding $u_{\frac{\alpha}{2}}$ can be easily found from the table.
4. Based on point 2, the confidence interval is

$$\underline{g}(\xi_1, \xi_2, \ldots, \xi_n) = \overline{\xi} - \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}, \qquad \overline{g}(\xi_1, \xi_2, \ldots, \xi_n) = \overline{\xi} + \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}.$$

Problem 1

Let $\xi_1, \xi_2, \ldots, \xi_n \sim \mathcal{N}(\mu, \sigma^2)$ be an independent sample, for which:
$n = 10$;
$\overline{\xi} = 1.12$;
$\sigma_0^2 = 0.3162$.
Determine the 95% confidence interval for μ.

Solution of Problem 1

$1-\alpha=0.95$, so $\alpha=0.05$.
We know that $\frac{\overline{\xi}-\mu}{\sigma_0}\sqrt{n}\sim\mathcal{N}(0,1)$. Then

$$\mathbb{P}\left(-u_{\frac{\alpha}{2}}\leqslant\frac{\overline{\xi}-\mu}{\sigma}\sqrt{n}\leqslant u_{\frac{\alpha}{2}}\right)=$$
$$=\Phi\left(u_{\frac{\alpha}{2}}\right)-\left(1-\Phi\left(u_{\frac{\alpha}{2}}\right)\right)=2\Phi\left(u_{\frac{\alpha}{2}}\right)-1=1-\alpha$$

from which we obtain

$$\Phi\left(u_{\frac{\alpha}{2}}\right)=1-\frac{\alpha}{2}=0.975.$$

Continuation of the Solution to Problem 1

$u_{\frac{\alpha}{2}}$ can be looked up in the table: $u_{\frac{\alpha}{2}} = 1.960$.

$$-u_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi} - \mu}{\sigma}\sqrt{n} \leqslant u_{\frac{\alpha}{2}}$$

then

$$|\overline{\xi} - \mu| < \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}$$

thus

$$\overline{\xi} - \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}} < \mu < \overline{\xi} + \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}$$

that is

$$1.12 - \frac{0.3162 \cdot 1.960}{\sqrt{10}} < \mu < 1.12 + \frac{0.3162 \cdot 1.960}{\sqrt{10}}$$

hence

$$0.92 < \mu < 1.32.$$

Problem 2

Let $\xi_1, \xi_2, \ldots, \xi_n \sim \mathcal{N}(\mu, \sigma^2)$ be an independent sample, for which:
$n = 10$;
$\overline{\xi} = 1.12$;
$\sigma_0 = 0.3162$.
Determine the 99% confidence interval for μ.

Solution of Problem 2

$1 - \alpha = 0.99$, so $\alpha = 0.01$.
We know that $\frac{\overline{\xi}-\mu}{\sigma_0}\sqrt{n} \sim \mathcal{N}(0,1)$. Then

$$\mathbb{P}\left(-u_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi}-\mu}{\sigma}\sqrt{n} \leqslant u_{\frac{\alpha}{2}}\right) =$$
$$= \Phi\left(u_{\frac{\alpha}{2}}\right) - \left(1 - \Phi\left(u_{\frac{\alpha}{2}}\right)\right) = 2\Phi\left(u_{\frac{\alpha}{2}}\right) - 1 = 1 - \alpha$$

from which we obtain

$$\Phi\left(u_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} = 0.995.$$

Continuation of the Solution to Problem 2

$u_{\frac{\alpha}{2}}$ can be looked up in the table: $u_{\frac{\alpha}{2}} = 2.576$.

$$-u_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi} - \mu}{\sigma}\sqrt{n} \leqslant u_{\frac{\alpha}{2}}$$

then

$$|\overline{\xi} - \mu| < \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}$$

thus

$$\overline{\xi} - \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}} < \mu < \overline{\xi} + \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}$$

that is

$$1.12 - \frac{0.3162 \cdot 2.576}{\sqrt{10}} < \mu < 1.12 + \frac{0.3162 \cdot 2.576}{\sqrt{10}}$$

hence

$$0,8624 < \mu < 1.3776.$$

Steps for Interval estimation for unknown μ, also σ^2 unknown

1. Determining $t_{\frac{\alpha}{2}}$

$$\mathbb{P}\left(-t_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi}-\mu}{s_n^*}\sqrt{n} < t_{\frac{\alpha}{2}}\right) = 2\mathbb{F}\left(t_{\frac{\alpha}{2}}\right) - 1 = 1-\alpha.$$

   From the appropriate table, $t_{\frac{\alpha}{2}}$ can be find out using the relevant table by

$$\mathbb{F}\left(t_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

2. Thus we obtain that

$$\overline{\xi} - \frac{t_{\frac{\alpha}{2}} s_n^*}{\sqrt{n}} < \mu < \overline{\xi} - \frac{t_{\frac{\alpha}{2}} s_n^*}{\sqrt{n}}.$$

Problem 3

Let $\xi_1, \xi_2, \ldots, \xi_n \sim \mathcal{N}(\mu, \sigma_0^2$ be an independent sample.
Now neither μ nor σ_0 are known.
Let
$n = 10$;
$\overline{\xi} = 1.12$;
$s_n^* = 0.3162$.
Provide a confidence interval for μ with 95% confidence level.

Solution of Problem 3

$$\frac{\overline{\xi} - \mu}{s_n^*}\sqrt{n} \sim t_{n-1}$$

$$\mathbb{P}\left(-x \leqslant \frac{\overline{\xi} - \mu}{s_n^*}\sqrt{n} \leqslant x\right) = \mathbb{F}(x) - \mathbb{F}(-x) = 1 - \alpha$$

thus

$$\mathbb{F}(x) = 1 - \frac{\alpha}{2}.$$

Since $\mathbb{F}(x) + G(x) = 1$, we have
$G(x) = 1 - \mathbb{F}(x) = 1 - \left(1 - \frac{\alpha}{2}\right) = \frac{\alpha}{2}$, so we must find the value x for which $G(x) = 0.25$.
For $n = 10$ we have $n - 1 = 9$, so look at row 9 and column 0.25.
$x = 2.262$, hence

$$-2.262 \leqslant \frac{\overline{\xi} - \mu}{s_n^*}\sqrt{n} \leqslant 2.262$$

Solution of Problem 3

Thus

$$|\bar{\xi} - \mu| \leqslant \frac{2.262 \cdot s_n^*}{\sqrt{n}}$$

whence we have that

$$1.12 - \frac{2.262 \cdot 0.3162}{\sqrt{10}} \leqslant \mu \leqslant 1.12 + \frac{2.262 \cdot 0.3162}{\sqrt{10}}$$

that is

$$0.8938 \leqslant \mu \leqslant 1.3462$$

Steps for Determining the Confidence Interval

1. Since the density function of the distribution χ^2_{n-1} is not symmetric with respect to the y-axis, we can not construct a confidence interval that is symmetric about the origin, and consequently we cannot obtain the shortest confidence interval either.
2. We know that $\frac{ns_n^2}{\sigma^2} \sim \chi^2_{n-1}$, thus from

$$\mathbb{P}\left(x_l \leqslant \frac{ns_n^2}{\sigma^2} \leqslant x_u\right) = 1 - \alpha$$

   we must find an appropriate pair of values x_l and x_u in the corresponding table.

Steps for Determining the Confidence Interval, continuation

3. Since

$$\mathbb{P}\left(x_l \leqslant \frac{ns_n^2}{\sigma^2} \leqslant x_u\right) = \mathbb{F}(x_u) - \mathbb{F}(x_l) = 1 - \alpha,$$

it is therefore sufficient to find the pair x_a, x_f such that

$$\mathbb{F}(x_u) = 1 - \frac{\alpha}{2}, \qquad \mathbb{F}(x_l) = \frac{\alpha}{2}.$$

With these numbers x_l and x_u, the confidence interval is:

$$\frac{ns_n^2}{x_u} \leqslant \sigma^2 \leqslant \frac{ns_n^2}{x_l}.$$

Problem 4.

Let $\xi_1, \xi_2, \ldots, \xi_n \sim \mathcal{N}(\mu, \sigma_0^2)$ be an independent sample.
$n = 10$;
$s_n^2 = 0.099$;
Construct a confidence interval for σ_0 with confidence level 0.95.

Solution of Problem 4.

We know that

$$\frac{ns_n^2}{\sigma^2} \sim \chi_{n-1}.$$

We are looking for constants x_l and x_u such that

$$\mathbb{P}\left(x_l \leqslant \frac{ns_n^2}{\sigma^2} \leqslant x_u\right) = 1 - \alpha,$$

however, the density function of the χ_{n-1} distribution (denote it by f) is defined on $\mathbb{R}_+$ and thus is not symmetric with respect to the y-axis, so we cannot use the relation $f(-x) = f(x)$.
Let the cumulative distribution function be $\mathbb{F}$, then it must satisfy

$$\mathbb{F}(x_u) - \mathbb{F}(x_l) = 1 - \alpha,$$

but it is enough if

$$\mathbb{F}(x_u) = 1 - \frac{\alpha}{2}, \qquad \mathbb{F}(x_l) = \frac{\alpha}{2}.$$

Solution of Exercise 4, continuation

However, the cumulative distribution function is not in the table, but

$$G(x) + \mathbb{F}(x) = 1, \qquad \text{so} \qquad G(x) = 1 - \mathbb{F}(x)$$

that is,

$$G(x_u) = \frac{\alpha}{2}, \qquad G(x_l) = 1 - \frac{\alpha}{2}.$$

$\frac{\alpha}{2} = 0.025.$
Since $n = 10$, we need to look at the 9th row and the 0.25 and 0.75 columns of the table.

$$G(x_u) = 0.025, \qquad G(x_l) = 0.975,$$

that is,

$$x_u = 19.023, \qquad x_l = 2.7004.$$

Solution of Exercise 4, continuation

Thus we obtain that

$$2.7004 \leqslant \frac{ns_n^2}{\sigma^2} \leqslant 19.023,$$

$$\frac{ns_n^2}{19.023} \leqslant \sigma^2 \leqslant \frac{ns_n^2}{2.7004}.$$

that is

$$0.228 \leqslant \sigma \leqslant 0.7023.$$

End of Seminar 11