Math Revision Session

Statistics (5): Population and Sample

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

May 28, 2025

Relationship Between Population and Sample

- **Population:** The entire set of individuals or observations under study.
  - Denoted as $\mathcal{P}$.
  - Parameters such as μ (mean) and σ^2 (variance) describe the population.
- **Sample:** A subset of the population used for analysis.
  - Denoted as $S = \{X_1, X_2, \ldots, X_n\}$.
  - Sample statistics (e.g., $\bar{X}$, s^2) estimate population parameters.
- **Key Relationships:**
  - $\mathbb{E}[\bar{X}] = \mu$ (Unbiased estimator of mean).
  - $\mathbb{E}[s^2] = \sigma^2$ (Unbiased estimator of variance).
  - Larger sample sizes provide better approximations of population characteristics.

Sample Mean and Estimation

- **Sample Mean:** Given a sample $S = \{X_1, X_2, \ldots, X_n\}$, the sample mean is defined as:
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
- **Estimator vs. Estimate:**
    - **Estimator:** A function of the sample used to infer a population parameter.
        - Example: $\bar{X}$ is an estimator of μ.
    - **Estimate:** A specific numerical value obtained from an estimator using observed data.
        - Example: If $\bar{X} = 5.2$ from a given sample, 5.2 is the estimate of μ.
- **Properties of Sample Mean:**
    - $\mathbb{E}[\bar{X}] = \mu$ (Unbiasedness).
    - Var$(\bar{X}) = \frac{\sigma^2}{n}$ (Smaller variance with larger n).
    - If $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \sigma^2/n)$.

Ex.) Coin Toss Example

- Consider a fair coin toss, where $X_i = 1$ if heads, $X_i = 0$ if tails.
- The probability of heads is p, but we do not know the true p.
- The sample mean is used to estimate p:

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

- **As a Random Variable:**
  - $\hat{p}$ varies depending on the sample.
  - $\mathbb{E}[\hat{p}] = p$ (Unbiased estimator).
  - $\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}$ (Variance decreases as n increases).
- **Law of Large Numbers:** As $n \to \infty$, $\hat{p} \to p$.

$$
\begin{aligned}
E[\hat{p}] = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
= \frac{1}{n}E[X_1 + \cdots + X_n] \\
= \frac{1}{n}\{E[X_1] + \cdots + E[X_n]\}
\end{aligned}
$$

The variance is:

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

Remamber that: $E[X_i] = 0 \times (1-p) + 1 \times p = p$ and $Var[X_i] = E[X_i^2] - E[X_i]^2 = 0^2 \times (1-p) + 1^2 \times -p^2 = p(1-p)$.

Population Distribution

- The **population distribution** refers to the probability distribution from which data is drawn.
- It describes the characteristics of the entire population, such as mean μ and variance σ^2.
- **Random Sampling:**
  - A sample is drawn randomly from the population.
  - The selected samples inherit the properties of the population distribution.
  - Each sample is an independent observation from the population.
- **Implications:**
  - If the population follows a normal distribution, each randomly drawn sample also follows a normal distribution.
  - Even if the population is not normal, large samples tend to approximate normality due to the Central Limit Theorem.
  - The assumption of **independent and identically distributed (i.i.d.)** samples is fundamental in statistical inference.

Sampling Distribution of the Sample Mean

- The **sample mean** $\bar{X}$ is a random variable that varies across different samples.
- The **sampling distribution** of

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

describes its probability distribution over repeated sampling.

- **Expectation and Variance:**
  - $\mathbb{E}[\bar{X}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$ (Unbiased Estimator)
  - Var$(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$ (Lower variance for larger n)
- **Normality:**
  - If $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
  - If X_i is not normal, the **Central Limit Theorem** states that $\bar{X}$ approaches normality as n increases.

Law of Large Numbers

- The **Law of Large Numbers (LLN)** states that the sample mean $\bar{X}_n$ converges to the population mean μ as the sample size n increases.
- **Weak Law of Large Numbers (WLLN):**

$$\bar{X}_n \xrightarrow[n\to\infty]{P} \mu \quad \text{(Convergence in Probability)}$$

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

$$\bar{X}_n \xrightarrow[n\to\infty]{a.s.} \mu \quad \text{(Almost Sure Convergence)}$$

- **Implication:**
  - As n increases, $\bar{X}_n$ gets closer to μ.
  - The larger the sample, the more accurate the estimate of μ.
  - Ensures stability of sample estimates over repeated sampling.

Intuition Behind the Law of Large Numbers

- As the sample size n increases, the variability of the sample mean $\bar{X}_n$ decreases.
- Mathematically, the variance of the sample mean is given by:

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

  which approaches 0 as $n \to \infty$.
- Intuition:
  - The sample mean is like an archer aiming at a target:
    - With a few shots (small n), the hits are scattered.
    - With many shots (large n), the hits concentrate near the true centre (μ).
  - The law ensures that, with enough samples, the deviation from μ becomes negligible.

Markov's and Chebyshev's Inequalities

- **Markov's Inequality:** For a non-negative random variable X and any $a > 0$,
$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$
This provides an upper bound on the probability that X takes large values.
- **Chebyshev's Inequality:** For any random variable X with mean μ and variance σ^2, and for any $k > 0$,
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$
This shows that most values of X concentrate around the mean μ.
- **Importance in the CLT Proof:**
  - Markov's inequality is a general tool for bounding probabilities of extreme values.
  - Chebyshev's inequality is used to show that sample means become tightly concentrated around μ.
  - These inequalities help establish the convergence required for the Central Limit Theorem (CLT).

Proof of Markov's Inequality

- **Proof:**
  - Consider the indicator function $I(X \geq a)$, which is 1 if $X \geq a$ and 0 otherwise.
  - Since $X \geq aI(X \geq a)$, taking expectations gives:

$$\mathbb{E}[X] \geq \mathbb{E}[aI(X \geq a)] = aP(X \geq a).$$

  - Rearranging, we obtain:

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

- **Conclusion:** This inequality provides an upper bound on the probability that X takes large values based on its expected value.

Proof of Chebyshev's Inequality

- **Proof:**
  - Apply Markov's inequality to the non-negative random variable $(X-\mu)^2$ with $a = k^2\sigma^2$:

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

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

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

  - Noting that $P(|X-\mu| \geq k\sigma) = P((X-\mu)^2 \geq k^2\sigma^2)$, we conclude:

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

- **Conclusion:** This inequality provides an upper bound on the probability that X deviates significantly from its mean in terms of its variance.

Weak Law of Large Numbers (WLLN)

Statement: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean $\mathbb{E}[X_i] = \mu$ and finite variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, for any $\epsilon > 0$,

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

Proof using Chebyshev's Inequality:

- The sample mean is defined as:

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

- Compute its expectation:

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

- Compute its variance:

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

- Apply Chebyshev ' s inequality:

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

- As $n \to \infty$, the right-hand side $\frac{\sigma^2}{n\epsilon^2} \to 0$, implying

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

Conclusion: The sample mean $\bar{X}_n$ converges to μ in probability as $n \to \infty$, proving the Weak Law of Large Numbers.

Sample Error and the Reliability of Sample Mean

- In reality, we cannot take an infinite sample size n. Some uncertainty always remains in our estimate.
- This remaining uncertainty is called the **sample error**.
- Despite this, the sample mean $\bar{X}_n$ has a crucial property:

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

  meaning that, on average, the sample mean equals the true mean.
- Even with some uncertainty, the sample mean provides a reliable way to estimate an unknown population mean.

Big-O and Little-o Notation

Big-O (O) and Little-o (o) notation are used to describe the asymptotic behaviour of functions, particularly in algorithm analysis.

- **Big-O (O) Notation**:

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

  means that $f(x)$ grows at most as fast as $g(x)$ for large values of x, up to a constant factor. This gives an upper bound on the growth of $f(x)$.

- **Example of Big-O:**

$$f(x) = 3x^2 + 5x, \quad g(x) = x^2$$

  Since $f(x) = 3x^2 + 5x$ is dominated by the x^2 term for large x, we can say:

$$f(x) = O(x^2)$$

  This means that for sufficiently large x, $f(x)$ grows at most as fast as x^2, up to a constant multiple.

- **Little-o (o) Notation**:

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

means that $f(x)$ grows strictly slower than $g(x)$ as $x \to \infty$. In other words, the ratio $\frac{f(x)}{g(x)}$ tends to 0 as $x \to \infty$.

- **Example of Little-o:**

$$f(x) = x, \quad g(x) = x^2$$

Here, $f(x) = x$ grows strictly slower than $g(x) = x^2$. As x becomes large, the ratio $\frac{f(x)}{g(x)} = \frac{x}{x^2} = \frac{1}{x}$ tends to 0. Thus:

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

This means that x grows much slower than x^2 as $x \to \infty$.

Key Difference:

- Big-O provides an upper bound on growth (asymptotic upper bound).
- Little-o indicates that one function grows strictly slower than the other.

Moment Generating Function (MGF)

Definition: The moment generating function (MGF) of a random variable X is defined as

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

Properties:

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

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

- The nth derivative at $t = 0$ gives the nth moment:

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

- If $M_X(t)$ exists in an open interval around $t = 0$, it uniquely determines the distribution.

Importance in CLT:

- Used to analyse the limiting behaviour of sums of i.i.d. random variables.
- Helps show that standardised sums converge to a normal distribution.
- Provides an alternative proof of the Central Limit Theorem.

Central Limit Theorem via Moment Generating Function

Setup: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean $\mathbb{E}[X_i] = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2$. Consider the standardised sum:

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

Step 1: Compute the MGF of Z_n

$$M_{Z_n}(t) = \mathbb{E}\left[e^{tZ_n}\right].$$

Using independence and the MGF of each X_i,

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

where $Y_i = \frac{X_i - \mu}{\sigma}$ has mean 0 and variance 1.

Step 2: Expand the MGF using Taylor series

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

Substituting $t/\sqrt{n}$:

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

Raising to the power n, we get

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

Using $(1+x)^n \approx e^{nx}$, this converges to

$$M_{Z_n}(t) \to e^{t^2/2}, \quad \text{as } n \to \infty.$$

Conclusion: Since $e^{t^2/2}$ is the MGF of $N(0,1)$, we conclude

$$Z_n \xrightarrow[n\to\infty]{d} N(0,1).$$

In some proofs, the **characteristic function** is used instead of the moment generating function (MGF).

- The moment generating function (MGF) is defined as:

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

  However, this function does not always exist for all values of t, making it unsuitable in some cases.
- The characteristic function is defined as:

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

  Unlike the MGF, the characteristic function always exists for any random variable because $|e^{itX}| = 1$.
- **Why use the characteristic function?**
  - It is always well-defined, even when the MGF does not exist.
  - It uniquely determines the distribution of a random variable.
  - It simplifies proofs, especially in limit theorems (e.g., Central Limit Theorem).

Conclusion: While MGFs are useful when they exist, characteristic functions provide a more general and robust approach in many theoretical proofs.

Characteristic Function

Definition: The characteristic function $\phi_X(t)$ of a random variable X is given by

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

Properties:

- $\phi_X(0) = 1$.
- $|\phi_X(t)| \leq 1$ for all t.
- If X and Y are independent, then

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

- If X has mean μ and variance σ^2, then for small t,

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

- The characteristic function uniquely determines the probability distribution.

Importance in CLT:

- Used to analyse the limiting distribution of standardised sums.
- Provides an elegant proof of the Central Limit Theorem.

Sampling Distribution of Sample Variance

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from a population with mean μ and variance σ^2. The sample variance is defined as:

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

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean.

- S^2 is an unbiased estimator of the population variance σ^2.
- The distribution of S^2 depends on the population distribution.
- If $X_i \sim N(\mu, \sigma^2)$, then:

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

  follows a chi-square distribution with $n-1$ degrees of freedom.

Expectation and Variance of Sample Variance

Expectation of S^2:

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

This shows that S^2 is an **unbiased estimator** of σ^2.

Variance of S^2:

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

- As $n \to \infty$, Var$(S^2) \to 0$, meaning S^2 becomes more concentrated around σ^2.
- The larger the sample size, the more precise the estimate.

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

Sample variance is defined as:

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

If we divide by n, the resulting estimator is biased, meaning:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Reason:

- The sample mean $\bar{X}$ is an estimate of μ, not the true mean.
- This introduces one constraint: once $n-1$ values are chosen, the last one is determined.
- This reduces the effective degrees of freedom from n to $n-1$.

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

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

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

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

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

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

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

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}(X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2
\end{aligned}
$$

Consistency of Sample Variance

Definition of Consistency: An estimator $\hat{\theta}_n$ is consistent for θ if:

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{as } n \to \infty$$

That is, $\hat{\theta}_n$ converges to θ in probability.

Consistency of S^2:

- Since $\mathbb{E}[S^2] = \sigma^2$ and $\mathsf{Var}(S^2) \to 0$ as $n \to \infty$,
- By Chebyshev's inequality:

$$P(|S^2 - \sigma^2| \geq \epsilon) \leq \frac{\mathsf{Var}(S^2)}{\epsilon^2} = \frac{2\sigma^4}{(n-1)\epsilon^2} \to 0$$

- Therefore, S^2 is a consistent estimator of σ^2.

Conclusion:

- S^2 is both an **unbiased** and **consistent** estimator of σ^2.
- As the sample size increases, the estimation accuracy improves.

Sample Variance under Normal Distribution

- Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. from $N(\mu, \sigma^2)$.
- The sample variance is defined as:

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

- We examine the distribution of S^2 and its relation to the chi-square and t-distributions.

Chi-Square Distribution of Squared Deviations

- Define the sum of squared deviations from the population mean:

$$W = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 .$$

- Since $X_i \sim N(\mu, \sigma^2)$, it follows that:

$$W \sim \chi^2(n).$$

- However, in practice, we do not know μ and use $\bar{X}$ instead.

Sample Variance and Chi-Square Distribution

- Using $\bar{X}$, define:

$$W' = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 .$$

- It can be shown that:

$$W' \sim \chi^2(n-1).$$

- Hence, the sample variance satisfies:

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

t-Distribution

- Define the standardized sample mean:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

- Since $\bar{X} \sim N(\mu, \sigma^2/n)$, it follows that $Z \sim N(0, 1)$.
- The t-statistic is defined as:

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

- Using the previous result, we obtain:

$$T \sim t(n-1).$$