# Math Revision Session

## Statistics (5): Population and Sample

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

April 9, 2026

# Relationship Between Population and Sample

- **Population:** the entire set of individuals or observations under study.
  - Often described by parameters such as $\mu$ (mean) and $\sigma^2$ (variance).
- **Sample:** a subset of the population used for analysis.
  - A sample may be written as $S = \{X_1, X_2, \ldots, X_n\}$.
  - Sample statistics such as $\bar{X}$ and $S^2$ are used to estimate population parameters.
- **Key relationships:**
  - $\mathbb{E}[\bar{X}] = \mu$,
  - $\mathbb{E}[S^2] = \sigma^2$,
  - larger samples usually provide more accurate approximations to population characteristics.

# Sample Mean and Estimation

- Given a sample $S = \{X_1, X_2, \ldots, X_n\}$, the **sample mean** is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- **Estimator vs. estimate:**
    - An **estimator** is a function of the sample used to infer a population parameter.
    - An **estimate** is the numerical value obtained from observed data.
- For example:
    - $\bar{X}$ is an estimator of $\mu$,
    - if $\bar{X} = 5.2$, then 5.2 is an estimate of $\mu$.

# Properties of the Sample Mean

- The sample mean is unbiased:

$$\mathbb{E}[\bar{X}] = \mu.$$

- Its variance is

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

- Hence, the variability of $\bar{X}$ becomes smaller as $n$ increases.
- If $X_i \sim N(\mu, \sigma^2)$, then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

# Example: Coin Toss

- Let

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th toss is heads,} \\ 0, & \text{if the } i\text{-th toss is tails.} \end{cases}$$

- Then $p = \Pr(X_i = 1)$ is the probability of heads.
- A natural estimator of $p$ is

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Since $\hat{p} = \bar{X}$, it is a random variable that depends on the sample.

# Expectation and Variance of $\hat{p}$

$$\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n} p = p.$$

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{p(1-p)}{n}.$$

Here,

$$\mathbb{E}[X_i] = p, \qquad \mathrm{Var}(X_i) = p(1-p).$$

# Law of Large Numbers and $\hat{p}$

- The estimator $\hat{p}$ is unbiased:

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

- Its variance decreases with $n$:

$$\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}.$$

- By the law of large numbers,

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

- So larger samples give more stable estimates of the true probability.

Math Revision Session

Jukina HATAKEYAMA

Relationship Between Population and Sample

Population Distribution and Sampling Distribution
Population Distribution
Sampling Distribution

Law of Large Numbers
Statement
Markov's and Chebyshev's Inequalities
Proof of the WLLN

Central Limit Theorem
Big-O and Little-o
Moment Generating Function
An Intuitive MGF-Based Proof
Characteristic Function

Sampling Distribution of

# Population Distribution

- The **population distribution** is the probability distribution from which the data are drawn.
- It is described by parameters such as the mean $\mu$ and variance $\sigma^2$.
- Under random sampling, each observation $X_i$ is drawn from this population distribution.

**Implications:**

- If the population is normal, each observation is normal.
- Even if the population is not normal, averages of large samples are often approximately normal.
- The i.i.d. assumption is fundamental in statistical inference.

# Sampling Distribution of the Sample Mean

- The sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  is itself a random variable.
- Its **sampling distribution** describes how $\bar{X}$ varies across repeated samples.

# Expectation and Variance of the Sampling Distribution

- The expectation of the sample mean is

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu.$$

- Its variance is

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

- Thus, $\bar{X}$ is unbiased and becomes more concentrated around $\mu$ as $n$ grows.

# Normality of the Sample Mean

- If
  $$X_i \sim N(\mu, \sigma^2),$$
  then
  $$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$
- If the $X_i$ are not normally distributed, the Central Limit Theorem tells us that $\bar{X}$ becomes approximately normal when $n$ is large.

# Law of Large Numbers

- The **Law of Large Numbers (LLN)** states that the sample mean converges to the population mean as the sample size increases.
- **Weak Law of Large Numbers (WLLN):**

$$\bar{X}_n \xrightarrow{P} \mu.$$

- **Strong Law of Large Numbers (SLLN):**

$$\bar{X}_n \xrightarrow{a.s.} \mu.$$

- In both cases, the main message is that large samples tend to reveal the true mean.

# Intuition Behind the LLN

- The variance of the sample mean is

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

- As $n \to \infty$, this variance goes to 0.
- So the sample mean becomes more concentrated around the true mean $\mu$.
- Intuitively, averaging many observations reduces randomness.

# Markov's and Chebyshev's Inequalities

- **Markov's inequality:** If $X \geq 0$ and $a > 0$, then
$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$
- **Chebyshev's inequality:** If $X$ has mean $\mu$ and variance $\sigma^2$, then for any $k > 0$,
$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$
- These inequalities are useful for proving concentration results such as the WLLN.

# Proof of Markov's Inequality

For the indicator function $I(X \geq a)$,

$$X \geq a\, I(X \geq a).$$

Taking expectations,

$$\mathbb{E}[X] \geq \mathbb{E}[a\, I(X \geq a)] = a\, \Pr(X \geq a).$$

Therefore,

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

# Proof of Chebyshev's Inequality

Apply Markov's inequality to the non-negative random variable $(X-\mu)^2$:

$$\Pr\big((X-\mu)^2 \geq k^2\sigma^2\big) \leq \frac{\mathbb{E}[(X-\mu)^2]}{k^2\sigma^2}.$$

Since

$$\mathbb{E}[(X-\mu)^2] = \sigma^2,$$

we obtain

$$\Pr\big((X-\mu)^2 \geq k^2\sigma^2\big) \leq \frac{1}{k^2}.$$

Hence,

$$\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

# Weak Law of Large Numbers

Let $X_1, \dots, X_n$ be i.i.d. with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

We want to show that for any $\epsilon > 0$,

$$\Pr(|\bar{X}_n - \mu| \geq \epsilon) \to 0 \qquad \text{as } n \to \infty.$$

# Proof of the WLLN Using Chebyshev's Inequality

First,

$$\mathbb{E}[\bar{X}_n] = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Applying Chebyshev's inequality,

$$\Pr(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

As $n \to \infty$,

$$\frac{\sigma^2}{n\epsilon^2} \to 0.$$

Therefore,

$$\Pr(|\bar{X}_n - \mu| \geq \epsilon) \to 0.$$

So

$$\bar{X}_n \xrightarrow{P} \mu.$$

# Sample Error and Reliability of the Sample Mean

- In practice, we never have an infinite sample.
- So there is always some uncertainty in estimation.
- This remaining uncertainty is often called **sampling error**.
- Even so, the sample mean has the important property

$$\mathbb{E}[\bar{X}_n] = \mu,$$

so on average it hits the true population mean.

# Big-O and Little-o Notation

These notations describe asymptotic behaviour.

**Big-O:**

$$f(x) = O(g(x)) \quad \text{as } x \to \infty$$

means that $f(x)$ grows no faster than a constant multiple of $g(x)$ for large $x$.

**Little-o:**

$$f(x) = o(g(x)) \quad \text{as } x \to \infty$$

means that

$$\frac{f(x)}{g(x)} \to 0.$$

# Examples of Big-O and Little-o

$$3x^2 + 5x = O(x^2),$$

because the $x^2$ term dominates for large $x$.
Also,

$$x = o(x^2),$$

because

$$\frac{x}{x^2} = \frac{1}{x} \to 0.$$

# Moment Generating Function (MGF)

The **moment generating function** of a random variable $X$ is

$$M_X(t) = \mathbb{E}[e^{tX}], \qquad t \in \mathbb{R},$$

whenever it exists in a neighbourhood of 0.

Properties:

- $M_X(0) = 1$,
- if $X$ and $Y$ are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t),$$

- if the MGF exists near 0, it determines the distribution uniquely.

# Moments from the MGF

If the MGF exists, then its derivatives at 0 give the moments:

$$M_X^{(n)}(0) = \mathbb{E}[X^n].$$

This is one reason the MGF is useful in probability theory.

# Central Limit Theorem via the MGF

Let $X_1, \ldots, X_n$ be i.i.d. with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2.$$

Define the standardised sum

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

We want to show that

$$Z_n \xrightarrow{d} N(0, 1).$$

# MGF Expansion

Let

$$Y_i = \frac{X_i - \mu}{\sigma},$$

so that

$$\mathbb{E}[Y_i] = 0, \qquad \mathrm{Var}(Y_i) = 1.$$

Then

$$M_{Z_n}(t) = \left( M_Y\left( \frac{t}{\sqrt{n}} \right) \right)^n .$$

Using a Taylor expansion around $0$,

$$M_Y(t) = 1 + \frac{t^2}{2} + O(t^3).$$

Hence

$$M_Y\left( \frac{t}{\sqrt{n}} \right) = 1 + \frac{t^2}{2n} + O(n^{-3/2}).$$

# Conclusion of the CLT Argument

Therefore,

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + O(n^{-3/2})\right)^n \to e^{t^2/2}.$$

But

$$e^{t^2/2}$$

is the MGF of $N(0,1)$. Hence,

$$Z_n \xrightarrow{d} N(0,1).$$

This is the Central Limit Theorem.

# Why Use the Characteristic Function?

In some proofs, the **characteristic function** is used instead of the MGF.
The MGF is

$$M_X(t) = \mathbb{E}[e^{tX}],$$

but it may fail to exist for some distributions.
The characteristic function is

$$\varphi_X(t) = \mathbb{E}[e^{itX}],$$

and it always exists because

$$|e^{itX}| = 1.$$

# Characteristic Function

The characteristic function of $X$ is

$$\varphi_X(t) = \mathbb{E}[e^{itX}], \qquad t \in \mathbb{R}.$$

Properties:

- $\varphi_X(0) = 1$,
- $|\varphi_X(t)| \leq 1$,
- if $X$ and $Y$ are independent, then

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t),$$

- it uniquely determines the distribution.

# Small-$t$ Expansion of the Characteristic Function

If $X$ has mean $\mu$ and variance $\sigma^2$, then for small $t$,

$$\varphi_X(t) = 1 + it\mu - \frac{t^2\sigma^2}{2} + o(t^2).$$

This is useful in proofs of the Central Limit Theorem.

# Sampling Distribution of the Sample Variance

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a population with mean $\mu$ and variance $\sigma^2$.

The sample variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \qquad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- $S^2$ is an unbiased estimator of $\sigma^2$.
- Its distribution depends on the population distribution.

# Normal Case

If

$$X_i \sim N(\mu, \sigma^2),$$

then

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

So the sample variance is closely related to the chi-squared distribution.

# Expectation and Variance of the Sample Variance

The sample variance satisfies

$$\mathbb{E}[S^2] = \sigma^2.$$

In the normal case,

$$\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}.$$

As $n \to \infty$, this variance goes to 0, so $S^2$ becomes more concentrated around $\sigma^2$.

# Why Divide by $n-1$ Instead of $n$?

If we define

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

then

$$\mathbb{E}[\tilde{S}^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

So dividing by $n$ gives a biased estimator.
Dividing by $n-1$ corrects this bias.

# Key Identity

A useful identity is

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

This identity is central to understanding why the factor $n-1$ appears.

# Proof of the Identity

Start from

$$X_i - \bar{X} = (X_i - \mu) - (\bar{X} - \mu).$$

Then

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} \Big( (X_i - \mu) - (\bar{X} - \mu) \Big)^2 .$$

Expanding,

$$= \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) + \sum_{i=1}^{n} (\bar{X} - \mu)^2.$$

Now use

$$\sum_{i=1}^{n} (X_i - \mu) = n(\bar{X} - \mu),$$

to obtain

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

# Why the Degrees of Freedom Become $n - 1$

The sample mean $\bar{X}$ is estimated from the same data.
So once $n - 1$ deviations are known, the last one is determined by the condition

$$\sum_{i=1}^{n}(X_i - \bar{X}) = 0.$$

Therefore, only $n - 1$ independent deviations remain.
This is why the sample variance uses $n - 1$ rather than $n$.

# Summary

In this lecture, we studied:

- the relationship between population and sample,
- the sample mean as an estimator,
- population and sampling distributions,
- the law of large numbers,
- Markov's and Chebyshev's inequalities,
- the Central Limit Theorem,
- moment generating functions and characteristic functions,
- the sampling distribution of the sample variance.