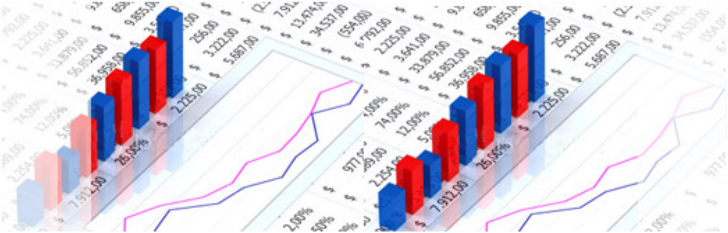
# MTMS.01.099 Mathematical Statistics

## Lecture 6. Methods for estimating parameters

Tõnu Kollo



Fall 2017

# Properties of estimators (reminder)

$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X})$ is an **unbiased** estimator of $\theta$ if $E(\hat{\boldsymbol{\theta}}) = \theta$.

$\hat{\boldsymbol{\theta}}$ is an **asymptotically unbiased** estimator of $\theta$ if $\lim_{n\to\infty} E(\hat{\boldsymbol{\theta}}) = \theta$.

If two estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased and

$$Var[\hat{\theta}_1(\boldsymbol{X})] \leq Var[\hat{\theta}_2(\boldsymbol{X})],$$

then $\hat{\theta}_1$ is said to be **more efficient** than $\hat{\theta}_2$.

The ratio $\frac{Var(\hat{\boldsymbol{\theta}}_1)}{Var(\hat{\boldsymbol{\theta}}_2)}$ denotes the **relative efficiency** of $\hat{\boldsymbol{\theta}}_1$ to $\hat{\boldsymbol{\theta}}_2$.

$\hat{\theta}$ is a **consistent** iff

$$lim_{n\to\infty} E(\hat{\boldsymbol{\theta}}) = \theta \quad (1)$$

$$lim_{n\to\infty} Var(\hat{\boldsymbol{\theta}}) = 0 \quad (2)$$

# Properties of estimators (2)

*Definition*. The *Mean Square Error* (**MSE**) of an estimator $\hat{\theta}$ for estimating $\theta$ is

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

**Proposition 1**

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + B^2$$

*Proof.*

$$\begin{aligned}
MSE(\hat{\theta}) &= E\left[(\hat{\theta} - \theta)^2\right] = E\left[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2\right] \\
&= E\left[((\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta))^2\right] \\
&= E\left[(\hat{\theta} - E(\hat{\theta}))^2\right] + 2E\left[\hat{\theta} - E(\hat{\theta})\right](E(\hat{\theta}) - \theta) + \\
&+ (E(\hat{\theta}) - \theta)^2 \\
&= Var(\hat{\theta}) + B^2.
\end{aligned}$$

# Reminder. Remark 1

We use the term *estimator* for a function of theoretical sample $\hat{\theta}(\mathbf{X})$. The result $\hat{\theta}(\mathbf{X} = \mathbf{x}) = \hat{\theta}$ where $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is an *estimate*.

Asymptotic properties are important and interesting, but we always have finite $n$ in practice. There may be no reason to prefer one estimator to the other asymptotically, but there may be large differences for small $n$. Often, simulations are needed to compare the small sample performance of estimators.

The function of interest $\theta$ may be some function of the usual parameters of a distribution. For example, $\theta$ could be $p/(1-p)$ (the odds ratio) in a binomial distribution or the mean $1/\lambda$ in an exponential distribution.

# Remark 2

The efficiency of an estimator refers to how much information it contains about the parameter from the sample. A more efficient estimator contains more information, in some sense, from a sample of a given size. Efficiency measures information determined by the variance of an unbiased estimator – smaller variance means greater efficiency.

Efficiency and relative efficiency are useful concepts only for unbiased estimators, in which case the MSE's of the estimators are equal to their variances. Since many estimators of interest are not unbiased, these concepts can sometimes be of limited usefulness. Other measures of accuracy may be more useful.

# Estimation of population mean

Consider a random variable $X$ with distribution $F$ with

$$E(X) = \mu \text{ and } Var(X) = \sigma^2.$$

Distribution $F$ is not known $\Rightarrow$ mean and variance are unknown.

**Theorem 1**

Let $x_1, x_2, \ldots, x_n$ be a sample from distribution $F$, then the sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

is an unbiased and consistent estimate of $\mu$.

# Estimation of population variance

Consider a random variable $X$ with distribution $F$ with

$$E(X) = \mu \text{ and } Var(X) = \sigma^2,$$

where $\mu$ and $\sigma^2$ are unknown. We want to estimate $\sigma^2$ from the sample.

**Theorem 2**

Let $x_1, x_2, \ldots, x_n$ be a sample from distribution $F$, then the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

is an unbiased estimate of $\sigma^2$.

Proof can be found in book (pg. 197).

# Estimation of population variance (2)

**Remark 1**

$s^2$ is also a consistent estimate of $\sigma^2$, e.g. $Var(s^2) \to 0$ if $n \to \infty$.

The standard deviation $\sqrt{Var(X)} = \sigma$ could be estimated by the sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

The sample standard deviation $s$ is not an unbiased estimate of the standard deviation $\sigma$. But $s$ is a consistent estimate of $\sigma$.

# Estimation of population variance (2)

**Remark 2**

If the population mean $\mu$ is known, we can use is to estimate the variance. It can be seen that $s_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ is an unbiased estimate of $\sigma^2$.

The following estimate is biased estimate of $\sigma^2$:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

$\hat{\sigma}^2 = \frac{n-1}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{n-1}{n}s^2 \Rightarrow$

$E(\frac{n-1}{n}\boldsymbol{s}^2) = \frac{n-1}{n}E(\boldsymbol{s}^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$

# Methods for estimating parameters

We have considered properties that we would want an estimator to have. Now, we will consider some general procedures for estimating parameters.

**Definition**

Let $x_1, \ldots, x_n$ be a sample from a distribution $F(x;\theta)$, which can be discrete or continuous. The function

$$L(\mathbf{x}, \theta) = \begin{cases} f(x_1;\theta) \times f(x_2;\theta) \times \ldots \times f(x_n;\theta) & \text{(continuous case)} \\ p(x_1;\theta) \times p(x_2;\theta) \times \ldots \times p(x_n;\theta) & \text{(discrete case).} \end{cases}$$

is called the **likelihood function**, where $f(x;\theta)$ is a density function (in continuous case), $p(x;\theta)$ is a probability function (in discrete case) and $\theta$ is an unknown parameter with parameter space $A$, $\theta \in A$.

# The Method of Maximum Likelihood

$$L(\mathbf{x}, \theta) = \begin{cases} f(x_1;\theta) \times f(x_2;\theta) \times \ldots \times f(x_n;\theta), & \text{continuous case} \\ p(x_1;\theta) \times p(x_2;\theta) \times \ldots \times p(x_n;\theta), & \text{discrete case.} \end{cases}$$

In the discrete case, $L(\mathbf{x}, \theta)$ is just the probability of obtaining the particular sample $\mathbf{x} = (\mathbf{x_1}, \ldots, \mathbf{x_n})$.
In the continuous case, $L(\mathbf{x}, \theta)$ is the value assumed by the $n$-dimensional density function of $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ at $x_1, \ldots, x_n$.
The goal is to find such $\theta$ that $L(\mathbf{x}, \theta)$ would get the largest value as the function of $\theta$.

# The Method of Maximum Likelihood (2)

**Definition**

The value $\hat{\theta}$ from sample space $A$ for which $L(\mathbf{x}, \theta)$ obtains its largest value within $A$ is called the **ML estimate** of $\theta$:

$$\hat{\theta}_{MLE} = max_{\theta \in A} L(\mathbf{x}, \theta).$$

A maximum likelihood estimate $\hat{\theta}_{MLE}$ is a value of $\theta$ that maximizes the likelihood function at the given sample **x.**

# The Method of Maximum Likelihood (3)

In many cases it is simpler to find maximum of the logarithmed function than of the function itself. Since the logarithm is an increasing function, if $\hat{\theta}$ is a maximum of $ln(L(\boldsymbol{x}, \theta))$, then $\hat{\theta}$ is the maximum of $L(\boldsymbol{x}, \theta)$ also.

**Definition**

The logarithmic likelihood function is

$$l(\boldsymbol{x}, \theta) = ln(L(\boldsymbol{x}, \theta)) = \begin{cases} \sum_{i=1}^{n} \ln f(x_i; \theta), & \text{continuous case} \\ \sum_{i=1}^{n} \ln p(x_i; \theta), & \text{discrete case.} \end{cases}$$

# ML Example

## Example

We switch on $n$ new electric bulbs and record their lifetimes as $x_1, \ldots, x_n$. These values are considered as random sample from an exponential distribution with density function

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad (x \geq 0),$$

where $\theta$ is unknown. Find the the likelihood and logarithmic likelihood functions.

# ML Example

## Example

We switch on $n$ new electric bulbs and record their lifetimes as $x_1, \ldots, x_n$. These values are considered as random sample from an exponential distribution with density function

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \ \ (x \geq 0),$$

where $\theta$ is unknown. Find the the likelihood and logarithmic likelihood functions.

Solution:

$$L(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \frac{1}{\theta^n} e^{-\sum x_i/\theta},$$

$$l(\boldsymbol{x}, \theta) = -n ln \theta - \sum_{i=1}^{n} \frac{x_i}{\theta}.$$

# The Method of Least Squares

Let $x_1, x_2, \ldots, x_n$ be a sample from a distribution with mean $E(X) = \mu(\theta)$, where $\mu(\theta)$ is a known function and $\theta$ an unknown parameter with parameter space $A$. Let

$$Q(\theta) = \sum_{i=1}^{n} [x_i - \mu(\theta)]^2$$

be the sum of the squares of the deviations of the observations from $\mu(\theta)$.

**Definition**

The value $\hat{\theta}$, for which $Q(\theta)$ obtains its minimal value within $A$, is called the **LS estimate** of $\theta$:

$$Q(\hat{\theta}) = min_{\theta \in A} Q(\theta)$$

# Generalization of the LS method

Let $x_1, \ldots, x_n$ be the observations on independent random variables $X_1, \ldots, X_n$ with different distributions depending on the same unknown parameter, where

$$E(X_i) = \mu_i(\theta), \quad Var(X_i) = \sigma_i^2.$$

In this case, we set

$$Q(\theta) = \sum_{i=1}^{n} \lambda_i [x_i - \mu(\theta)]^2,$$

where $\lambda_1, \ldots, \lambda_n$ are weights, for example $\lambda_i = \frac{1}{\sigma_i^2}$. An observation on a rv with large variance will then have less influence on the estimate than an observation with smaller variance.

# LS Example

## Example

An experimenter measures the constant of free fall and drops an object $n$ times independently from a given point and determines after time $t$ the distance it has fallen. The distances $x_1, \ldots, x_n$ vary because of various errors connected with the experiment. We regard the $x_i$'s as a random sample from a distribution with mean $\theta t^2/2$ and standard deviation $\sigma$ where $\sigma$ measures the experimental errors. Estimate the acceleration $\theta$ by means of the LS method.

# LS Example

**Example**

An experimenter measures the constant of free fall and drops an object $n$ times independently from a given point and determines after time $t$ the distance it has fallen. The distances $x_1, \ldots, x_n$ vary because of various errors connected with the experiment. We regard the $x_i$'s as a random sample from a distribution with mean $\theta t^2/2$ and standard deviation $\sigma$ where $\sigma$ measures the experimental errors. Estimate the acceleration $\theta$ by means of the LS method.

We find

$$Q(\theta) = \sum_{i=1}^{n} (x_i - \frac{\theta t^2}{2})^2.$$

# LS Example continues

Differentiating $Q(\theta)$ we obtain

$$\frac{dQ(\theta)}{d\theta} = \sum_{i=1}^{n} 2(x_i - \frac{\theta t^2}{2})(-\frac{t^2}{2}).$$

Setting this derivative equal to zero and solving for $\theta$ we obtain

$$\sum_{i=1}^{n} x_i - \frac{\theta n t^2}{2} = 0 \Rightarrow \quad \theta = \frac{2\sum_{i=1}^{n} x_i}{nt^2} = \frac{2\overline{x}}{t^2}$$

So the least square estimate for $\theta$ is

$$\hat{\theta} = \frac{2\overline{x}}{t^2}.$$

# Method of Moments

Let $X_1, X_2, \ldots, X_n$, $X_i \sim X$, be a theoretical sample from a distribution with $r$ unknown parameters, $(\theta_1, \theta_2, \ldots, \theta_r)$ and the $k$-th moments

$$\mu_k(X) = E(X^k), \quad k = 1, \ldots, r.$$

The $k$-th moment depends on the parameters $\theta_1, \ldots, \theta_r$, $\mu_k(X) = \mu_k(\theta_1, \ldots, \theta_r) = \mu_k(\boldsymbol{\theta})$.

**The $k$-th sample moment** is defined as

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

# Method of Moments (2)

The method of moments procedure equates the $k$-th theoretical moment with the corresponding $k$-th sample moment to obtain a system of $r$ equations in $r$ unknowns,

$$\mu_k(\boldsymbol{\theta}) = m_k, \quad k = 1, \ldots, r; \quad (*).$$

**Definition**

The solution to the system of equations $(*)$, $(\hat{\theta}_1, \ldots, \hat{\theta}_r)$, gives the **method of moments estimates** of $\theta_1, \ldots, \theta_r$.

**Remark**

$k$th sample moment $m_k$ is an unbiased estimate of a $k$th theoretical moment $\mu_k$.

# MM Example

**Example**

Let $X_1, X_2, \ldots, X_n \sim U(0, \theta)$, where $\theta$ is unknown. The first theoretical moment is $E(X_i) = \theta/2$, while the sample mean $\overline{x}$ is the first sample moment.
Thus, setting

$$\frac{\theta}{2} = \overline{x}$$

yields the method of moments estimator

$$\hat{\theta} = 2\overline{x}.$$