Estimation Theory

Point Estimation

Benchikh Tawfik

Faculty of Medicine
1st Year Medicine

February 08, 2026

PLAN DE COURS

ESTIMATION: DEFINITION

- An important aspect of inferential statistics is to obtain **reliable estimates** of the characteristics of a population based on a sample drawn from that population.
- Estimating a parameter means providing an **approximate value** of this parameter using the information obtained from a random sample taken from the population.

Parameters to Be Estimated

- Estimation is a decision problem concerning parameters such as:
  - the mathematical expectation, denoted by m or μ (for numerical variables),
  - the variance σ^2 or the standard deviation σ (for numerical variables),
  - the proportion p (for count or binary variables).
- Since a sample provides only partial information about the population, the resulting estimates are inevitably subject to error, which should be minimized as much as possible.

Example

- A typical problem in estimation theory is the estimation of parameters of the distribution followed by a random variable X, when the **form of the distribution is known**.
- For example, the number of accidents occurring in a workshop during one week may reasonably be modeled by a Poisson distribution.
- This distribution depends on a parameter λ, whose value is unknown.
- To estimate this parameter, we record the number of accidents observed over n weeks, forming a random sample of size n.
- The sample mean of the number of accidents provides a **point estimate** of the parameter λ (a property of the Poisson distribution).

Formulation of the Problem

- In general terms, the estimation problem can be formulated as follows:
  - Let X be a characteristic of a random phenomenon. The variable X is therefore a random variable whose characteristics (mean, variance, etc.) are unknown.
  - We observe the realizations of a random sample drawn from the population of interest.
  - These observations are used to infer values or estimates of the parameters of the probability distribution governing the random variable X.

Formulation of the Problem

- These estimates can take two forms:
  - a single value: **point estimation**, representing the most plausible value of the parameter,
  - a range of values: **interval estimation** (confidence interval). A confidence interval is constructed so as to have a high probability of containing the true value of the parameter; it is always associated with a risk of error α.

DEFINITION OF A POINT ESTIMATOR

- The objective of point estimation theory is to select, among all possible statistics, the **best estimator**, that is, the one that provides a point estimate as close as possible to the parameter θ, regardless of the particular sample observed.

Definition

- An **estimator** is a statistic possessing well-defined theoretical properties.
- A sequence of statistics T_n, with $T_n = \varphi(\underline{X}_n)$, is called an estimator of the parameter θ if T_n converges to θ as $n \to \infty$, where the convergence may be in probability, almost surely, or in mean square (depending on the chosen criterion).
- Once the sample is observed, the estimator takes a numerical value, called the **point estimate** of the parameter θ.
- An estimator of θ is usually denoted by $\hat{\theta}(\underline{X})$, $\hat{\theta}_n$, or simply $\hat{\theta}$.

Estimator of the Mean $\mathbb{E}(X) = m$

- Consider the statistic $\overline{X}$ defined by

$$\overline{X} = T(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

  that is, the arithmetic mean of the n observations in the sample.
- This statistic can reasonably be considered an estimator of the population mean $\mathbb{E}(X) = m$, according to the **Weak Law of Large Numbers**.

- Indeed:
  - the statistic $\overline{X}$ takes into account all the observations,
  - it satisfies the following properties:

$$\mathbb{E}(T) = \mathbb{E}(\overline{X}) = \mathbb{E}(X) = m, \qquad \mathrm{Var}(T) = \frac{\mathrm{Var}(X)}{n},$$

    - if the distribution of the random variable X is known, the distribution of $\overline{X}$ can be derived.
- It will be shown later that the first two properties above (unbiasedness and decreasing variance) are, in fact, **desirable properties** for an estimator.

UNBIASED ESTIMATOR: DEFINITION

- An estimator T is **unbiased** if $\mathbb{E}(T) = \theta$.
- An estimator is **biased** if $\mathbb{E}(T) \neq \theta$.
- An estimator is **asymptotically unbiased** if $\mathbb{E}(T) \to \theta$ as the sample size $n \to \infty$.

CONSISTENT ESTIMATOR (GOOD ESTIMATOR)

- An estimator is **consistent** (in probability) if its distribution becomes increasingly concentrated around the true value of the parameter as the sample size increases.
- An unbiased estimator whose variance tends to zero as $n \to \infty$ is therefore consistent.
- Consistency is not necessarily unique; different estimators may satisfy these conditions.

REMARK ON UNBIASED ESTIMATORS

- A biased estimator produces estimates that systematically deviate from the true parameter value; therefore, it is generally less desirable than an unbiased estimator.
- However, the absence of bias alone does not guarantee a "good" estimator. One must also consider the estimator's variance.

Unbiased Estimator, Consistent Estimator: Example 1

- **Comparison of estimators of the mean:**
  - The usual sample mean

$$\overline{X} = T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \mathbb{E}(T_1) = m, \quad \mathrm{Var}(T_1) = \frac{\sigma^2}{n}$$

    is unbiased, and since its variance tends to 0 as $n \to \infty$, it is consistent.
  - Another estimator T_2 can be defined as the mean of odd-ranked observations (assuming n is even):

$$T_2 = \frac{2}{n}\sum_{i=1}^{p-1} X_{2i+1}, \quad n = 2p, \quad \mathbb{E}(T_2) = m, \quad \mathrm{Var}(T_2) = \frac{2\sigma^2}{n}.$$

    T_2 is also unbiased and consistent.

Consistent Estimator: Example 1 (Conclusion)

- Both estimators T_1 and T_2 are unbiased and consistent.
- However, T_1 is "better" because:
  - it uses all the information in the sample,
  - it has smaller variance: $\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$.

Unbiased Estimator: Example 2

- **Estimator of the variance:**
  - The statistic $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a **biased** estimator of the variance:

$$\mathbb{E}(S^2) = \frac{n-1}{n}\sigma^2 = \left(1 - \frac{1}{n}\right)\sigma^2,$$

    so the bias is σ^2/n.
  - In contrast, the corrected statistic

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

    is an **unbiased estimator** of the variance:

$$\mathbb{E}(S^{*2}) = \sigma^2.$$

Unbiased Estimator: Example 3

- **Comparison of variance estimators:**
  - Suppose the population mean m is known, and we want the best estimator of the variance.
  - Compare $T = \frac{1}{n}\sum(X_i - m)^2$ and S^{*2}:

$$\mathrm{Var}(T) = \frac{1}{n}(\mu_4 - \sigma^4), \quad \mathrm{Var}(S^{*2}) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right),$$

    where μ_4 is the fourth central moment.
  - Hence, $\mathrm{Var}(T) < \mathrm{Var}(S^{*2})$, so T is a "better" estimator of variance when m is known.

Practical Example 4

1. Suppose we measured the heart rates of 100 randomly selected normal individuals: 64, 80, 72, 88, ..., 78, 88, 78, 88, 88, 72, 60.
2. We want to estimate the population mean and variance of heart rate.
3. Then:
   - Sample mean (point estimate of population mean): $\overline{X} = \frac{1}{100}\sum x_i = 77.82$ (unbiased and consistent).
   - Sample variance: $S^2 = 90.51$, corrected population variance estimate: $S'^2 = \frac{n}{n-1}S^2 = 99.56$ (unbiased and consistent).
   - Population standard deviation estimate: $S' = \sqrt{99.56} = 9.98$.

Practical Example 5: Estimating BMI Mean

- The Body Mass Index (BMI) is calculated as:

$$\text{BMI} = \frac{\text{weight (kg)}}{\text{height (m)}^2},$$

  and is used to assess overweight status.
- Suppose the BMI of men has a population standard deviation $\sigma = 3kg/m^2$, and we want to estimate the population mean μ using a sample of size $n = 10$.
- The observed BMI values in the sample are:

$$28, 31, 26, 29, 27, 32, 25, 30, 28, 31$$

- The sample mean

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 28.7$$

  provides a **point estimate** of the population mean μ. This estimator is **unbiased** and **consistent**.

Practical Example 5: Estimating BMI Mean

- Suppose that σ^2 is unknown; provide an estimator for σ^2.
- An unbiased estimator of σ^2 is the **sample variance**:

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}S^2 = 5.23$$

Example: Estimating the Parameter p of a Binomial Distribution

- Suppose we study a medical characteristic X in a population P of size N, e.g., the proportion p of individuals with a disease or allergy.
- To know p exactly, we would need to survey all N individuals, which is usually impossible. Instead, we survey a random sample of size n, obtaining values x_i:
  - $x_i = 1$ if individual i has the characteristic X,
  - $x_i = 0$ otherwise.

Example: Binomial Parameter Estimation

- Let k_n be the number of "successes" in the sample.
- Let $f_n = \frac{k_n}{n} = \frac{\sum X_i}{n} = \overline{X}_n$, the sample proportion.
- By the law of large numbers, $f_n \to p$ as $n \to \infty$:

$$\mathbb{E}(k_n) = np \Rightarrow \mathbb{E}(f_n) = p,$$

$$\mathrm{Var}(k_n) = np(1-p) \Rightarrow \mathrm{Var}(f_n) = \frac{p(1-p)}{n}.$$

- The sample proportion f_n is an unbiased and consistent estimator of p. By choosing a suitable n, we can reduce the approximation error.

ESTIMATOR PRECISION

- The precision of an estimator is measured by its **mean squared error (MSE)**:

$$\mathbb{E}\left[(T-\theta)^2\right].$$

- The most desirable property for an estimator is a low MSE, which does not necessarily require it to be unbiased.

Estimator Precision

- By writing $T - \mathbb{E}(T) + \mathbb{E}(T) - \theta$, we can decompose the MSE as:

$$\mathbb{E}\left[(T-\theta)^2\right] = \mathsf{Var}(T) + \left[\mathbb{E}(T) - \theta\right]^2.$$

- To minimize the mean squared error, we should choose:
  - an **unbiased estimator** so that $\mathbb{E}(T) = \theta$,
  - an estimator with a **small variance** $\mathsf{Var}(T)$.

Estimator Precision: Summary

- Among unbiased estimators, the one with the smallest variance is preferred; this property is called the **efficiency** of the estimator.
- For example, if the population mean m is known, the best estimator for the variance σ^2 is the statistic

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

  rather than S^{*2}.
- **Remark:** It is possible to find biased estimators that are more precise than the best unbiased estimator.
- **Note:** Calculating the variance of an estimator, and thus identifying the minimum variance estimator, generally requires knowledge of the joint probability distribution of the random sample.

Summary: Convergent and Unbiased Estimators

- For a mean $\mathbb{E}(X)$: the sample mean

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

  - is an unbiased estimator: $\mathbb{E}(\overline{X}) = m$,
  - has an asymptotically vanishing MSE: $\mathbb{E}\left[(\overline{X} - m)^2\right] = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$,
  - is a convergent estimator: $\overline{X} \xrightarrow{\Pr} m$, by the law of large numbers.

- For a variance: the statistic

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

  is an unbiased and convergent estimator.

- For a proportion p: the sample frequency

$$f_n = \frac{k_n}{n} \quad (k_n = \text{number of successes in } n \text{ trials})$$

  is an unbiased and convergent estimator, with asymptotically vanishing MSE.

Summary: Convergent and Unbiased Estimators

- **Remark:** $S^* = \sqrt{S^{*2}}$ is *not* an unbiased estimator of the population standard deviation σ, because

$$\mathbb{E}(\sqrt{S^{*2}}) \neq \sqrt{\mathbb{E}(S^{*2})}.$$

METHOD OF MOMENTS: DEFINITION

- Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n drawn from a population with probability distribution $f(x;\theta)$, where θ is a vector of p unknown parameters to be estimated.
- Define the **empirical moments** and **theoretical moments** of order r as:

$$\overline{X^r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r \quad \text{and} \quad m_r(\theta) = \mathbb{E}_\theta[X^r],$$

respectively.

- Here, $\overline{X^1}$ is the sample mean and $m_1(\theta)$ is the population mean. Higher-order moments (variance, skewness, etc.) can also be used.
- The theoretical moments $m_r(\theta)$ depend on the unknown parameter(s) θ.

Method of Moments: Estimator Definition

- The **method of moments estimator** Θ_n is defined as the solution to the system of equations:

$$\overline{X^r} = m_r(\theta), \quad r = 1, \ldots, p.$$

- If this system can be solved, the estimator is a function of the empirical moments:

$$\Theta_n = k(\overline{X^1}, \overline{X^2}, \ldots, \overline{X^p}).$$

- Once a sample $(x_1, \ldots, x_n)$ is observed, the numerical value of the estimator is:

$$\hat{\theta}_n = k(\overline{x^1}, \overline{x^2}, \ldots, \overline{x^p}).$$

Method of Moments: Remarks

1. The method of moments generally provides **consistent estimators**, i.e., $\Theta_n \xrightarrow{\Pr} \theta$ as $n \to \infty$.
2. However, for moderate sample sizes, the method of moments may produce **less precise estimators** compared to other methods, such as maximum likelihood.
3. The method is simple to implement and intuitive, especially when the moments of the theoretical distribution are easy to compute.
4. It can be applied to both univariate and multivariate parameter estimation problems.

EXERCISE 1

Consider the statistical sample:

$$(1, 0, 2, 1, 1, 0, 1, 0, 0)$$

1. Assuming these data are realizations of a random variable with an unknown distribution, provide an **unbiased estimator** for the mean and variance of this distribution.
   Solution:

$$\overline{X} = \frac{1+0+2+1+1+0+1+0+0}{9} = \frac{6}{9} = 0.6667$$

$$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{9}(X_i - \overline{X})^2 = \frac{1}{8}\left[(1-0.6667)^2 + (0-0.6667)^2 + \cdots\right.$$

$$S^{*2} = 0.5185$$

2. Suppose we model the sample using a binomial distribution $\mathcal{B}(2,p)$. Use the sample mean to propose a **point estimate** for p.
Solution: For a binomial $\mathcal{B}(2,p)$, $\mathbb{E}(X) = 2p$, so

$$\hat{p}_{\text{mean}} = \frac{\overline{X}}{2} = \frac{0.6667}{2} = 0.3333$$

3. Using the same binomial model, use the sample variance to propose another estimate of p.
Solution:
Set $S^{*2} = 2p(1-p) = 0.5185 \Rightarrow p^2 - p + 0.25925 = 0.$
Solving the quadratic equation:

$$p = \frac{1 \pm \sqrt{1 - 4 \cdot 0.25925}}{2} = \frac{1 \pm 0.439}{2}$$

$$\Rightarrow p \approx 0.2805 \quad \text{or} \quad p \approx 0.7195$$

We choose $p < 0.5$ as it is consistent with the mean estimate: $p \approx 0.2805$.

4. Suppose we model the sample using a Poisson distribution $\mathcal{P}(\lambda)$. What is a suitable **point estimate** for λ?
   Solution: For a Poisson distribution, $\mathbb{E}(X) = \lambda$, so the natural estimator is

$$\hat{\lambda} = \overline{X} = 0.6667$$