Math Revision Session

Statistics (7): Estimation

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

April 10, 2026

Point Estimation

In statistics, **point estimation** means estimating an unknown parameter by a single number.

Let θ be an unknown parameter.

- An **estimator** $\hat{\theta}$ is a function of the sample data used to estimate θ.
- The realised value of $\hat{\theta}$ for a given sample is called an **estimate**.

Examples of Estimators

Typical examples are:

- the sample mean
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
as an estimator of the population mean μ,
- the sample variance
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
as an estimator of the population variance σ^2.

So an estimator is a random variable, because it depends on the sample.

Unbiasedness

An estimator $\hat{\theta}$ is **unbiased** for θ if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

This means that over repeated sampling, the estimator is correct on average.

Bias

The **bias** of an estimator $\hat{\theta}$ is

$$\mathrm{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta.$$

- If $\mathrm{Bias}(\hat{\theta}) = 0$, the estimator is unbiased.
- If $\mathrm{Bias}(\hat{\theta}) \neq 0$, the estimator is biased.

A small bias may sometimes be acceptable if it greatly reduces variance.

Consistency

An estimator $\hat{\theta}_n$ is **consistent** for θ if

$$\hat{\theta}_n \xrightarrow{p} \theta \qquad \text{as } n \to \infty.$$

Equivalently, for every $c > 0$,

$$\Pr(|\hat{\theta}_n - \theta| > c) \to 0 \qquad \text{as } n \to \infty.$$

So consistency means that the estimator gets close to the true parameter as the sample size becomes large.

Efficiency

Among unbiased estimators, an estimator is more **efficient** if it has smaller variance.
If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ, then $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2).$$

So efficiency means precision among unbiased estimators.

Cramér–Rao Lower Bound

The **Cramér–Rao lower bound** gives a theoretical lower bound for the variance of an unbiased estimator:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)},$$

where $I(\theta)$ is the Fisher information:

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta; X)\right].$$

This bound describes the best precision an unbiased estimator can achieve.

Mean Squared Error (MSE)

The **mean squared error** of an estimator $\hat{\theta}$ is

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\big[(\hat{\theta} - \theta)^2\big].$$

It can be decomposed as

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \big(\mathbb{E}[\hat{\theta}] - \theta\big)^2.$$

So MSE reflects both variance and bias.

Why Does MSE Matter?

MSE is often useful because it summarises two important aspects of an estimator:

- **variance:** how much the estimator fluctuates across samples,
- **bias:** how far its average is from the true parameter.

So even a biased estimator may be preferred if it has much smaller variance.

Confidence Interval

A **confidence interval** is a range of values, computed from a sample, that is used to estimate an unknown population parameter.
It reflects sampling uncertainty.

Confidence Coefficient

The **confidence coefficient** is the long-run probability that intervals constructed in the same way contain the true parameter.
For example, a 95% confidence interval means that, in repeated sampling, about 95% of such intervals would contain the true parameter.

How Should We Interpret a Confidence Interval?

A confidence interval does **not** mean:

"there is a 95% probability that this fixed interval contains μ."

Instead:

- the parameter μ is fixed,
- the interval is random before we see the data,
- and 95% refers to the long-run success rate of the procedure.

General Form of a Confidence Interval

A confidence interval often has the form

$$\text{estimator} \pm \text{critical value} \times \text{standard error}.$$

For the population mean, a typical form is

$$\hat{\mu} \pm \text{critical value} \times \frac{\text{standard deviation}}{\sqrt{n}}.$$

Confidence Interval for the Mean (Known Variance)

Suppose the population is normal and the population variance σ^2 is known. Then the $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

where

- $\bar{X}$ is the sample mean,
- $z_{\alpha/2}$ is the standard normal critical value,
- σ is the known population standard deviation.

Example Setup: Known Variance

Suppose we have $n = 10$ observations and know the population variance. Assume that

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} = 2.5.$$

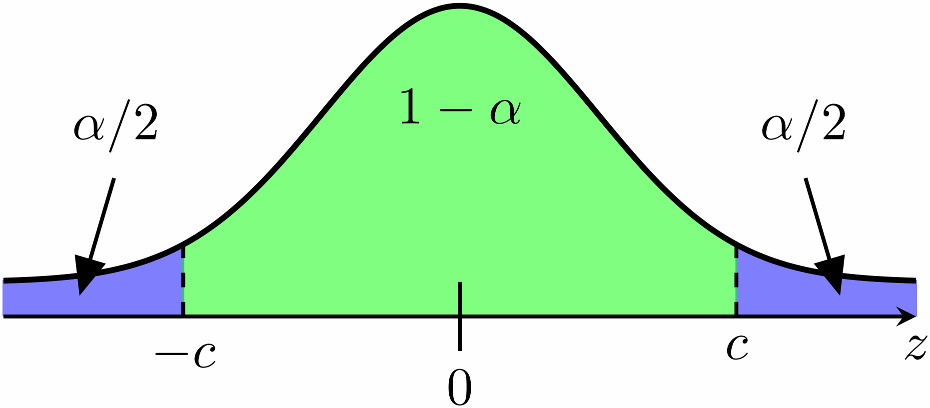
Then

$$\bar{X} \sim N(\mu, 2.5).$$

Standardising,

$$Z = \frac{\bar{X} - \mu}{\sqrt{2.5}} \sim N(0, 1).$$

Central Region of the Standard Normal Distribution



$$\Pr(-c \le Z \le c) = 1 - \alpha.$$

Finding the Critical Value

We choose c so that

$$\Pr(Z \le c) = \Phi(c) = 1 - \frac{\alpha}{2}.$$

If $\alpha = 0.05$, then

$$\Phi(c) = 0.975.$$

From the standard normal table,

$$c \approx 1.96.$$

Confidence Interval Calculation

$$
\begin{aligned}
0.95 &= \Pr(-1.96 \leq Z \leq 1.96) \\
&= \Pr\left(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{2.5}} \leq 1.96\right) \\
&= \Pr\left(\bar{X} - 1.96\sqrt{2.5} \leq \mu \leq \bar{X} + 1.96\sqrt{2.5}\right).
\end{aligned}
$$

If the realised sample mean is $\bar{x}$, the confidence interval is

$$\bar{x} - 1.96\sqrt{2.5} \leq \mu \leq \bar{x} + 1.96\sqrt{2.5}.$$

If $\bar{x} = 100$, then

$$96.9 \leq \mu \leq 103.1.$$

Confidence Interval for the Mean (Unknown Variance)

Now suppose the population is normal, but σ^2 is unknown.
Then we use the sample variance S^2, where

$$S = \sqrt{S^2}.$$

The statistic

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows a t-distribution with $n-1$ degrees of freedom:

$$T \sim t_{n-1}.$$

So the $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}.$$

Relationship Between Sample Mean and Sample Variance

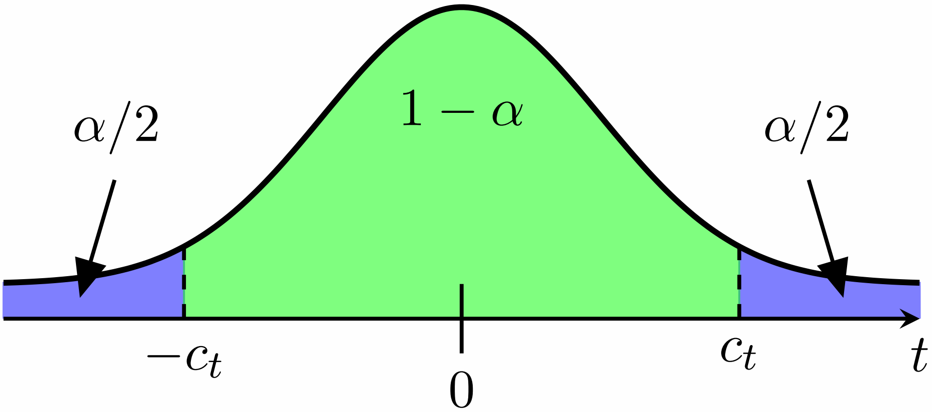
For a normal population:

- $\bar{X}$ and S^2 are both computed from the same sample,
- but in the normal case they are in fact **independent**,
- and this leads to the t-statistic

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

This is one of the key reasons the t-distribution appears in classical inference.

Central Region of the t-Distribution



$$\Pr(-c_t \leq T \leq c_t) = 1 - \alpha.$$

Finding the t Critical Value

We choose c_t so that

$$\Pr(T \leq c_t) = F_{t_{n-1}}(c_t) = 1 - \frac{\alpha}{2}.$$

If $\alpha = 0.05$ and $n = 10$, then the degrees of freedom are 9, and from the t-table we obtain

$$c_t \approx 2.262.$$

Confidence Interval with Unknown Variance

$$
\begin{aligned}
0.95 &= \Pr(-2.262 \leq T \leq 2.262) \\
&= \Pr\left(-2.262 \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq 2.262\right) \\
&= \Pr\left(\bar{X} - 2.262\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + 2.262\frac{S}{\sqrt{n}}\right).
\end{aligned}
$$

If the realised values are $\bar{x} = 100$, $S = 5$, and $n = 10$, then

$$100 \pm 2.262\frac{5}{\sqrt{10}}$$

gives

$$96.42 \leq \mu \leq 103.58.$$

When the Population Distribution is Unknown

If the population distribution is unknown, the sample mean can still be approximately normal when n is large, by the Central Limit Theorem. In that case,

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow{d} N(0,1).$$

So for large n, an approximate confidence interval for μ is

$$\bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}.$$

How to Construct a Confidence Interval (1)

1. **Find the variance of the estimator.** Compute or estimate the variance of $\hat{\theta}$.
2. **Standardise the estimator.** Form a standardised quantity such as

$$Z = \frac{\hat{\theta} - \theta}{\text{standard error}}.$$

How to Construct a Confidence Interval (2)

3. **Substitute unknown quantities if necessary.** Replace unknown parameters by estimators when needed.
4. **Find the critical value.** Use the relevant distribution to obtain the quantile corresponding to $1 - \alpha$.
5. **Construct the interval.** Write

$$\hat{\theta} \pm \text{critical value} \times \text{standard error}.$$

Summary

In this lecture, we studied:

- point estimation,
- examples of estimators,
- unbiasedness, bias, consistency, efficiency, and MSE,
- the Cramér–Rao lower bound,
- confidence intervals,
- confidence intervals for the mean with known and unknown variance,
- large-sample confidence intervals based on the Central Limit Theorem.