UNIVERSITY OF TWENTE.

Chapter 2: Estimation

Statistical Techniques for CS/BIT

18 November 2021

Overview

Statistical inference

Estimates and estimators

Comparing estimators

Statistical inference

Let $\{x_1, x_2, \ldots, x_n\}$ be a random sample of observations coming from some distribution X.

Population **Sample**

$$\boxed{X} \longrightarrow x_1, x_2, \ldots, x_n$$

Objective

Learn something about **the population** by studying **the sample**.

This is called **statistical inference**.

Unlike descriptive statistics, inference is <u>not</u> interested in the sample.

Fact 1

The average height of men in the Netherlands is 180.8 cm.

Source: `https://en.wikipedia.org/wiki/Average_human_height_by_country`

How do we know?

- We obtained a sample from the population
- We observed the sample
- We made conclusions about the population

Fact 2

All cows in the Netherlands are black and white.

Source: Personal observation

How do I know?

- I obtained a sample from the population
- I observed the sample
- I made conclusions about the population

Two reasoning processes:

- Deductive reasoning: *general* $\longrightarrow$ *particular*
- Inductive reasoning: *particular* $\longrightarrow$ *general*

Fact 2 is an example of naïve inductive reasoning

It simply does not work!

But why do we accept **Fact 1** as true?

- We don't claim **absolute knowledge**
- We assume an underlying **probability model**
- We quantify our **uncertainty**

Warning

If the assumptions are not satisfied the conclusions are invalid!

Beware of accidentally doing naïve induction!!

Fact 1 (precise formulation)

With 95% confidence, the average height of men in the Netherlands is 180.8 ± 0.2 cm.

Source: `https://opendata.cbs.nl/statline/#/CBS/en/dataset/81175ENG/table`

Fact 2 (precise formulation)

At least five cows in the Netherlands are black and white.

Source: Personal observation, after a long reflection

Statistical inference

Estimates and estimators

Comparing estimators

Objective of statistical inference

Learn something about **the population** by studying **a random sample**.

Consider a property of a population X, measured by a single real number θ.

Suppose we have a random sample $x_1, \ldots, x_n$.

Goal:

Estimate the value of θ using some function $T(x_1, \ldots, x_n)$.

Definition

A function $T(x_1, \ldots, x_n)$ of a random sample $\{x_1, \ldots, x_n\}$ is called **a statistic**.

The statistic $T(x_1, \ldots, x_n)$ is **an estimate** for θ.

Notation

We write $T(x_1, \ldots, x_n) = \hat{\theta}$. This means that $\hat{\theta}$ is an estimate for θ.

A *good* choice for T should give $\hat{\theta} \approx \theta$. **(!)**

Before we have observed any data, we can consider $X_1, X_2, \ldots, X_n$ as random variables.

These must be **independent** and **identically distributed** (in short: iid).

Definition

An **estimator** is a function $T(X_1, X_2, \ldots, X_n)$ of iid random variables.

Note that $T(X_1, X_2, \ldots, X_n)$ is a **random variable** itself!

Estimates vs estimator

An **estimator** is a random process, namely, the function $T(X_1, \ldots, X_n)$.

An **estimate** is a real number, namely, the observed value $T(x_1, \ldots, x_n)$.

Estimates vs estimators

Warning

If T is an estimator for a parameter θ, it is common to write $\hat{\theta}$ for both estimate and estimator:

$$\hat{\theta} = T(X_1, \ldots, X_n) \quad \text{but also} \quad \hat{\theta} = T(x_1, \ldots, x_n)\ !$$

Let X be a random variable with mean $\mu = \mathrm{E}(X)$.

Assume μ is a **fixed** but **unknown** real number.

Definition

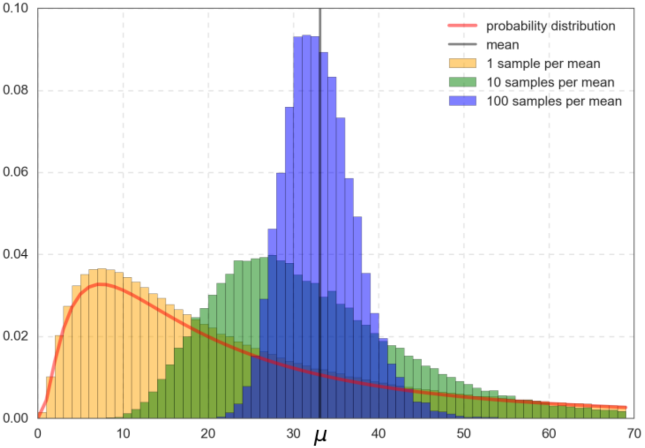
Let $X_1, \ldots, X_n$ be a random sample of X. The **sample mean** is

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

- The random variable $\overline{X}$ is an **estimator** for μ,
- Once data is observed, $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is an **estimate** for μ,

Why is $\overline{X}$ a good estimator?

The bigger the sample size n, the more *"concentrated"* the distribution of $\overline{X}$ around μ.



Why is $\overline{X}$ a good estimator?

Remark

- The **mean** of $\overline{X}$ is:

$$\mathrm{E}(\overline{X}) = \mathrm{E}(X) = \mu$$

- The **variance** of $\overline{X}$ is:

$$\mathrm{Var}(\overline{X}) = \frac{1}{n}\,\mathrm{Var}(X) = \frac{\sigma^2}{n}$$

General principle

The more data we collect, the better our estimates will be.

Overview

Statistical inference

Estimates and estimators

Comparing estimators

Bias

Let $T(X_1, \ldots, X_n)$ be an estimator for the population parameter θ.

Definition

- We say that T is **unbiased** if $\mathrm{E}(T) = \theta$.
- Otherwise we say that T is **biased**.
- The **bias** of T is the difference: $\mathrm{Bias}(T) = \mathrm{E}(T) - \theta$.

Let $T(X_1, \ldots, X_n)$ be an estimator for the population parameter θ.

Definition

The **standard error** of T is defined as

$$\mathrm{se}(T) = \sqrt{\mathrm{Var}(T)}.$$

Skewness		Kurtosis	
Statistic	Std. Error	Statistic	Std. Error
0.673	0.414	-0.017	0.809

Example: Consider $X_1, \ldots, X_{10}$ a random sample from $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown.

Suppose we have observed data $x_1, x_2, \ldots, x_{10}$, and we estimate μ and σ^2 to be

$$\overline{x} = 2.08, \quad s^2 = 1.44.$$

The standard error of $\overline{X}$ is

$$\mathrm{se}(\overline{X}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{10}},$$

and the estimated standard error of $\overline{X}$ is

$$\widehat{\mathrm{se}}(\overline{X}) = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{1.44}{10}} = 0.38.$$

Ideally we want our estimator T to satisfy:

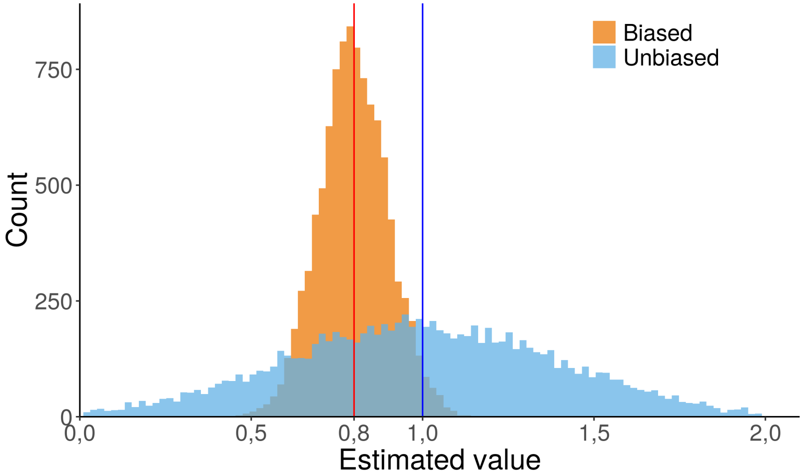
1. The bias $\mathrm{Bias}(T)$ is **zero or close to zero**,
2. The variance $\mathrm{Var}(T)$ is **as small as possible**.

Questions

Can we achieve both?

Which one is more important?

Estimating a parameter $\theta = 1$.



Mean-squared error

Let $T(X_1, \ldots, X_n)$ be an estimator for the population parameter θ.

The **error** of T is the random variable $T - \theta$.

The **squared error** of T is $(T - \theta)^2$, also a random variable.

Clearly, $(T - \theta)^2 \geq 0$.

Definition

The **mean squared error** of T is

$$\mathrm{MSE}(T) = \mathrm{E}[(T - \theta)^2] \in \mathbb{R}.$$

Properties

- $\mathrm{MSE}(T) \geq 0$,
- $\mathrm{MSE}(T) = 0$ if and only if $T \equiv \theta$,
- The MSE is a measure of the performance of T.

Note

We only want to work with estimators that satisfy

$$\mathrm{MSE}(T) \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty.$$

Theorem

Let $T(X_1, \ldots, X_n)$ *be an estimator for the population parameter* θ*. Then*

$$\mathrm{MSE}(T) = \mathrm{Bias}^2(T) + \mathrm{Var}(T).$$

Let T_1, T_2 be estimators for the same parameter θ.

Definition

We say that T_1 is **better** than T_2 if

$$\mathrm{MSE}(T_1) < \mathrm{MSE}(T_2).$$

Example

Let X_1, X_2 be a random sample from $N(\mu, \sigma^2)$.

Suppose we don't like $\overline{X}$, and we decide to estimate μ by another linear combination

$$T(X_1, X_2) = \frac{1}{3}X_1 + \frac{2}{3}X_2.$$

We can guess that $\overline{X}$ is a better estimator than T.

Poll: Why is $\overline{X}$ a better estimator?

1. Because T is biased and $\overline{X}$ is not.
2. Because T is unbiased but has a greater variance than $\overline{X}$.
3. Because T has greater variance and greater bias than $\overline{X}$.

Example

$$\begin{aligned}
\mathrm{E}(T) &= \mathrm{E}\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right)\\
&= \mathrm{E}\left(\frac{1}{3}X_1\right) + \mathrm{E}\left(\frac{2}{3}X_2\right)\\
&= \frac{1}{3}\,\mathrm{E}(X_1) + \frac{2}{3}\,\mathrm{E}(X_2)\\
&= \frac{1}{3}\mu + \frac{2}{3}\mu\\
&= \mu
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(T) &= \mathrm{Var}\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right)\\
&\overset{\text{indep.}}{=} \mathrm{Var}\left(\frac{1}{3}X_1\right) + \mathrm{Var}\left(\frac{2}{3}X_2\right)\\
&= \frac{1}{9}\,\mathrm{Var}(X_1) + \frac{4}{9}\,\mathrm{Var}(X_2)\\
&= \frac{1}{9}\sigma^2 + \frac{4}{9}\sigma^2\\
&= \frac{5}{9}\sigma^2
\end{aligned}$$

Conclusion

$$\mathrm{Bias}(T) = 0 \qquad \mathrm{Var}(T) = \frac{5}{9}\sigma^2$$