Point Estimates

Jiří Neubauer

Department of Econometrics FVL UO Brno
office 69a, tel. 973 442029
email:Jiri.Neubauer@unob.cz

Introduction

Suppose that we manufacture lightbulbs and we want to state the average lifetime on the box. Let us say that we have following five observed lifetimes (in hours)

$$983 \quad 1063 \quad 1241 \quad 1040 \quad 1103$$

which have the average 1086. If it is all the information we have, it seems to be reasonable to state 1086 as the average lifetime.

Introduction

Let the random variable X be the lifetime of a lightbulb, and let $E(X) = \mu$. Here μ is an **unknown parameter**. We decide to repeat the experiment to measure a lifetime 5 times and will then get an outcome on the five random variables $X_1, \ldots, X_5$ that are i.i.d. (independent identically distributed). We now estimate μ by

$$\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$$

which is the sample mean.

Point Estimator

Definition

Let $X_1, \ldots, X_n$ be a random sample. The statistic (random variable)

$$T = T(X_1, X_2, \ldots, X_n) = T(\mathbf{X}),$$

which is a function of the random sample and is used to estimate an unknown parameter θ, is called a point estimator of θ. We write $T(\mathbf{X}) = \hat{\theta}$.

Unbiased Estimator

Definition

The estimator $T(\mathbf{X})$ is said to be **unbiased** estimator the parameter θ if

$$E\left[T(\mathbf{X})\right] = \theta.$$

Unbiased Estimator

Definition

The estimator $T(\mathbf{X})$ is said to be **unbiased** estimator the parameter θ if

$$E\left[T(\mathbf{X})\right] = \theta.$$

The difference

$$B(\theta, T) = E\left[T(\mathbf{X})\right] - \theta$$

is called **a bias** of the estimator $T(\mathbf{X})$.

Example

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

- The sample mean $\overline{X}$ is an unbiased estimator of μ, because

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu.$$

Example

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

- The sample mean $\overline{X}$ is an unbiased estimator of μ, because

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu.$$

- The sample variance S^2 is an unbiased estimator of σ^2, because

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \cdots = \sigma^2.$$

Example

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

- The (moment) variance S_n^2 is a biased estimator of σ^2, because

$$E(S_n^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \cdots = \frac{n-1}{n}\sigma^2.$$

Example

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

- The (moment) variance S_n^2 is a biased estimator of σ^2, because

$$E(S_n^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \cdots = \frac{n-1}{n}\sigma^2.$$

The bias of the estimator S_n^2 is

$$B(\sigma^2, S_n^2) = E(S_n^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{1}{n}\sigma^2.$$

The bias decreases for large n.

Asymptotically Unbiased Estimator

Some estimators are biased but their bias decrease when n increases.

Definition

If

$$\lim_{n\to\infty} E\left[T(\mathbf{X})\right] = \theta,$$

then the estimator $T(\mathbf{X})$ is said to be asymptotically unbiased estimator of the parameter θ.

It easy to see that

$$\lim_{n\to\infty} E\left[T(\mathbf{X}) - \theta\right] = 0.$$

Example

The (moment) variance is an asymptotically unbiased estimator of σ^2, because

$$\lim_{n\to\infty} E(S_n^2) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2.$$

Consistent Estimator

Definition

The statistic $T(\mathbf{X})$ is a **consistent estimator** of the parameter θ if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(|T(\mathbf{X}) - \theta| < \epsilon) = 1.$$

Consistent Estimator

Definition

The statistic $T(\mathbf{X})$ is a **consistent estimator** of the parameter θ if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(|T(\mathbf{X}) - \theta| < \epsilon) = 1.$$

If

$$\lim_{n\to\infty} B(\theta, T) = 0 \quad \text{and} \quad \lim_{n\to\infty} D[T(\mathbf{X})] = 0,$$

then $T(\mathbf{X})$ is the consistent estimator of θ.

Example

Prove that the sample mean is a consistent estimator of the expected value μ.

Example

Prove that the sample mean is a consistent estimator of the expected value μ.

According to $E(\overline{X}) = \mu$ and $D(\overline{X}) = \sigma^2/n$ we obtain

$$B(\mu, \overline{X}) = E(\overline{X}) - \mu = 0 \quad \text{a} \quad \lim_{n\to\infty} D(\overline{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

Efficiency of Estimators

If we have two unbiased estimators $T_1(\mathbf{X}) = \hat{\theta}$ and $T_2(\mathbf{X}) = \tilde{\theta}$, which should we choose? Intuitively, we should choose the one that tends to be closer to θ, and since $E(T_1) = E(T_2) = \theta$, it makes sense to choose the estimator with the smaller variance.

Definition

Suppose that $T_1(\mathbf{X}) = \hat{\theta}$ and $T_2(\mathbf{X}) = \tilde{\theta}$ are two unbiased estimators of θ. If

$$D(T_1(\mathbf{X})) < D(T_2(\mathbf{X}))$$

then $T_1(\mathbf{X}) = \hat{\theta}$ is said to be **more efficient** than $T_2(\mathbf{X}) = \tilde{\theta}$.

Example

We can find two unbiased estimators of a parameter λ of Poisson distribution

$$E(\overline{X}) = \lambda \quad \text{and} \quad E(S^2) = \lambda.$$

It is possible to calculate that

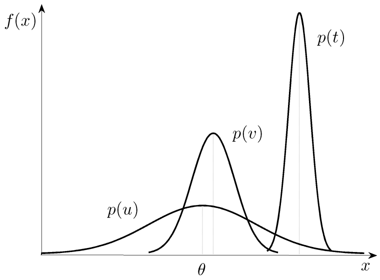
$$D(\overline{X}) < D(S^2).$$

The estimator $\overline{X}$ is more efficient then the estimator S^2.

How to Compare Estimators?

Let us suppose we would like to compare unbiased and biased estimators of the parameter θ. In this case might not be suitable to choose one of the smallest variance.

The estimator T has the smallest variance but has a large bias. Even the estimator with the smallest bias is not necessary the best one. The estimator U has no bias but its variance is to large. The estimator V seems to be the best.



Mean Square Error

Definition

The mean square error of the estimator T of a parameter θ is defined as

$$MSE(T) = E(T-\theta)^2 = D(T) + B^2(\theta, T)$$

$$(MSE \text{ of estimator} = \text{variance of estimator} + \text{bias}^2),$$

where $T - \theta$ is a sample error.

Mean Square Error

The mean square error

- indicates the "average" sample error of estimates which can be calculated for all possible random sample of the size n.

Mean Square Error

The mean square error

- indicates the "average" sample error of estimates which can be calculated for all possible random sample of the size n.
- is a combination of 2 required properties (a small bias and a small variance), that why it is an universal criterion.

Mean Square Error

The mean square error

- indicates the "average" sample error of estimates which can be calculated for all possible random sample of the size n.
- is a combination of 2 required properties (a small bias and a small variance), that why it is an universal criterion.

If T is an unbiased estimator then $MSE(T) = D(T)$.
Another possibility how to measure an accuracy of estimators is **standard error**

$$SE = \sqrt{D(T)}.$$

Example

The sample mean is an unbiased estimator of the expected value μ, the standard error is equal to the standard deviation of the sample mean

$$SE = \sqrt{D(\overline{X})} = \sigma(\overline{X}) = \frac{\sigma(X)}{\sqrt{n}}.$$

$\sigma(X)$ is unknown, we have to estimate it by the sample standard deviation and we get the estimation

$$\widehat{SE} = \frac{\hat{\sigma}(X)}{\sqrt{n}} = \frac{S}{\sqrt{n}}.$$

Example

Find the mean square error of S^2 and S_n^2. Let us start with the statistic S^2 which is an unbiased estimator of σ^2.

$$MSE(S^2) = D(S^2) = E(S^2 - \sigma^2)^2 = E(S^4) - 2\sigma^2 E(\sigma^2) + \sigma^4 =$$
$$= E(S^4) - \sigma^4 = \frac{2\sigma^4}{n-1}.$$

Example

Find the mean square error of S^2 and S_n^2. Let us start with the statistic S^2 which is an unbiased estimator of σ^2.

$$MSE(S^2) = D(S^2) = E(S^2-\sigma^2)^2 = E(S^4) - 2\sigma^2 E(\sigma^2) + \sigma^4 = \\ = E(S^4) - \sigma^4 = \tfrac{2\sigma^4}{n-1}.$$

The MSE of the estimator S_n^2 is

$$MSE(S_n^2) = E(S_n^2-\sigma^2)^2 = E(S_n^4) - 2\tfrac{n-1}{n}\sigma^4 + \sigma^4 = \\ = E(S_n^4) - \tfrac{2-n}{n}\sigma^4 = \tfrac{2n-1}{n^2}\sigma^4,$$

$$MSE(S_n^2) < MSE(S^2) \quad \text{because} \quad \frac{2n-1}{n^2} < \frac{2}{n-1}$$

.

Methods of Point Estimations

The definitions of unbiasness and other properties of estimators do not provide any guidance about how good estimators can be obtained. In this part, we discuss two methods for obtaining point estimators:

- the method of moments,
- the method of maximum likelihood.

Maximum likelihood estimates are generally preferable to moment estimators because they have better efficiency properties. However, moment estimators are sometimes easier to compute. Both methods can produce unbiased point estimators.

Method of Moments

The general idea behind the method of moments is to equate population moments, which are defined in terms of expected values, to the corresponding sample moments. The population moments will be functions of the unknown parameters. Then these equations are solved to yield estimators of the unknown parameters.

Method of Moments

Let us assume the distribution with $m \geq 1$ real parameters $\theta_1, \theta_2, \ldots, \theta_m$ and let $X_1, X_2, \ldots, X_n$ be a random sample from this distribution. Let us suppose that exist moments

$$\mu_r' = E(X_i^r) \quad \text{for} \quad r = 1, 2, \ldots, m.$$

These moments depend on the parameters $\theta_1, \theta_2, \ldots, \theta_m$. Sample moments are defined by the formula

$$M_r' = \frac{1}{n} \sum_{i=1}^{n} X_i^r, \quad r = 1, 2 \ldots.$$

Method of Moments

Let $X_1, \ldots, X_n$ be a random sample from either a probability function or probability density function with m unknown parameters $\theta_1, \ldots, \theta_m$. The moment estimators are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters

$$\mu_r' = M_r'.$$

Example

Estimation of the parameter λ – Poisson distribution.

Suppose that $X_1, \ldots, X_n$ is a random sample from the Poisson distribution $Po(\lambda)$, we get an equation

$$\mu_1' = M_1' \quad \Rightarrow \quad E(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

the estimator $\hat{\lambda}$ of the parameter λ is

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

Example

Estimation of the parameters μ and σ^2 – normal distribution.
Suppose that $X_1, \ldots, X_n$ is a random sample from the normal distribution $N(\mu, \sigma^2)$.

$$\mu_1' = M_1' \quad \Rightarrow \quad E(X_i) = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

Example

Estimation of the parameters μ and σ^2 – normal distribution.
Suppose that $X_1, \ldots, X_n$ is a random sample from the normal distribution $N(\mu, \sigma^2)$.

$$\mu_1' = M_1' \quad \Rightarrow \quad E(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\mu_2' = M_2' \quad \Rightarrow \quad E(X_i^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \Leftrightarrow D(X_i) + E(X_i)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

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

Example

Estimation of the parameters μ and σ^2 – normal distribution.
Suppose that $X_1, \dots, X_n$ is a random sample from the normal distribution $N(\mu, \sigma^2)$.

$$\mu_1' = M_1' \quad \Rightarrow \quad E(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\mu_2' = M_2' \quad \Rightarrow \quad E(X_i^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \Leftrightarrow D(X_i) + E(X_i)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

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

We obtain estimators

$$\hat{\mu} = \overline{X}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S_n^2 = \frac{n-1}{n} S^2$$

.

Method of Maximum Likelihood

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability density function $f(x, \theta)$ or a probability function $p(x, \boldsymbol{\theta})$ with an unknown parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$. A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has either a joint probability density function or probability function

$$g(\mathbf{x}, \boldsymbol{\theta}) = g(x_1, x_2, \ldots, x_n, \boldsymbol{\theta}) = f(x_1, \boldsymbol{\theta}) f(x_2, \boldsymbol{\theta}) \cdots f(x_n, \boldsymbol{\theta})$$

or

$$g(\mathbf{x}, \boldsymbol{\theta}) = g(x_1, x_2, \ldots, x_n, \boldsymbol{\theta}) = p(x_1, \boldsymbol{\theta}) p(x_2, \boldsymbol{\theta}) \cdots p(x_n, \boldsymbol{\theta}).$$

Method of Maximum Likelihood

The density $g(\mathbf{x}, \boldsymbol{\theta})$ is a function of $\mathbf{x}$ with a given value of $\boldsymbol{\theta}$. If values $\mathbf{x}$ are given (observed data) than $g(\mathbf{x}, \boldsymbol{\theta})$ is a function of a variable $\boldsymbol{\theta}$. We denote it $\mathcal{L}(\boldsymbol{\theta}, \mathbf{x})$ and call it a **likelihood function**.
If exists some $\hat{\boldsymbol{\theta}}$ which fulfils

$$\mathcal{L}(\hat{\boldsymbol{\theta}}, \mathbf{x}) \geq \mathcal{L}(\boldsymbol{\theta}, \mathbf{x}),$$

then $\hat{\boldsymbol{\theta}}$ is a **maximum likelihood estimator** of the parameter $\boldsymbol{\theta}$.

Method of Maximum Likelihood

The density $g(\mathbf{x}, \boldsymbol{\theta})$ is a function of $\mathbf{x}$ with a given value of $\boldsymbol{\theta}$. If values $\mathbf{x}$ are given (observed data) than $g(\mathbf{x}, \boldsymbol{\theta})$ is a function of a variable $\boldsymbol{\theta}$. We denote it $\mathcal{L}(\boldsymbol{\theta}, \mathbf{x})$ and call it a **likelihood function**.
If exists some $\hat{\boldsymbol{\theta}}$ which fulfils

$$\mathcal{L}(\hat{\boldsymbol{\theta}}, \mathbf{x}) \geq \mathcal{L}(\boldsymbol{\theta}, \mathbf{x}),$$

then $\hat{\boldsymbol{\theta}}$ is a **maximum likelihood estimator** of the parameter $\boldsymbol{\theta}$.
Sometimes is reasonable to use a logarithm of the likelihood function $L(\boldsymbol{\theta}, \mathbf{x}) = \ln \mathcal{L}(\boldsymbol{\theta}, \mathbf{x})$. For the maximum likelihood estimator we can write

$$L(\hat{\boldsymbol{\theta}}, \mathbf{x}) \geq L(\boldsymbol{\theta}, \mathbf{x}),$$

because the logarithm is an increasing function.

Method of Maximum Likelihood

The Maximum likelihood estimator of the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ we obtain by solving a system of equations

$$\frac{\partial L(\boldsymbol{\theta}, \mathbf{x})}{\partial \theta_i} = 0, \quad i = 1, 2, \ldots, m.$$

Example

Let X be a Bernoulli random variable. The probability function is

$$p(x) = \begin{cases} \pi^x(1-\pi)^{1-x} & x = 0, 1, \\ 0 & \text{otherwise.} \end{cases}$$

The likelihood function is

$$\mathcal{L}(\pi, \mathbf{x}) = \pi^{x_1}(1-\pi)^{1-x_1}\pi^{x_2}(1-\pi)^{1-x_2}\dots\pi^{x_n}(1-\pi)^{1-x_n} = \\ = \pi^{\sum_{i=1}^n x_i}(1-\pi)^{n-\sum_{i=1}^n x_i}$$

The logarithm of $\mathcal{L}(\pi, \mathbf{x})$ is

$$L(\pi, \mathbf{x}) = \sum_{i=1}^n x_i \ln \pi + \left(n - \sum_{i=1}^n x_i\right) \ln(1-\pi).$$

Example

We calculate the maximum of $L(\pi, \mathbf{x})$

$$\frac{dL(\pi, \mathbf{x})}{d\pi} = \frac{\sum_{i=1}^{n} x_i}{\pi} - \frac{n - \sum_{i=1}^{n} x_i}{1 - \pi} = 0,$$

and get the estimator

$$\hat{\pi} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}.$$

Example

Find a maximum likelihood estimator of a parameter λ of Poisson distribution $Po(\lambda)$.

$$\mathcal{L}(\lambda, \mathbf{x}) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \cdots x_n!},$$

$$L(\lambda, \mathbf{x}) = \ln \mathcal{L}(\lambda, \mathbf{x}) = -n\lambda + \sum_{i=1}^{n} x_i \ln \lambda - \ln(x_1! x_2! \cdots x_n!)$$

$$\frac{dL(\lambda, \mathbf{x})}{d\lambda} = -n + \sum_{i=1}^{n} x_i \cdot \frac{1}{\lambda} = 0$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}.$$