STAT 423/523 Statistical Methods for Engineers and Scientists

Lecture 2: Point Estimation I

Chencheng Cai

Washington State University

Sample Mean

Proposition

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and standard deviation σ. Then

- $E(\bar{X}) = \mu$
- $Var(\bar{X}) = \frac{\sigma^2}{n}$

In addition, with $T = X_1 + \cdots + X_n$, we have

- $E(T) = n\mu$
- $Var(T) = n\sigma^2$

Interpretation:
The sample mean's expectation is the population mean, and its variance is the population variance divided by the sample size.

Sample Mean — Concepts

- **Population**: In statistics, a population is the entire pool from which a statistical sample is drawn. It is the complete set of individuals or objects that we are interested in.
- **Sample**: A sample is a subset of the population. It is the group of individuals or objects that we actually collect data from.
- **Random Sample**: A random sample is a sample in which each individual or object in the population has an equal chance of being selected.
- An alternative expression is $X_1, X_2, \ldots, X_n$ *are independent and identically distributed (i.i.d.) random variables with mean* μ *and variance* σ^2*.*

Sample Mean — Justification

- By linearity of expectation, we have

$$E(T) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = n\mu.$$

- By independence, we have

$$Var(T) = Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = n\sigma^2$$

- Since $\bar{X} = T/n$, we have

$$E(\bar{X}) = E(T/n) = E(T)/n = \mu$$

and

$$Var(\bar{X}) = Var(T/n) = Var(T)/n^2 = \sigma^2/n.$$

Example: Bernoulli and Binomial

Suppose we have an unfair coin whose probability of landing heads is p.
We toss the coin n times and let X_i be the indicator of the i-th toss.

- $X_i = 1$ if the i-th toss is head
- $X_i = 0$ if the i-th toss is tail

X_i follows a **Bernoulli distribution** with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$.

- $E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = p$
- $Var(X_i) = E(X_i^2) - [E(X_i)]^2 = E(X_i) - [E(X_i)]^2 = p - p^2 = p(1-p)$

Example: Bernoulli and Binomial

Let $T = X_1 + X_2 + \cdots + X_n$ be the number of heads from n tosses.
By definition, T follows a **binomial distribution** with parameters n and p, denoted as $T \sim \text{Bin}(n, p)$. (number of successes in n independent Bernoulli trials)
From our previous statement:

- $E(T) = n \cdot E(X_i) = np$
- $Var(T) = n \cdot Var(X_i) = np(1 - p)$

Similary, let $\bar{X} = T/n$ be the proportion of heads from n tosses. Then

- $E(\bar{X}) = E(X_i) = p$
- $Var(\bar{X}) = Var(X_i)/n = p(1 - p)/n$

Normal Population Distribution

Proposition

Let $X_1, X_2, \ldots, X_n$ be a random sample from a **normal** *distribution with mean μ and standard deviation σ. Then for any n, $\bar{X}$ is normally distributed with mean μ and variance σ^2/n.*

A random variable X is said to have a normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, if its probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then

$$c_1 X_1 + c_2 X_2 \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2^2\sigma_2^2).$$

Example

The distribution of egg weights of certian type is normal with mean value 53 and standard deviation 0.3.
Let $X_1, X_2, \ldots, X_{12}$ be the weights of a dozen randomly selected eggs.
Let $T = X_1 + X_2 + \cdots + X_{12}$ be the total weight of the dozen eggs.

$$E(T) = 12 \times 53 = 636, \quad Var(T) = 12 \times 0.3^2 = 1.08.$$

The probability that the total weight of the dozen eggs is between 635 and 640 is

$$P(635 < T < 640) = P\left(\frac{635 - 636}{\sqrt{1.08}} < Z < \frac{640 - 636}{\sqrt{1.08}}\right) = P(-0.96 < Z < 3.85) = 0.8315,$$

where $Z \sim N(0,1)$ follows the **standard normal distribution**.

Central Limit Theorem

Theorem (Central Limit Theorem (CLT))

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. Then if n is sufficiently large, $\bar{X}$ has approximately a normal distribution with mean μ and variance σ^2/n, and T also has approximately a normal distribution with mean $n\mu$ and variance $n\sigma^2$. The larger the value of n, the better the approximation.

A shorter version:

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1) \quad \text{as } n \to \infty$$

Central Limit Theorem

- **Implicit Assumption**: The population distribution has finite mean and finite variance.
- **Sequential Interpretation**: The CLT applies to a sequence of i.i.d. random variables.
- **Reparametrization**: When n is large enough, $Y = \sqrt{n}(\bar{X} - \mu)/\sigma$ is approximately standard normal, where the limit distribution does not depend on n.
- **Approximation**: The approximated distribution should be interpreted that the c.d.f. of Y, $P(Y \leq t)$, converges to the c.d.f. of $N(0,1)$ as $n \to \infty$ for any t.
- **Proof**: The proof of the Central Limit Theorem is beyond the scope of this course. It is a result from the characteristic function and the Lévy's convergence theorem.
- **Rule of Thumb**: $n \geq 30$ is often considered as a sufficiently large sample size.

Example

Let Y be a Binomial random variable with parameters $n = 100$ and $p = 0.5$.
We want to estimate the probability $P(40 < Y < 60)$.

Recall our discussion on tossing a coin. Let X_i be the indicator of the i-th toss.
Then $T = X_1 + X_2 + \cdots + X_{100}$ follows a Binomial distribution with parameters $n = 100$ and $p = 0.5$. That is, $T \sim Y$.
From the central limit theorem,

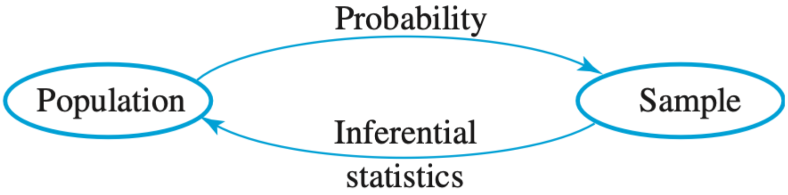
$$\bar{X} \approx N(p, p(1-p)/n) \sim N(0.5, 0.0025).$$

Therefore, $T = n\bar{X} \sim N(50, 25)$.
We have

$$P(40 < Y < 60) = P(40 < T < 60) \approx P\left(\frac{40-50}{\sqrt{25}} < Z < \frac{60-50}{\sqrt{25}}\right) = P(-2 < Z < 2)$$

From Probabilitic Model to Statistical Inference

- **Probabilitic Model**:
  Given the (known) parameters of the population (e.g. mean and variance), we can model the distribution of the sample.
- **Statistical Inference**:
  Given the sample, we want to estimate the(unknown) parameters of the population.



Point Estimation

A **point estimate** of a parameter θ is a single number that can be regarded as a sensible value of θ. It is obtained by selecting a suitable sta tistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ.

Terminology:

- **Parameter / Estimand**: A numerical characteristic of a population.
- **Estimator**: A function of the sample data used to estimate a parameter.
- **Estimate**: The value of the estimator computed from the sample data.

Properties:

- Estimand is usually a fixed and unknown value.
- Estimator is a random variable whose value depends on the sample data.
- Estimate is a realization of the estimator.

Example

- An automobile manufacturer has developed a new type of bumper.
- The manufacturer has used this bumper in a sequence of $n = 25$ controlled crashes against a wall, each at 10 mph, using one of its compact car models.
- Let $X =$ the number of crashes that result in no visible damage to the automobile.
- The estimand is the probability of no visible damage in a crash, denoted as p.
- The estimator is

$$\hat{p} = \frac{X}{n}$$

- If X is observed to be $x = 15$, the estimate is

$$\frac{x}{n} = \frac{15}{25} = 0.6$$

Example

X = voids filled with asphalt(%) for 52 specimens of a certain type of hot-mix asphalt:

74.33	71.07	73.82	77.42	79.35	82.27	77.75	78.65	77.19
74.69	77.25	74.84	60.90	60.75	74.09	65.36	67.84	69.97
68.83	75.09	62.54	67.47	72.00	66.51	68.21	64.46	64.34
64.93	67.33	66.08	67.31	74.87	69.40	70.83	81.73	82.50
79.87	81.96	79.51	84.12	80.61	79.89	79.70	78.74	77.28
79.97	75.09	74.38	77.67	83.73	80.39	76.90		

- Estimand: the variance of the voids filled with asphalt.

Example

- Estimator 1: the sample variance

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

- The estimate is

$$s^2 = \frac{\sum_{i=1}^{52}(x_i - \bar{x})^2}{52-1} = 41.126$$

- Esimator 2:

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

- The estimate is

$$s^2 = \frac{\sum_{i=1}^{52}(x_i - \bar{x})^2}{52} = 40.336$$

Evaluate an Estimator

Recall θ is the parameter to be estimated, $\hat{\theta}$ is an estimator.

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

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta.$$

- The **variance** of an estimator $\hat{\theta}$ is defined as

$$\text{Var}(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2] = E(\hat{\theta}^2) - E(\hat{\theta})^2.$$

- The **standard error** of an estimator $\hat{\theta}$ is defined as

$$\text{se}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}.$$

- The **mean squared error** (MSE) of an estimator $\hat{\theta}$ is defined as

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2.$$

Example

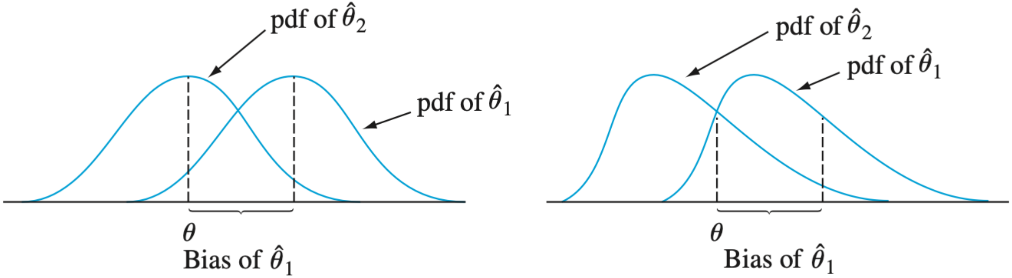
Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. We want to estimate μ.

- Estimator 1: $\hat{\mu} = X_1$.
  Bias: 0, Variance: σ^2, MSE: σ^2.
- Estimator 2: $\hat{\mu} = 0$.
  Bias: $-\mu$, Variance: 0, MSE: μ^2.
- Estimator 3: $\hat{\mu} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \bar{X}$.
  Bias: 0, Variance: $\frac{\sigma^2}{n}$, MSE: $\frac{\sigma^2}{n}$.
- Estimator 4: $\hat{\mu} = \alpha \bar{X}$ for a constant $0 < \alpha < 1$.
  Bias: $(\alpha - 1)\mu$, Variance: $\frac{\alpha^2 \sigma^2}{n}$, MSE: $(1-\alpha)^2 \mu^2 + \alpha^2 \frac{\sigma^2}{n}$.

Unbiased Estimator

An estimator with zero bias is called an **unbiased estimator**. That is, an estimator $\hat{\theta}$ is unbiased if

$$E(\hat{\theta}) = \theta.$$



Unbiased Estimator

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2.

Proposition

The sample mean $\bar{X} = n^{-1} \sum_i X_i$ is an unbiased estimator of the population mean μ. That is,

$$E(\bar{X}) = \mu.$$

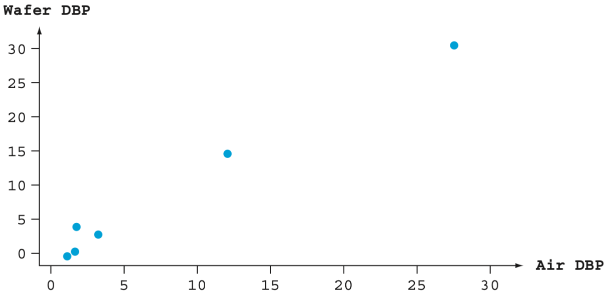
The sample variance $S^2 = (n-1)^{-1} \sum_i (X_i - \bar{X})^2$ is an unbiased estimator of the population variance σ^2. That is,

$$E(S^2) = \sigma^2.$$

The proposition also implies $n^{-1} \sum_i (X_i - \bar{X})^2$ is biased for the population variance σ^2.

Example (textbook Example 6.5)

- Investigation on how contaminant concentration in air related to concentration on a wafer surface after prolonged exposure.
- Collect data for $i = 1, 2, \ldots, n = 6$ experiments.
- Set X_i: DBP concentration in air.
- Observe Y_i: DBP concentration on wafer surface after 4 hours.



Example (textbook Example 6.5)

We assume

$$Y_i = \beta X_i + \epsilon_i,$$

with ϵ_i be the random error term with $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2$.
Consider the following three estimators:

- Estimator 1:

$$\hat{\beta} = \frac{1}{n} \sum_i \frac{Y_i}{X_i}.$$

- Estimator 2:

$$\hat{\beta} = \frac{\sum_i Y_i}{\sum_i X_i}.$$

- Estimator 3:

$$\hat{\beta} = \frac{\sum_i X_i Y_i}{\sum_i X_i^2}.$$

All three estimators are unbiased.

Principles in Choosing Estimators

Principle of unbiased Estimation:
When choosing among several different estimators of u, select one that is unbiased.

Principle of Minimum Variance Unbiased Estimation:
Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting θ is called the minimum variance unbiased estimator (MVUE) of θ.

Example (textbook Example 6.5) Cont.

The variances for the three estimators are

- Estimator 1:

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{n^2} \sum_i \frac{1}{X_i^2}.$$

- Estimator 2:

$$\mathrm{Var}(\hat{\beta}) = \frac{n\sigma^2}{\left(\sum_i X_i\right)^2}.$$

- Estimator 3:

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_i X_i^2}.$$

The third estimator has the smallest variance among the three.