Multivariate Random Variable & Estimation

Prof. Pranay Kumar Saha

March 13, 2025

What is a Multivariate Random Variable?

- A **random variable** is a measurable function from a sample space to the real numbers.
- A **multivariate random variable** (or random vector) consists of more than one random variable considered together.
- Examples:
  - Bivariate: (X, Y)
  - Multivariate: $(X_1, X_2, \dots, X_n)$
- Studying multiple random variables jointly allows us to capture their dependence or independence.

Joint PMF/PDF

Discrete Case: Joint PMF

$$p_{X,Y}(x,y) \; = \; \Pr(X = x,\; Y = y).$$

Continuous Case: Joint PDF

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} \Pr(X \le x,\; Y \le y).$$

Marginal Probability (Distribution)

- For **discrete** random variables X and Y with a joint PMF $p_{X,Y}(x, y)$:

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad p_Y(y) = \sum_x p_{X,Y}(x, y).$$

- For **continuous** random variables X and Y with a joint PDF $f_{X,Y}(x, y)$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

- **Interpretation:** The marginal distribution of one variable is obtained by "summing out" or "integrating out" the other variable(s).

Conditional Probability (Distribution)

Conditional Probability (Discrete case):

$$p_{Y|X}(y \mid x) \;=\; \frac{p_{X,Y}(x,y)}{p_X(x)}.$$

- This is read as: "The probability that $Y = y$ given $X = x$."

Conditional Density (Continuous case):

$$f_{Y|X}(y \mid x) \;=\; \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

- The conditional PDF describes how Y is distributed once we know the value of X.

Interpretation and Usage

- **Marginal distribution:**
  - Describes the distribution of a single variable independently of the others.
  - Example: If you're only interested in X (regardless of Y), you use $p_X(x)$ or $f_X(x)$.
- **Conditional distribution:**
  - Describes the distribution of one variable *given* information about another.
  - Example: If you know $X = x$, it updates your understanding of how Y behaves.
- These two ideas are linked by:

$$p_{X,Y}(x, y) = p_X(x)\, p_{Y|X}(y \mid x) \quad \text{or} \quad f_{X,Y}(x, y) = f_X(x)\, f_{Y|X}(y \mid x).$$

Discrete Bivariate Random Variable

Consider two discrete random variables X and Y, each taking values 0 or 1, with the joint PMF:

	$Y = 0$	$Y = 1$
$X = 0$	0.2	0.3
$X = 1$	0.1	0.4

Marginal PMFs:

$$P(X = 0) = 0.2 + 0.3 = 0.5, \quad P(X = 1) = 0.1 + 0.4 = 0.5$$

$$P(Y = 0) = 0.2 + 0.1 = 0.3, \quad P(Y = 1) = 0.3 + 0.4 = 0.7$$

Conditional PMF:

$$P(X = 1|Y = 1) = \frac{P(X = 1, Y = 1)}{P(Y = 1)} = \frac{0.4}{0.7} \approx 0.571$$

Discrete Bivariate Random Variable

Covariance:

$$E[XY] = 0 \cdot 0 \cdot 0.2 + 0 \cdot 1 \cdot 0.3 + 1 \cdot 0 \cdot 0.1 + 1 \cdot 1 \cdot 0.4 = 0.4$$

$$E[X] = 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5, \quad E[Y] = 0 \cdot 0.3 + 1 \cdot 0.7 = 0.7$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.4 - 0.5 \cdot 0.7 = 0.05$$

Continuous Bivariate Random Variable

Let X and Y be continuous random variables with joint PDF:

$$f(x, y) = x + y \quad \text{for} \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

and 0 otherwise. This is a valid PDF since:

$$\int_0^1 \int_0^1 (x + y)\, dx\, dy = 1$$

Marginal PDFs:

$$f_X(x) = \int_0^1 (x + y)\, dy = x + \frac{1}{2} \quad \text{for} \quad 0 \leq x \leq 1$$

$$f_Y(y) = \int_0^1 (x + y)\, dx = y + \frac{1}{2} \quad \text{for} \quad 0 \leq y \leq 1$$

Continuous Bivariate Random Variable

Since $f(x, y) \neq f_X(x)f_Y(y)$, e.g., $f(0,0) = 0 \neq \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$, the variables are dependent. **Example Probability:**

$$P(X \leq 0.5, Y \leq 0.5) = \int_0^{0.5} \int_0^{0.5} (x + y)\, dx\, dy = 0.125$$

Compare: $P(X \leq 0.5) = \int_0^{0.5} \left(x + \frac{1}{2}\right) dx = 0.375$, and similarly $P(Y \leq 0.5) = 0.375$, so:

$$P(X \leq 0.5)P(Y \leq 0.5) = 0.375^2 = 0.140625 \neq 0.125$$

This confirms dependence.

Example: A Discrete Bivariate Random Variable

Setup: Suppose we have two discrete random variables X and Y, each taking values in $\{0, 1\}$. Their joint PMF is given by:

$$p_{X,Y}(x, y) = \begin{cases} 0.1, & (x = 0,\ y = 0), \\ 0.3, & (x = 0,\ y = 1), \\ 0.2, & (x = 1,\ y = 0), \\ 0.4, & (x = 1,\ y = 1). \end{cases}$$

- Note that $0.1 + 0.3 + 0.2 + 0.4 = 1$, so it is a valid probability distribution.
- We will use this table to illustrate how to compute marginal and conditional probabilities.

Compute Marginals

Marginal of X:

$$p_X(0) = p_{X,Y}(0,0) + p_{X,Y}(0,1) = 0.1 + 0.3 = 0.4,$$

$$p_X(1) = p_{X,Y}(1,0) + p_{X,Y}(1,1) = 0.2 + 0.4 = 0.6.$$

So,

$$p_X(x) = \begin{cases} 0.4, & x = 0, \\ 0.6, & x = 1. \end{cases}$$

Compute Marginals

Marginal of Y:

$$p_Y(0) = p_{X,Y}(0,0) + p_{X,Y}(1,0) = 0.1 + 0.2 = 0.3,$$

$$p_Y(1) = p_{X,Y}(0,1) + p_{X,Y}(1,1) = 0.3 + 0.4 = 0.7.$$

So,

$$p_Y(y) = \begin{cases} 0.3, & y = 0, \\ 0.7, & y = 1. \end{cases}$$

Compute Conditional Probabilities

Conditionals of Y given X:

$$p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x,y)}{p_X(x)}.$$

- For $X = 0$:

$$p_{Y|X}(0 \mid 0) = \frac{p_{X,Y}(0,0)}{p_X(0)} = \frac{0.1}{0.4} = 0.25,$$

$$p_{Y|X}(1 \mid 0) = \frac{p_{X,Y}(0,1)}{p_X(0)} = \frac{0.3}{0.4} = 0.75.$$

Compute Conditional Probabilities

- For $X = 1$:

$$p_{Y|X}(0 \mid 1) = \frac{p_{X,Y}(1,0)}{p_X(1)} = \frac{0.2}{0.6} \approx 0.33,$$

$$p_{Y|X}(1 \mid 1) = \frac{p_{X,Y}(1,1)}{p_X(1)} = \frac{0.4}{0.6} \approx 0.67.$$

Interpretation of the Conditional Distributions

- If you know $X = 0$, then the probability that $Y = 0$ is 0.25, and $Y = 1$ is 0.75.
- If you know $X = 1$, then the probability that $Y = 0$ is about 0.33, and $Y = 1$ is about 0.67.
- This shows how knowledge of X *changes* the distribution of Y.

Are X and Y independent?

- Independence requires $p_{X,Y}(x, y) = p_X(x)\,p_Y(y)$ for all (x, y).
- For instance, $p_{X,Y}(0, 0) = 0.1$, but $p_X(0)p_Y(0) = 0.4 \times 0.3 = 0.12 \neq 0.1$.
- Therefore, X and Y are **not independent**.

A Continuous Example

Bivariate Uniform on $[0,1]^2$:

$$f_{X,Y}(x,y) = 1, \quad 0 \le x \le 1,\ 0 \le y \le 1,$$

and 0 otherwise.

- The marginal PDFs are:

$$f_X(x) = \int_0^1 1\,dy = 1, \quad (0 \le x \le 1),$$

$$f_Y(y) = \int_0^1 1\,dx = 1, \quad (0 \le y \le 1).$$

A Continuous Example

- The conditionals are:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{1}{1} = 1, \quad 0 \leq y \leq 1.$$

- Here, X and Y are *independent* since $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

Independence

Two random variables X and Y are independent if:

$$\text{Discrete:} \quad p_{X,Y}(x,y) = p_X(x)\, p_Y(y),$$

$$\text{Continuous:} \quad f_{X,Y}(x,y) = f_X(x)\, f_Y(y).$$

- **Interpretation:** Knowing one does not change the distribution of the other.
- For $\{X_1, X_2, \ldots, X_n\}$ to be **jointly independent**, *every* subset must also be independent.
- Independence $\implies$ Zero Covariance, but zero covariance does *not* necessarily imply independence.

Expectation of a Function of Two RVs

Discrete Case:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y)\, p_{X,Y}(x,y).$$

Continuous Case:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy.$$

- Special cases:
  - $E[X]$, $E[Y]$
  - Mixed moments like $E[XY]$

Covariance and Correlation

Covariance

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]\,E[Y].$$

- $\text{Cov}(X, Y) = 0$ implies X and Y are **uncorrelated**.

Correlation Coefficient

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}} \quad \in [-1, 1].$$

- $\rho_{X,Y} = 0$ means no *linear* correlation, but not necessarily independence.

Conditional Distribution

Discrete:

$$p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Continuous:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

- Similar definitions hold for $p_{X|Y}(x \mid y)$ or $f_{X|Y}(x \mid y)$.
- The shape of one variable's distribution can change once the other is known.

Conditional Expectation

$$E[Y \mid X = x] = \begin{cases} \sum_y y \, p_{Y|X}(y \mid x), & \text{(discrete)} \\ \int_{-\infty}^{\infty} y \, f_{Y|X}(y \mid x) \, dy, & \text{(continuous)} \end{cases}$$

- This is a *function* of x, often denoted $g(x)$.
- The **Law of Total Expectation**:

$$E[Y] = E\big[E[Y \mid X]\big].$$

Multivariate Normal Distribution

- A random vector $\mathbf{X} \in \mathbb{R}^n$ is **multivariate normal** if any linear combination of its components is normally distributed.
- Specified by:
  - Mean vector $\mu \in \mathbb{R}^n$
  - Covariance matrix Σ ($n \times n$ and positive semi-definite)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right).$$

Multinomial Distribution (Discrete)

- Generalization of the Binomial distribution to k categories.
- Suppose you have n independent trials, each trial resulting in one of k categories with probabilities $p_1, \ldots, p_k$.
- Let X_i be the count of trials that fall in category i, so $\sum_{i=1}^{k} X_i = n$.

$$p(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} \, p_1^{x_1} \cdots p_k^{x_k}.$$

Point Estimation

What is Point Estimation?

Point estimation is a statistical method that uses sample data to calculate a single value, called a **point estimate**, to approximate an unknown population parameter.

- **Population**: The entire group of interest (e.g., all students in a school).
- **Sample**: A smaller subset of the population.
- **Population Parameter**: A numerical value describing the population (e.g., mean μ, variance σ^2, proportion p).

A point estimate is a single number calculated from the sample to estimate the parameter.

Why is Point Estimation Important?

Point estimation is the foundation for statistical inference, which involves drawing conclusions about a population based on a sample. It is widely used in:

- Decision-making (e.g., estimating average customer wait times).
- Modeling (e.g., providing initial values for regression coefficients).
- Understanding data (e.g., summarizing population traits).

However, a point estimate alone does not provide information about its uncertainty. For that, we use interval estimation (e.g., confidence intervals).

Properties of Good Point Estimators

Good point estimators should have certain properties:

1. **Unbiasedness**: On average, the estimator equals the true parameter.
2. **Consistency**: The estimator gets closer to the true parameter as the sample size increases.
3. **Efficiency**: Among unbiased estimators, it has the smallest variance.
4. **Robustness**: It is less affected by outliers or violations of assumptions.

Unbiasedness

An estimator T is **unbiased** if its expected value equals the true parameter θ:

$$E[T] = \theta$$

Example: The sample mean $\bar{X} = \frac{1}{n}\sum X_i$ is an unbiased estimator of the population mean μ because $E[\bar{X}] = \mu$. Unbiasedness ensures the estimator does not systematically over- or underestimate the parameter.

Consistency

An estimator T is **consistent** if it converges to the true parameter θ as the sample size n increases:

$$T \xrightarrow{p} \theta \quad \text{as} \quad n \to \infty$$

Example: By the Law of Large Numbers, the sample mean $\bar{X}$ is consistent for μ. Consistency ensures the estimator becomes more accurate with larger samples.

Efficiency

Among unbiased estimators, the **most efficient** one has the smallest variance. The **Cramér-Rao Lower Bound** provides the minimum possible variance for an unbiased estimator. **Example**: For a normal distribution, the sample mean $\bar{X}$ is efficient because it achieves the Cramér-Rao bound. Efficiency ensures the estimator is as precise as possible.

Methods to Compute Point Estimates

Two common methods are:

1. **Method of Moments**: Match sample moments to population moments.
2. **Maximum Likelihood Estimation (MLE)**: Choose the parameter that maximizes the likelihood of observing the sample data.

Introduction to Moments in Statistics

Moments are statistical measures that capture key properties of a probability distribution, helping us understand the behavior of a random variable.

Sample Moments:

- For a sample $\{X_1, X_2, \ldots, X_n\}$, the k-th sample moment is:

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

- This is computed directly from the sample data to estimate population moments.

Moments provide insights into the shape, central tendency, and spread of a distribution, making them essential tools in statistics.

Method of Moments

The method of moments involves equating sample moments to population moments and solving for the parameter. **Example**: For a normal distribution:

- Population mean: μ. Estimate with sample mean $\bar{X}$.
- Population variance: σ^2. Estimate with sample variance $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$.

This method is simple but may not always be the most efficient.

Problem Statement

Given

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where both μ and σ are unknown.

Objective

Find Method of Moment (MoM) estimators for μ and σ.

Method of Moments: Concept

Key Idea

The Method of Moments equates sample moments to the corresponding population moments to derive parameter estimators.

- The k^{th} population moment is defined as $\mu'_k = E[X^k]$
- The k^{th} sample moment is $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
- For p unknown parameters, we need p equations by equating the first p moments

Population Moments for Normal Distribution

Normal Distribution Properties

For $X \sim N(\mu, \sigma^2)$:

- First moment (mean): $\mu_1' = E[X] = \mu$
- Second moment: $\mu_2' = E[X^2] = \mu^2 + \sigma^2$

Note

Higher moments also exist, but for our problem with two parameters (μ and σ), we only need the first two moments.

Step 1: Calculate Sample Moments

Sample Moments

From our random sample $X_1, X_2, \ldots, X_n$, we calculate:

- First sample moment (sample mean):

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

- Second sample moment:

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Step 2: Equate Sample and Population Moments

Moment Equations

Setting sample moments equal to population moments:

$$m_1 = \mu_1'$$

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

$$m_2 = \mu_2'$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2$$

Step 3: Solve for Parameter Estimators

$$\hat{\mu} = m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\begin{aligned}
\mu^2 + \sigma^2 &= m_2 \\
\sigma^2 &= m_2 - \mu^2 \\
&= m_2 - (m_1)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2
\end{aligned}$$

Step 3: Solve for Parameter Estimators

Method of Moment Estimator for σ

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

Alternative Expression for $\hat{\sigma}^2$

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

Method of Moment Estimators: Final Result

MoM Estimators for Normal Distribution

For a random sample $X_1, X_2, \ldots, X_n$ from $N(\mu, \sigma^2)$:

$$\hat{\mu}_{MoM} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2_{MoM} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\hat{\sigma}_{MoM} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Maximum Likelihood Estimation (MLE)

MLE chooses the parameter value that maximizes the likelihood function, which is the probability of observing the sample data. **Example**: For Bernoulli trials (e.g., coin flips), the sample proportion $\hat{p} = \frac{\text{number of successes}}{n}$ is the MLE for p. MLEs are often consistent and efficient, especially with large samples, but can be biased in small samples.

Example 1: Estimating a Population Mean

Suppose you survey 5 people about their commute times (in minutes): 20, 25, 30, 15, 40. **Sample Mean**: $\bar{X} = \frac{20+25+30+15+40}{5} = 26$ minutes. **Point Estimate**: 26 minutes estimates the population mean μ. This estimator is unbiased, consistent, and efficient (for normal data).

Example 2: Estimating a Population Proportion

You inspect 100 light bulbs and find 5 defective. **Sample Proportion**: $\hat{p} = \frac{5}{100} = 0.05$. **Point Estimate**: 0.05 estimates the population proportion p. This estimator is unbiased, consistent, and approximately efficient for large n.

Mathematical Representation: MLE

likelihood

Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with a probability density function (PDF) or probability mass function (PMF) $f(x;\theta)$, where θ represents the parameter(s) to estimate. The **likelihood function** is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Here, $f(x;\theta)$ is the PDF for continuous distributions or the PMF for discrete ones. The likelihood $L(\theta)$ measures how well the model with parameter θ fits the observed data.

Maximizing the Likelihood

The MLE, denoted $\hat{\theta}$, is the value of θ that maximizes the likelihood:

$$\hat{\theta} = \arg\max_{\theta} L(\theta)$$

Since products can be complex to optimize, we often maximize the **log-likelihood** instead:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$$

Because the logarithm is a monotonic function, maximizing $\ell(\theta)$ yields the same result as maximizing $L(\theta)$.

Mathematical Procedure to Find the MLE

1. Write the likelihood function $L(\theta)$ based on the sample and distribution.
2. Compute the log-likelihood $\ell(\theta) = \ln L(\theta)$.
3. Differentiate $\ell(\theta)$ with respect to θ:
$$\frac{d\ell(\theta)}{d\theta}$$
4. Set the derivative to zero to find critical points:
$$\frac{d\ell(\theta)}{d\theta} = 0$$
5. Solve for θ to get $\hat{\theta}$.
6. Confirm it's a maximum by ensuring:
$$\frac{d^2\ell(\theta)}{d\theta^2} < 0$$

For multiple parameters, use partial derivatives to solve a system of equations.

Key Takeaways

- Point estimation uses sample data to estimate population parameters with a single value.
- Good estimators are unbiased, consistent, and efficient.
- Common methods include the method of moments and MLE.
- Practice calculating point estimates and understanding their properties.

Next steps: Explore interval estimation (e.g., confidence intervals) to quantify uncertainty.

What is Interval Estimation?

- **Interval estimation** is a range of plausible values for a parameter (e.g., a mean or a probability) constructed from observed data.
- In contrast to a *point estimate*, which gives a single value, an *interval* provides a measure of uncertainty or confidence.
- Often reported as **confidence intervals** (CIs).

Confidence Interval: Basic Concept

- A **confidence interval** for a parameter θ is given by random interval $[\mathsf{L}(\mathbf{X}), \mathsf{U}(\mathbf{X})]$, where **X** represents the sample data.
- In the case of a $(1-\alpha)\times 100\%$ confidence interval, we have:

$$\mathsf{Pr}\big(\mathsf{L}(\mathbf{X}) \leq \theta \leq \mathsf{U}(\mathbf{X})\big) = 1-\alpha,$$

  for repeated sampling from the same population.
- Commonly, $\alpha = 0.05$ for a 95% confidence interval.

Types of Confidence Intervals

- **For Mean (when population standard deviation is known)**:

$$\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$

- **For Mean (when population standard deviation is unknown)**:

$$\bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

  where t^* is the critical value from the t-distribution.

- **For Proportion**:

$$\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where:

- $\hat{p}$ is the sample proportion.
- z^* is the critical value from the standard normal distribution.

Example: Confidence Interval for a Mean (Normal Case)

Assumptions:

- $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a sample from a Normal distribution $N(\mu, \sigma^2)$.
- σ^2 is known, for simplicity (Z-interval).

Example: Confidence Interval for a Mean (Normal Case)

Steps to build a $(1-\alpha)\times 100\%$ CI for μ:

1. Compute sample mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. Use standard error SE $= \frac{\sigma}{\sqrt{n}}$.
3. The $(1-\alpha)\times 100\%$ **Z-interval** is:

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

where $z_{\alpha/2}$ is the critical value from the standard normal distribution (e.g., $z_{0.025} \approx 1.96$ for a 95% CI).

Example

Suppose you conduct a survey and find the average height of 100 individuals to be 170 cm with a sample standard deviation of 10 cm. You want to estimate the average height of the population with a 95% confidence level.

$$\bar{x} = 170, \quad s = 10, \quad n = 100, \quad \text{Confidence Level} = 95\%$$

Since the population standard deviation is unknown, use the t-distribution:

$$SE = \frac{10}{\sqrt{100}} = 1$$

With a 95% confidence level and 99 degrees of freedom ($n - 1$), the t-value is approximately 1.984.

$$CI = 170 \pm 1.984 \cdot 1 = 170 \pm 1.984$$

$$CI = [168.016, 171.984]$$

Interpretation of the Confidence Interval

- If we repeat the experiment many times:
  - A certain percentage (e.g., 95%) of the intervals computed will contain the true mean μ.
- *Important*: The parameter μ is *fixed*; it's the interval that varies from sample to sample.
- Similar ideas extend to more complex scenarios (unknown σ^2, or building intervals for other parameters like proportions, difference of means, regression coefficients, etc.).