COMPUTATIONAL FINANCE & RISK MANAGEMENT

UNIVERSITY *of* WASHINGTON

Department of Applied Mathematics

CFRM 410: Probability and Statistics for Computational Finance

Week 9 Estimation

Jake Price
Instructor, Computational Finance and Risk Management
University of Washington
Slides originally produced by Kjell Konis

Fundamental Ideas of Statistics

Statistical Models

Point Estimation

Evaluating Point Estimators

Interval Estimators

Outline

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model: Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model: Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Data Set: Data $x_1, \ldots, x_n$ are assumed to be a realization of the random sample $X_1, \ldots, X_n \overset{iid}{\sim} f_{X|\theta}$ $\quad$ ($\Rightarrow$ representative)

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model: Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Data Set: Data $x_1, \ldots, x_n$ are assumed to be a realization of the random sample $X_1, \ldots, X_n \overset{iid}{\sim} f_{X|\theta}$ ($\Rightarrow$ representative)

Statistic: Random variable $T = t(X_1, \ldots, X_n)$

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model: Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Data Set: Data $x_1, \dots, x_n$ are assumed to be a realization of the random sample $X_1, \dots, X_n \overset{iid}{\sim} f_{X|\theta}$ $\quad$ ($\Rightarrow$ representative)

Statistic: Random variable $T = t(X_1, \dots, X_n)$

Estimator: Statistic used to estimate the parameter θ

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model	Density $f_{X\|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)
Data Set	Data $x_1, \ldots, x_n$ are assumed to be a realization of the random sample $X_1, \ldots, X_n \overset{iid}{\sim} f_{X\|\theta}$ ($\Rightarrow$ representative)
Statistic	Random variable $T = t(X_1, \ldots, X_n)$
Estimator	Statistic used to estimate the parameter θ
Notation	$T = t(X_1, \ldots, X_n)$ **statistic** (a random variable)

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Data Set Data $x_1, \ldots, x_n$ are assumed to be a realization of the random sample $X_1, \ldots, X_n \overset{iid}{\sim} f_{X|\theta}$ ($\Rightarrow$ representative)

Statistic Random variable $T = t(X_1, \ldots, X_n)$

Estimator Statistic used to estimate the parameter θ

Notation

$T = t(X_1, \ldots, X_n)$	**statistic** (a random variable)
$t = t(x_1, \ldots, x_n)$	**observed** value of the statistic T

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Data Set Data $x_1, \ldots, x_n$ are assumed to be a realization of the random sample $X_1, \ldots, X_n \stackrel{iid}{\sim} f_{X|\theta}$ ($\Rightarrow$ representative)

Statistic Random variable $T = t(X_1, \ldots, X_n)$

Estimator Statistic used to estimate the parameter θ

Notation

$T = t(X_1, \ldots, X_n)$	**statistic** (a random variable)
$t = t(x_1, \ldots, x_n)$	**observed** value of the statistic T
$\hat{\theta}(X_1, \ldots, X_n)$	**estimator** of the parameter θ

Statistical Models

Goal: learn something about a population by looking at a subset of the population called a *sample*

Probability Model: Density $f_{X|\theta}$ is assumed to be a member of a known parametric family (parameter θ is unknown)

Data Set: Data $x_1, \ldots, x_n$ are assumed to be a realization of the random sample $X_1, \ldots, X_n \overset{iid}{\sim} f_{X|\theta}$ $\quad$ ($\Rightarrow$ representative)

Statistic: Random variable $T = t(X_1, \ldots, X_n)$

Estimator: Statistic used to estimate the parameter θ

Notation:

$T = t(X_1, \ldots, X_n)$	**statistic** (a random variable)
$t = t(x_1, \ldots, x_n)$	**observed** value of the statistic T
$\hat{\theta}(X_1, \ldots, X_n)$	**estimator** of the parameter θ
$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$	**estimate** of the parameter θ

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ
- The observed value of the estimator $\hat{\mu}$ is $\bar{x}$

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ
- The observed value of the estimator $\hat{\mu}$ is $\bar{x}$
- $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an estimator of σ^2

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ
- The observed value of the estimator $\hat{\mu}$ is $\bar{x}$
- $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an estimator of σ^2
- The observed value of the estimator $\hat{\sigma}^2$ is $\sum_{i=1}^{n} (x_i - \bar{x})^2$

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ
- The observed value of the estimator $\hat{\mu}$ is $\bar{x}$
- $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an estimator of σ^2
- The observed value of the estimator $\hat{\sigma}^2$ is $\sum_{i=1}^{n} (x_i - \bar{x})^2$

Remarks

- Since a statistic T is a function of the random variables $X_1, \ldots, X_n$, T is itself a random variable

Statistical Models

Example Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ
- The observed value of the estimator $\hat{\mu}$ is $\bar{x}$
- $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ is an estimator of σ^2
- The observed value of the estimator $\hat{\sigma}^2$ is $\sum_{i=1}^n (x_i - \bar{x})^2$

Remarks

- Since a statistic T is a function of the random variables $X_1, \dots, X_n$, T is itself a random variable
- The distribution of T depends on the density of the X_i and is called the sampling distribution of T

Statistical Models

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- $\hat{\mu} = \bar{X}$ is an estimator of μ
- The observed value of the estimator $\hat{\mu}$ is $\bar{x}$
- $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an estimator of σ^2
- The observed value of the estimator $\hat{\sigma}^2$ is $\sum_{i=1}^{n} (x_i - \bar{x})^2$

Remarks

- Since a statistic T is a function of the random variables $X_1, \ldots, X_n$, T is itself a random variable
- The distribution of T depends on the density of the X_i and is called the sampling distribution of T
- $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ provide partial information on the distribution of T and are particularly useful when the distribution of T can be approximated

Problems to Study

Given a statistical model, might wish to:

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model
- Answer questions about the parameters of the model: for example, does $\theta = 0$?

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model
- Answer questions about the parameters of the model: for example, does $\theta = 0$?
- Make predictions about future observations

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model
- Answer questions about the parameters of the model: for example, does $\theta = 0$?
- Make predictions about future observations

There are several methods to estimate parameters, will discuss:

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model
- Answer questions about the parameters of the model: for example, does $\theta = 0$?
- Make predictions about future observations

There are several methods to estimate parameters, will discuss:

- The method of *moments* (easy - almost always works)

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model
- Answer questions about the parameters of the model: for example, does $\theta = 0$?
- Make predictions about future observations

There are several methods to estimate parameters, will discuss:

- The method of *moments* (easy - almost always works)
- The method of *least squares* (easy)

Problems to Study

Given a statistical model, might wish to:

- Estimate the parameters of the model
- Answer questions about the parameters of the model: for example, does $\theta = 0$?
- Make predictions about future observations

There are several methods to estimate parameters, will discuss:

- The method of *moments* (easy - almost always works)
- The method of *least squares* (easy)
- The method of *maximum likelihood* (more general - optimal in many situations)

Outline

Statistical Models

Point Estimation

Evaluating Point Estimators

Interval Estimators

Method of Moments

Let $X_1, \ldots, X_n$ be a representative random sample from a population characterized by a distribution with density $f_{X|\theta}$

Method of Moments

Let $X_1, \ldots, X_n$ be a representative random sample from a population characterized by a distribution with density $f_{X|\theta}$

To compute the method of moments estimator, equate the *empirical* moments (calculated from the sample) and the *theoretical* moments (calculated from the population, i.e, $f_{X|\theta}$)

Method of Moments

Let $X_1, \ldots, X_n$ be a representative random sample from a population characterized by a distribution with density $f_{X|\theta}$

To compute the method of moments estimator, equate the *empirical* moments (calculated from the sample) and the *theoretical* moments (calculated from the population, i.e, $f_{X|\theta}$)

The k^{th} moments are:

Theoretical $M'_k = \mathrm{E}(X^k)$

Method of Moments

Let $X_1, \ldots, X_n$ be a representative random sample from a population characterized by a distribution with density $f_{X|\theta}$

To compute the method of moments estimator, equate the *empirical* moments (calculated from the sample) and the *theoretical* moments (calculated from the population, i.e, $f_{X|\theta}$)

The k^{th} moments are:

Theoretical $M'_k = \mathrm{E}(X^k)$

Empirical $m'_k = \frac{1}{n} \sum_{i=1}^{k} X_i^k$

Method of Moments

Let $X_1, \ldots, X_n$ be a representative random sample from a population characterized by a distribution with density $f_{X|\theta}$

To compute the method of moments estimator, equate the *empirical* moments (calculated from the sample) and the *theoretical* moments (calculated from the population, i.e, $f_{X|\theta}$)

The k^{th} moments are:

Theoretical $M'_k = \mathrm{E}(X^k)$

Empirical $m'_k = \frac{1}{n}\sum_{i=1}^{k} X_i^k$

Need one moment for each parameter in the model

Example: Method of Moments for uniform$(0, \theta)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{uniform}(0, \theta)$

Example: Method of Moments for uniform$(0, \theta)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{uniform}(0, \theta)$

Find the method of moments estimator of θ

Example: Method of Moments for uniform$(0, \theta)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{uniform}(0, \theta)$

Find the method of moments estimator of θ

The first theoretical moment is: $M_1' = \mathrm{E}(X) = \frac{\theta+0}{2} = \frac{\theta}{2}$

Example: Method of Moments for uniform$(0, \theta)$

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{uniform}(0, \theta)$

Find the method of moments estimator of θ

The first theoretical moment is: $M_1' = \mathsf{E}(X) = \frac{\theta+0}{2} = \frac{\theta}{2}$

The first empirical moment is $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

Example: Method of Moments for uniform$(0, \theta)$

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ uniform$(0, \theta)$

Find the method of moments estimator of θ

The first theoretical moment is: $M_1' = \mathrm{E}(X) = \frac{\theta+0}{2} = \frac{\theta}{2}$

The first empirical moment is $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

The method of moments estimator for θ is thus

$$\hat{\theta}(X_1, \ldots, X_n) = \frac{2}{n}\sum_{i=1}^{n} X_i = 2\bar{X}$$

Example: Method of Moments for uniform$(0, \theta)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{uniform}(0, \theta)$

Find the method of moments estimator of θ

The first theoretical moment is: $M_1' = \mathrm{E}(X) = \frac{\theta+0}{2} = \frac{\theta}{2}$

The first empirical moment is $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

The method of moments estimator for θ is thus

$$\hat{\theta}(X_1, \ldots, X_n) = \frac{2}{n}\sum_{i=1}^{n} X_i = 2\bar{X}$$

Caveat: Suppose $\theta = 1$ and that $x_1 = 0.98$, $x_2 = 0.34$, $x_3 = 0.12$, $x_4 = 0.48$ and $x_5 = 0.08$ is a realization of a random sample. The method of moments estimate of θ is

$$\hat{\theta} = 2\bar{x} = 0.8$$

which is clearly not consistent with the model!

Example: Method of Moments for $\mathcal{N}(\mu, \sigma^2)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Example: Method of Moments for $\mathcal{N}(\mu, \sigma^2)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the method of moments estimators of μ and σ^2

Example: Method of Moments for $\mathcal{N}(\mu, \sigma^2)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the method of moments estimators of μ and σ^2

Theoretical moments

- $M_1' = \mathrm{E}(X) = \mu$
- $M_2' = \mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2 = \sigma^2 + \mu^2$

Example: Method of Moments for $\mathcal{N}(\mu, \sigma^2)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the method of moments estimators of μ and σ^2

Theoretical moments

- $M_1' = \mathsf{E}(X) = \mu$
- $M_2' = \mathsf{E}(X^2) = \mathsf{Var}(X) + [\mathsf{E}(X)]^2 = \sigma^2 + \mu^2$

Empirical moments

- $m_1' = \frac{1}{n}\sum_{i=1}^n X_i$
- $m_2' = \frac{1}{n}\sum_{i=1}^n X_i^2$

Example: Method of Moments for $\mathcal{N}(\mu, \sigma^2)$

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the method of moments estimators of μ and σ^2

Theoretical moments

- $M_1' = \mathsf{E}(X) = \mu$
- $M_2' = \mathsf{E}(X^2) = \mathsf{Var}(X) + [\mathsf{E}(X)]^2 = \sigma^2 + \mu^2$

Empirical moments

- $m_1' = \frac{1}{n}\sum_{i=1}^n X_i$
- $m_2' = \frac{1}{n}\sum_{i=1}^n X_i^2$

Method of moments estimators for μ and σ^2:

- $\hat{\mu}(X_1, \ldots, X_n) = m_1' = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$
- $\hat{\sigma}^2(X_1, \ldots, X_n) = m_2' - (m_1')^2 = \dfrac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$

Method of Least Squares

Let $X_1, \ldots, X_n$ be a random sample

Method of Least Squares

Let $X_1, \ldots, X_n$ be a random sample

Want to estimate θ, the expected value of X_i

Method of Least Squares

Let $X_1, \ldots, X_n$ be a random sample

Want to estimate θ, the expected value of X_i

Each random variable X_i should be "*close*" to θ

Method of Least Squares

Let $X_1, \ldots, X_n$ be a random sample

Want to estimate θ, the expected value of X_i

Each random variable X_i should be "*close*" to θ

Each difference $|X_i - \theta|$ should be *as small as possible*

Method of Least Squares

Let $X_1, \ldots, X_n$ be a random sample

Want to estimate θ, the expected value of X_i

Each random variable X_i should be "*close*" to θ

Each difference $|X_i - \theta|$ should be *as small as possible*

A reasonable estimator for θ would be the value minimizing

$$S(\theta) = \sum_{i=1}^{n} (X_i - \theta)^2$$

Example: Method of Least Squares

Let $X_1, \ldots, X_n \overset{iid}{\sim} f$ such that $\mathrm{E}(X_i) = \theta$

Example: Method of Least Squares

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} f$ such that $\mathrm{E}(X_i) = \theta$

Find the least squares estimator of θ

Example: Method of Least Squares

Let $X_1, \ldots, X_n \overset{iid}{\sim} f$ such that $\mathrm{E}(X_i) = \theta$

Find the least squares estimator of θ

$$S(\theta) = \sum_{i=1}^{n} (X_i - \theta)^2$$

Example: Method of Least Squares

Let $X_1, \ldots, X_n \overset{iid}{\sim} f$ such that $\mathrm{E}(X_i) = \theta$

Find the least squares estimator of θ

$$\begin{aligned} S(\theta) &= \sum_{i=1}^{n} (X_i - \theta)^2 \\ &= \sum_{i=1}^{n} (X_i - \bar{X} + \bar{X} - \theta)^2 \end{aligned}$$

Example: Method of Least Squares

Let $X_1, \ldots, X_n \overset{iid}{\sim} f$ such that $\mathrm{E}(X_i) = \theta$

Find the least squares estimator of θ

$$\begin{aligned}
S(\theta) &= \sum_{i=1}^{n}(X_i - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}[(X_i - \bar{X}) + (\bar{X} - \theta)]^2
\end{aligned}$$

Example: Method of Least Squares

Let $X_1, \ldots, X_n \overset{iid}{\sim} f$ such that $\mathrm{E}(X_i) = \theta$

Find the least squares estimator of θ

$$
\begin{aligned}
S(\theta) &= \sum_{i=1}^{n} (X_i - \theta)^2 \\
&= \sum_{i=1}^{n} (X_i - \bar{X} + \bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n} [(X_i - \bar{X}) + (\bar{X} - \theta)]^2 \\
&= \sum_{i=1}^{n} (X_i - \bar{X})^2 + 2\sum_{i=1}^{n} (X_i - \bar{X})(\bar{X} - \theta) + \sum_{i=1}^{n} (\bar{X} - \theta)^2
\end{aligned}
$$

Example: Method of Least Squares (continued)

$$= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta) + \sum_{i=1}^{n}(\bar{X} - \theta)^2$$

Example: Method of Least Squares (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\sum_{i=1}^{n}(X_i - \bar{X}) + \sum_{i=1}^{n}(\bar{X} - \theta)^2
\end{aligned}
$$

Example: Method of Least Squares (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\sum_{i=1}^{n}(X_i - \bar{X}) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\left[\left(\sum_{i=1}^{n} X_i\right) - n\bar{X}\right] + \sum_{i=1}^{n}(\bar{X} - \theta)^2
\end{aligned}
$$

Example: Method of Least Squares (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\sum_{i=1}^{n}(X_i - \bar{X}) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\left[\left(\sum_{i=1}^{n} X_i\right) - n\bar{X}\right] + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(\bar{X} - \theta)^2
\end{aligned}
$$

Example: Method of Least Squares (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\sum_{i=1}^{n}(X_i - \bar{X}) + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \theta)\left[\left(\sum_{i=1}^{n} X_i\right) - n\bar{X}\right] + \sum_{i=1}^{n}(\bar{X} - \theta)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(\bar{X} - \theta)^2
\end{aligned}
$$

$$\Longrightarrow \quad \hat{\theta}(X_1, \ldots, X_n) = \bar{X}$$

Method of Maximum Likelihood

The joint density of a random sample from a parametric family with parameter θ is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Method of Maximum Likelihood

The joint density of a random sample from a parametric family with parameter θ is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Let $x_1,\ldots,x_n$ (the data) be the realization of a random sample $X_1,\ldots,X_n \overset{iid}{\sim} f_X(x|\theta)$

Method of Maximum Likelihood

The joint density of a random sample from a parametric family with parameter θ is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Let $x_1,\ldots,x_n$ (the data) be the realization of a random sample $X_1,\ldots,X_n \stackrel{iid}{\sim} f_X(x|\theta)$

Definition The *likelihood* of the parameter θ is

$$L(\theta|x_1,\ldots,x_n) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Method of Maximum Likelihood

The joint density of a random sample from a parametric family with parameter θ is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Let $x_1,\ldots,x_n$ (the data) be the realization of a random sample $X_1,\ldots,X_n \overset{iid}{\sim} f_X(x|\theta)$

Definition The *likelihood* of the parameter θ is

$$L(\theta|x_1,\ldots,x_n) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Definition The *maximum likelihood estimator* $\hat{\theta}_{ML}$ of a parameter θ is a value of θ giving the largest likelihood possible:

$$L(\hat{\theta}_{ML}|x_1,\ldots,x_n) \geq L(\theta|x_1,\ldots,x_n)$$

for all admissible values of the parameter θ

Calculation of $\hat{\theta}_{ML}$

Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

Calculation of $\hat{\theta}_{ML}$

Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

General algorithm for finding $\hat{\theta}$

Calculation of $\hat{\theta}_{ML}$

Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

General algorithm for finding $\hat{\theta}$

1. Write down the likelihood function $L(\theta|x_1, \ldots, x_n)$

Calculation of $\hat{\theta}_{ML}$

Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

General algorithm for finding $\hat{\theta}$

1. Write down the likelihood function $L(\theta|x_1, \ldots, x_n)$
2. Define $\ell(\theta|x_1, \ldots, x_n) = \log L(\theta|x_1, \ldots, x_n)$ (the log likelihood of θ)

Calculation of $\hat{\theta}_{ML}$

Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

General algorithm for finding $\hat{\theta}$

1. Write down the likelihood function $L(\theta|x_1, \ldots, x_n)$
2. Define $\ell(\theta|x_1, \ldots, x_n) = \log L(\theta|x_1, \ldots, x_n)$ (the log likelihood of θ)
3. Find $\hat{\theta}_{ML}$ such that

$$\frac{d}{d\theta}\ell(\theta|x_1, \ldots, x_n)\bigg|_{\hat{\theta}_{ML}} = 0$$

Calculation of $\hat{\theta}_{ML}$

Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

General algorithm for finding $\hat{\theta}$

1. Write down the likelihood function $L(\theta|x_1, \ldots, x_n)$
2. Define $\ell(\theta|x_1, \ldots, x_n) = \log L(\theta|x_1, \ldots, x_n)$ (the log likelihood of θ)
3. Find $\hat{\theta}_{ML}$ such that

$$\frac{d}{d\theta}\ell(\theta|x_1, \ldots, x_n)\bigg|_{\hat{\theta}_{ML}} = 0$$

4. Verify that $\hat{\theta}_{ML}$ is a maximum

Example: Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p) be a random sample

Example: Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p) be a random sample

Find the maximum likelihood estimator of p

Example: Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ be a random sample

Find the maximum likelihood estimator of p

The likelihood function is

$$L(p|x_1, \ldots, x_n) \quad = \quad f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|p)$$

Example: Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ be a random sample

Find the maximum likelihood estimator of p

The likelihood function is

$$L(p|x_1, \ldots, x_n) \quad = \quad f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|p) = \prod_{i=1}^{n} f_X(x_i|p)$$

Example: Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ be a random sample

Find the maximum likelihood estimator of p

The likelihood function is

$$\begin{aligned} L(p|x_1, \dots, x_n) &= f_{X_1,\dots,X_n}(x_1, \dots, x_n|p) = \prod_{i=1}^{n} f_X(x_i|p) \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} \end{aligned}$$

Example: Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ be a random sample

Find the maximum likelihood estimator of p

The likelihood function is

$$\begin{aligned} L(p|x_1, \ldots, x_n) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|p) = \prod_{i=1}^{n} f_X(x_i|p) \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} \end{aligned}$$

The log likelihood function is

$$\ell(p|x_1, \ldots, x_n) = \log L(p|x_1, \ldots, x_n) = \log \left[\prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} \right]$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

$$= \sum_{i=1}^{n} \log \left[p^{x_i} (1-p)^{(1-x_i)} \right]$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n} \log\left[p^{x_i}(1-p)^{(1-x_i)}\right] \\
&= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i)
\end{aligned}
$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n} \log\left[p^{x_i}(1-p)^{(1-x_i)}\right] \\
&= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i) \\
&= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\left[n - \sum_{i=1}^{n} x_i\right]
\end{aligned}
$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

$$
\begin{aligned}
&= \sum_{i=1}^{n} \log\left[p^{x_i}(1-p)^{(1-x_i)}\right] \\
&= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i) \\
&= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\left[n - \sum_{i=1}^{n} x_i\right]
\end{aligned}
$$

Take the derivative of the log likelihood wrt the parameter

$$
\frac{d}{dp}\ell(p|x_1,\ldots,x_n) = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\left[n - \sum_{i=1}^{n} x_i\right]
$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Set the derivative equal to 0 and solve for $\hat{p}_{ML}$

$$\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i - \frac{1}{1 - \hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right] \quad \stackrel{\text{set}}{=} \quad 0$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Set the derivative equal to 0 and solve for $\hat{p}_{ML}$

$$
\begin{aligned}
\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i - \frac{1}{1-\hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right] &\stackrel{\text{set}}{=} 0 \\
\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i &= \frac{1}{1-\hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right]
\end{aligned}
$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Set the derivative equal to 0 and solve for $\hat{p}_{ML}$

$$
\begin{aligned}
\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i - \frac{1}{1-\hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right] &\overset{\text{set}}{=} 0 \\
\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i &= \frac{1}{1-\hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right] \\
(1-\hat{p}_{ML}) \sum_{i=1}^{n} x_i &= \hat{p}_{ML} \left[n - \sum_{i=1}^{n} x_i \right]
\end{aligned}
$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Set the derivative equal to 0 and solve for $\hat{p}_{ML}$

$$\begin{aligned}
\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i - \frac{1}{1-\hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right] &\overset{\text{set}}{=} 0 \\
\frac{1}{\hat{p}_{ML}} \sum_{i=1}^{n} x_i &= \frac{1}{1-\hat{p}_{ML}} \left[n - \sum_{i=1}^{n} x_i \right] \\
(1-\hat{p}_{ML}) \sum_{i=1}^{n} x_i &= \hat{p}_{ML} \left[n - \sum_{i=1}^{n} x_i \right] \\
\sum_{i=1}^{n} x_i - \hat{p}_{ML} \sum_{i=1}^{n} x_i &= n\hat{p}_{ML} - \hat{p}_{ML} \sum_{i=1}^{n} x_i
\end{aligned}$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Set the derivative equal to 0 and solve for $\hat{p}_{ML}$

$$\begin{aligned}
\frac{1}{\hat{p}_{ML}}\sum_{i=1}^{n} x_i - \frac{1}{1-\hat{p}_{ML}}\left[n - \sum_{i=1}^{n} x_i\right] &\overset{\text{set}}{=} 0 \\
\frac{1}{\hat{p}_{ML}}\sum_{i=1}^{n} x_i &= \frac{1}{1-\hat{p}_{ML}}\left[n - \sum_{i=1}^{n} x_i\right] \\
(1-\hat{p}_{ML})\sum_{i=1}^{n} x_i &= \hat{p}_{ML}\left[n - \sum_{i=1}^{n} x_i\right] \\
\sum_{i=1}^{n} x_i - \hat{p}_{ML}\sum_{i=1}^{n} x_i &= n\hat{p}_{ML} - \hat{p}_{ML}\sum_{i=1}^{n} x_i \\
\hat{p}_{ML} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Finally, verify that $\hat{p}_{ML}$ is indeed a maximum

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Finally, verify that $\hat{p}_{ML}$ is indeed a maximum

$$\frac{d^2}{dp^2}\ell(p|x_1,\ldots,x_n) = -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)^2}\left[n - \sum_{i=1}^{n} x_i\right]$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Finally, verify that $\hat{p}_{ML}$ is indeed a maximum

$$
\begin{aligned}
\frac{d^2}{dp^2}\ell(p|x_1,\ldots,x_n) &= -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)^2}\left[n - \sum_{i=1}^{n} x_i\right] \\
&< 0 \quad \text{for all } p \in (0,1)
\end{aligned}
$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Finally, verify that $\hat{p}_{ML}$ is indeed a maximum

$$
\begin{aligned}
\frac{d^2}{dp^2}\ell(p|x_1,\ldots,x_n) &= -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)^2}\left[n - \sum_{i=1}^{n} x_i\right] \\
&< 0 \quad \text{for all } p \in (0,1)
\end{aligned}
$$

$\Longrightarrow$ $\hat{p}_{ML}$ is indeed a maximum of the log likelihood

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Finally, verify that $\hat{p}_{ML}$ is indeed a maximum

$$
\begin{aligned}
\frac{d^2}{dp^2}\ell(p|x_1,\ldots,x_n) &= -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)^2}\left[n - \sum_{i=1}^{n} x_i\right] \\
&< 0 \quad \text{for all } p \in (0,1)
\end{aligned}
$$

$\Longrightarrow$ $\hat{p}_{ML}$ is indeed a maximum of the log likelihood

Have shown that for every possible $(X_1 = x_1, \ldots, X_n = x_n)$

$\hat{p}_{ML} = \bar{x}$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (continued)

Finally, verify that $\hat{p}_{ML}$ is indeed a maximum

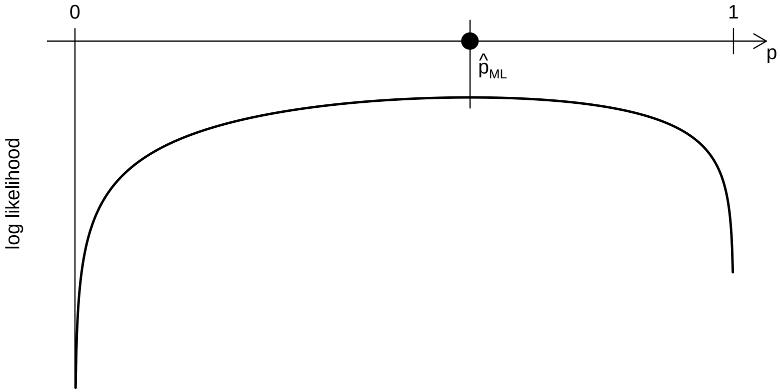
$$\begin{aligned}\frac{d^2}{dp^2}\ell(p|x_1,\ldots,x_n) &= -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)^2}\left[n - \sum_{i=1}^{n} x_i\right] \\ &< 0 \quad \text{for all } p \in (0,1)\end{aligned}$$

$\implies$ $\hat{p}_{ML}$ is indeed a maximum of the log likelihood

Have shown that for every possible $(X_1 = x_1, \ldots, X_n = x_n)$

$$\hat{p}_{ML} = \bar{x} \quad \implies \quad \hat{p}_{ML} = \bar{X}$$

Find $\hat{p}_{ML}$ for a Sequence of Bernoulli Trials (illustration)



Example: Find $\hat{\mu}$ for a Normal Population

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Example: Find $\hat{\mu}$ for a Normal Population

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the maximum likelihood estimator of μ

Example: Find $\hat{\mu}$ for a Normal Population

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the maximum likelihood estimator of μ

The likelihood function is

$$L(\mu, \sigma^2 | x_1, \ldots, x_n) = f_{X_1, \ldots, X_n}(x_1, \ldots, x_n | \mu, \sigma^2)$$

Example: Find $\hat{\mu}$ for a Normal Population

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the maximum likelihood estimator of μ

The likelihood function is

$$\begin{aligned} L(\mu, \sigma^2 | x_1, \ldots, x_n) &= f_{X_1, \ldots, X_n}(x_1, \ldots, x_n | \mu, \sigma^2) \\ &= \prod_{i=1}^{n} f_X(x_i | \mu, \sigma^2) \end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the maximum likelihood estimator of μ

The likelihood function is

$$\begin{aligned} L(\mu, \sigma^2 | x_1, \ldots, x_n) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n | \mu, \sigma^2) \\ &= \prod_{i=1}^{n} f_X(x_i | \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

The log likelihood function is

$$\ell(\mu, \sigma^2 | x_1, \ldots, x_n) = \log L(\mu, \sigma^2 | x_1, \ldots, x_n)$$

Example: Find $\hat{\mu}$ for a Normal Population

The log likelihood function is

$$\begin{aligned}
\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \log L(\mu, \sigma^2 | x_1, \ldots, x_n) \\
&= \log \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right]
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

The log likelihood function is

$$
\begin{aligned}
\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \log L(\mu, \sigma^2 | x_1, \ldots, x_n) \\
&= \log \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] \\
&= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n} \left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right)
\end{aligned}
$$

Example: Find $\hat{\mu}$ for a Normal Population

The log likelihood function is

$$\begin{aligned}
\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \log L(\mu, \sigma^2 | x_1, \ldots, x_n) \\
&= \log \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] \\
&= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n} \left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \\
&= -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

The log likelihood function is

$$
\begin{aligned}
\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \log L(\mu, \sigma^2 | x_1, \ldots, x_n) \\
&= \log \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] \\
&= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n} \left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \\
&= -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \\
&\propto -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
\end{aligned}
$$

Example: Find $\hat{\mu}$ for a Normal Population

Take the derivative of the log likelihood

$$\frac{d}{d\mu}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) \quad = \quad \frac{d}{d\mu}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

Example: Find $\hat{\mu}$ for a Normal Population

Take the derivative of the log likelihood

$$\begin{aligned}
\frac{d}{d\mu}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

Take the derivative of the log likelihood

$$
\begin{aligned}
\frac{d}{d\mu}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)
\end{aligned}
$$

Set the derivative equal to 0 and solve for $\hat{\mu}_{ML}$

$$
\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu}_{ML}) \stackrel{\text{set}}{=} 0
$$

Example: Find $\hat{\mu}$ for a Normal Population

Take the derivative of the log likelihood

$$\begin{aligned}
\frac{d}{d\mu}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)
\end{aligned}$$

Set the derivative equal to 0 and solve for $\hat{\mu}_{ML}$

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu}_{ML}) &\overset{\text{set}}{=} 0 \\
\sum_{i=1}^{n} x_i &= n\hat{\mu}_{ML}
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

Take the derivative of the log likelihood

$$\begin{aligned} \frac{d}{d\mu}\ell(\mu,\sigma^2|x_1,\ldots,x_n) &= \frac{d}{d\mu}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \end{aligned}$$

Set the derivative equal to 0 and solve for $\hat{\mu}_{ML}$

$$\begin{aligned} \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu}_{ML}) &\overset{\text{set}}{=} 0 \\ \sum_{i=1}^{n}x_i &= n\hat{\mu}_{ML} \quad \Longrightarrow \quad \hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x} \end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

Finally, need to verify that $\hat{\mu}_{ML}$ is indeed a maximum

$$\frac{d^2}{d\mu^2}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) \quad = \quad \frac{d}{d\mu}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right]$$

Example: Find $\hat{\mu}$ for a Normal Population

Finally, need to verify that $\hat{\mu}_{ML}$ is indeed a maximum

$$\begin{aligned}
\frac{d^2}{d\mu^2}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right] \\
&= -\frac{n}{\sigma^2}
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

Finally, need to verify that $\hat{\mu}_{ML}$ is indeed a maximum

$$\begin{aligned}
\frac{d^2}{d\mu^2}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right] \\
&= -\frac{n}{\sigma^2} \\
&< 0
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

Finally, need to verify that $\hat{\mu}_{ML}$ is indeed a maximum

$$
\begin{aligned}
\frac{d^2}{d\mu^2}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right] \\
&= -\frac{n}{\sigma^2} \\
&< 0
\end{aligned}
$$

$\implies$ $\hat{\mu}_{ML}$ is a local maximum of the log likelihood

Example: Find $\hat{\mu}$ for a Normal Population

Finally, need to verify that $\hat{\mu}_{ML}$ is indeed a maximum

$$
\begin{aligned}
\frac{d^2}{d\mu^2}\ell(\mu, \sigma^2 | x_1, \ldots, x_n) &= \frac{d}{d\mu}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right] \\
&= -\frac{n}{\sigma^2} \\
&< 0
\end{aligned}
$$

$\implies$ $\hat{\mu}_{ML}$ is a local maximum of the log likelihood

The maximum likelihood estimator of μ is

$\hat{\mu}_{ML} = \bar{X}$

Outline

Statistical Models

Point Estimation

Evaluating Point Estimators

Interval Estimators

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{Bias}_{\hat{\theta}}(\theta) = \mathsf{E}(\hat{\theta}) - \theta$$

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{Bias}_{\hat{\theta}}(\theta) = \mathsf{E}(\hat{\theta}) - \theta$$

Interpretation of bias:

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

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

Interpretation of bias:

- If $\text{Bias}_{\hat{\theta}}(\theta) < 0$ then $\hat{\theta}$ tends to underestimate θ

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

$$\mathrm{Bias}_{\hat{\theta}}(\theta) = \mathsf{E}(\hat{\theta}) - \theta$$

Interpretation of bias:

- If $\mathrm{Bias}_{\hat{\theta}}(\theta) < 0$ then $\hat{\theta}$ tends to underestimate θ
- If $\mathrm{Bias}_{\hat{\theta}}(\theta) > 0$ then $\hat{\theta}$ tends to overestimate θ

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

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

Interpretation of bias:

- If $\text{Bias}_{\hat{\theta}}(\theta) < 0$ then $\hat{\theta}$ tends to underestimate θ
- If $\text{Bias}_{\hat{\theta}}(\theta) > 0$ then $\hat{\theta}$ tends to overestimate θ
- If $\text{Bias}_{\hat{\theta}}(\theta) \equiv 0$ then $\hat{\theta}$ is an *unbiased* estimator of θ

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

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

Interpretation of bias:

- If $\text{Bias}_{\hat{\theta}}(\theta) < 0$ then $\hat{\theta}$ tends to underestimate θ
- If $\text{Bias}_{\hat{\theta}}(\theta) > 0$ then $\hat{\theta}$ tends to overestimate θ
- If $\text{Bias}_{\hat{\theta}}(\theta) \equiv 0$ then $\hat{\theta}$ is an *unbiased* estimator of θ

A qualitative measure of an estimator $\hat{\theta}$ is the absence of a systematic relationship between $\hat{\theta}$ and θ

Bias

The *bias* of an estimator $\hat{\theta}$ is defined to be

$$\mathrm{Bias}_{\hat{\theta}}(\theta) = \mathrm{E}(\hat{\theta}) - \theta$$

Interpretation of bias:

- If $\mathrm{Bias}_{\hat{\theta}}(\theta) < 0$ then $\hat{\theta}$ tends to underestimate θ
- If $\mathrm{Bias}_{\hat{\theta}}(\theta) > 0$ then $\hat{\theta}$ tends to overestimate θ
- If $\mathrm{Bias}_{\hat{\theta}}(\theta) \equiv 0$ then $\hat{\theta}$ is an *unbiased* estimator of θ

A qualitative measure of an estimator $\hat{\theta}$ is the absence of a systematic relationship between $\hat{\theta}$ and θ

Later: better measure is given by the average size of $(\hat{\theta} - \theta)^2$

Example: Bias

Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Example: Bias

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the bias and the variance of $\hat{\mu} = \bar{X}$

Example: Bias

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the bias and the variance of $\hat{\mu} = \bar{X}$

$$\text{Bias}_{\hat{\mu}}(\mu) = \text{E}(\bar{X}) - \mu$$

Example: Bias

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the bias and the variance of $\hat{\mu} = \bar{X}$

$$
\begin{aligned}
\text{Bias}_{\hat{\mu}}(\mu) &= \text{E}(\bar{X}) - \mu \\
&= \text{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu
\end{aligned}
$$

Example: Bias

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the bias and the variance of $\hat{\mu} = \bar{X}$

$$\begin{aligned}
\text{Bias}_{\hat{\mu}}(\mu) &= \mathrm{E}(\bar{X}) - \mu \\
&= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) - \mu
\end{aligned}$$

Example: Bias

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the bias and the variance of $\hat{\mu} = \bar{X}$

$$
\begin{aligned}
\text{Bias}_{\hat{\mu}}(\mu) &= \mathrm{E}(\bar{X}) - \mu \\
&= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) - \mu \\
&= \frac{1}{n} n\,\mu - \mu
\end{aligned}
$$

Example: Bias

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find the bias and the variance of $\hat{\mu} = \bar{X}$

$$
\begin{aligned}
\text{Bias}_{\hat{\mu}}(\mu) &= \text{E}(\bar{X}) - \mu \\
&= \text{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu \\
&= \frac{1}{n}\sum_{i=1}^{n} \text{E}(X_i) - \mu \\
&= \frac{1}{n} n\,\mu - \mu \\
&= 0 \implies \bar{X} \text{ is an unbiased estimator of } \mu
\end{aligned}
$$

Example: Bias (continued)

$$\mathrm{Var}(\bar{X}) \quad = \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

Example: Bias (continued)

$$
\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)
\end{aligned}
$$

Example: Bias (continued)

$$
\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \qquad \text{(b/c the } X_i \text{ are independent)}
\end{aligned}
$$

Example: Bias (continued)

$$
\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) && \text{(b/c the } X_i \text{ are independent)} \\
&= \frac{1}{n^2}\, n\, \sigma^2
\end{aligned}
$$

Example: Bias (continued)

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) && \text{(b/c the } X_i \text{ are independent)} \\
&= \frac{1}{n^2}\, n\, \sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Example: Bias (continued)

Find the bias of $\hat{\sigma}^2 = \dfrac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

Example: Bias (continued)

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

$$\text{Bias}_{\hat{\sigma}^2}(\sigma^2) = \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2$$

Example: Bias (continued)

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

$$
\begin{aligned}
\mathrm{Bias}_{\hat{\sigma}^2}(\sigma^2) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\mathrm{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2
\end{aligned}
$$

Example: Bias (continued)

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

$$
\begin{aligned}
\text{Bias}_{\hat{\sigma}^2}(\sigma^2) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\mathrm{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\mathrm{E}(S^2) - \sigma^2
\end{aligned}
$$

Example: Bias (continued)

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

$$
\begin{aligned}
\text{Bias}_{\hat{\sigma}^2}(\sigma^2) &= \text{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\text{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\text{E}(S^2) - \sigma^2 \\
&= \frac{n-1}{n}\sigma^2 - \sigma^2
\end{aligned}
$$

Example: Bias (continued)

Find the bias of $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$\begin{aligned}
\text{Bias}_{\hat{\sigma}^2}(\sigma^2) &= \text{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\text{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 \\
&= \frac{n-1}{n}\text{E}(S^2) - \sigma^2 \\
&= \frac{n-1}{n}\sigma^2 - \sigma^2 \\
&= \frac{-\sigma^2}{n} < 0 \quad \Rightarrow \quad \hat{\sigma}^2 \text{ tends to underestimate } \sigma^2
\end{aligned}$$

Mean Squared Error

The *mean squared error* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Mean Squared Error

The *mean squared error* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Bias-variance tradeoff: the mean squared error has the following decomposition

Mean Squared Error

The *mean squared error* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Bias-variance tradeoff: the mean squared error has the following decomposition

$$\mathsf{MSE}(\hat{\theta}) \quad = \quad \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Mean Squared Error

The *mean squared error* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Bias-variance tradeoff: the mean squared error has the following decomposition

$$\mathsf{MSE}(\hat{\theta}) \;\; = \;\; \mathsf{E}\left[(\hat{\theta} - \theta)^2\right] = \mathsf{E}\left[\left([\hat{\theta} - \mathsf{E}(\hat{\theta})] + [\mathsf{E}(\hat{\theta}) - \theta]\right)^2\right]$$

Mean Squared Error

The *mean squared error* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Bias-variance tradeoff: the mean squared error has the following decomposition

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

Mean Squared Error

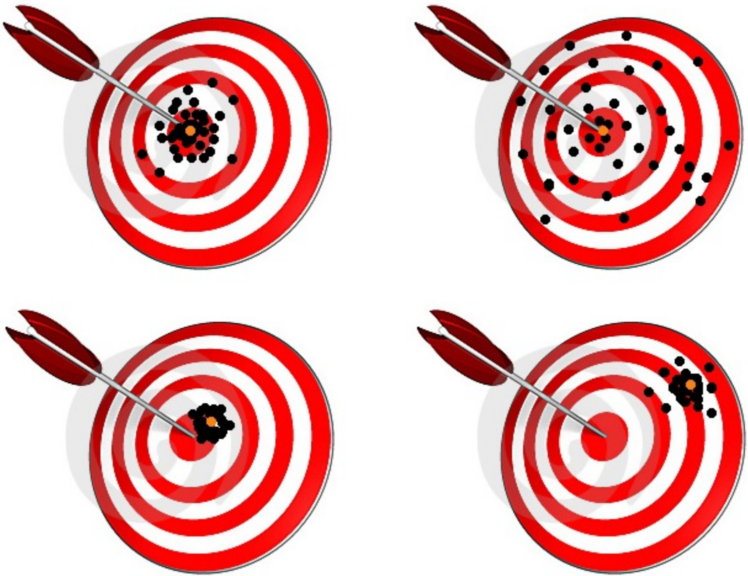
The *mean squared error* of an estimator $\hat{\theta}$ is defined to be

$$\mathsf{MSE}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$$

Bias-variance tradeoff: the mean squared error has the following decomposition

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

Bias-Variance Tradeoff (illustration)



Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

$\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$ if

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

$\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$ if

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

In this case, the estimator $\hat{\theta}_1$ is preferable

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

$\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$ if

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

In this case, the estimator $\hat{\theta}_1$ is preferable

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

$\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$ if

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

In this case, the estimator $\hat{\theta}_1$ is preferable

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- Sampling distribution of the median M is $\approx \mathcal{N}(\mu, \frac{\sigma^2 \pi}{2n})$

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

$\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$ if

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

In this case, the estimator $\hat{\theta}_1$ is preferable

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- Sampling distribution of the median M is $\approx \mathcal{N}(\mu, \frac{\sigma^2 \pi}{2n})$
- Sampling distribution of the sample mean $\bar{X}$ is $\mathcal{N}(\mu, \frac{\sigma^2}{n})$

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of same parameter θ

- $\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) + [\text{Bias}_{\hat{\theta}_1}(\theta)]^2 = \text{Var}(\hat{\theta}_1)$
- $\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) + [\text{Bias}_{\hat{\theta}_2}(\theta)]^2 = \text{Var}(\hat{\theta}_2)$

$\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$ if

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

In this case, the estimator $\hat{\theta}_1$ is preferable

Example Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- Sampling distribution of the median M is $\approx \mathcal{N}(\mu, \frac{\sigma^2 \pi}{2n})$
- Sampling distribution of the sample mean $\bar{X}$ is $\mathcal{N}(\mu, \frac{\sigma^2}{n})$
- Which estimator of μ is preferable, $\bar{X}$ or M?

Consistency

Let $X_1, X_2, \ldots$ be a sequence of random variables

Consistency

Let $X_1, X_2, \ldots$ be a sequence of random variables

Let $\bar{X}_1 = X_1$, $\bar{X}_2 = (X_1 + X_2)/2$, $\bar{X}_3 = (X_1 + X_2 + X_3)/3$, $\ldots$

Consistency

Let $X_1, X_2, \ldots$ be a sequence of random variables

Let $\bar{X}_1 = X_1$, $\bar{X}_2 = (X_1 + X_2)/2$, $\bar{X}_3 = (X_1 + X_2 + X_3)/3$, $\ldots$

Call $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots$ a sequence of estimators

Consistency

Let $X_1, X_2, \dots$ be a sequence of random variables

Let $\bar{X}_1 = X_1$, $\bar{X}_2 = (X_1 + X_2)/2$, $\bar{X}_3 = (X_1 + X_2 + X_3)/3$, $\dots$

Call $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots$ a sequence of estimators

A sequence of estimators $W_n = W_n(X_1, \dots, X_n)$ is a *consistent sequence of estimators* of the parameter θ if for every $\epsilon > 0$ and every θ

$$\lim_{n \to \infty} \mathrm{P}(|W_n - \theta| < \epsilon) = 1$$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

$$\mathrm{P}(|\bar{X}_n - \theta| < \epsilon) = \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)(\bar{x}_n - \theta)^2} \, d\bar{x}_n$$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

$$\begin{aligned} \mathrm{P}(|\bar{X}_n - \theta| < \epsilon) &= \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)(\bar{x}_n-\theta)^2}\, d\bar{x}_n \\ &= \int_{-\epsilon}^{\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)y^2}\, dy \end{aligned}$$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

$$
\begin{aligned}
\mathrm{P}(|\bar{X}_n - \theta| < \epsilon) &= \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)(\bar{x}_n-\theta)^2}\, d\bar{x}_n \\
&= \int_{-\epsilon}^{\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)y^2}\, dy \\
&= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt
\end{aligned}
$$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

$$
\begin{aligned}
\mathrm{P}(|\bar{X}_n - \theta| < \epsilon) &= \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)(\bar{x}_n-\theta)^2}\, d\bar{x}_n \\
&= \int_{-\epsilon}^{\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)y^2}\, dy \\
&= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \\
&= \mathrm{P}(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n})
\end{aligned}
$$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

$$\begin{aligned}
\mathrm{P}(|\bar{X}_n - \theta| < \epsilon) &= \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)(\bar{x}_n-\theta)^2}\, d\bar{x}_n \\
&= \int_{-\epsilon}^{\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)y^2}\, dy \\
&= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \\
&= \mathrm{P}(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \\
&\quad \to 1 \;\text{ as }\; n \to \infty
\end{aligned}$$

Example: Consistency

Let $X_1, X_2, \ldots$ be a sequence of iid $\mathcal{N}(\theta, 1)$ random variables

Recall $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

$$
\begin{aligned}
\mathrm{P}(|\bar{X}_n - \theta| < \epsilon) &= \int_{\theta-\epsilon}^{\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)(\bar{x}_n-\theta)^2}\, d\bar{x}_n \\
&= \int_{-\epsilon}^{\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-(n/2)y^2}\, dy \\
&= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \\
&= \mathrm{P}(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \\
&\quad \to 1 \text{ as } n \to \infty
\end{aligned}
$$

$\Longrightarrow$ $\bar{X}_n$ is a consistent sequence of estimators of θ

The Delta Method

Let $X_1, X_2, \ldots$ be iid random variables with $\mathsf{E}(X_i) = \mu$, $\mathsf{Var}(X_i) = \sigma^2$

The Delta Method

Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Central Limit Theorem says

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0, 1)$$

The Delta Method

Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Central Limit Theorem says

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

Let $g(t) = at + b$ be a linear function

The Delta Method

Let $X_1, X_2, \dots$ be iid random variables with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Central Limit Theorem says

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

Let $g(t) = at + b$ be a linear function

- $\mathrm{E}\big(g(\bar{X}_n)\big) = \mathrm{E}\big(a\bar{X}_n + b\big) = a\,\mathrm{E}(\bar{X}_n) + b = a\mu + b$

The Delta Method

Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Central Limit Theorem says

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

Let $g(t) = at + b$ be a linear function

- $\mathrm{E}\big(g(\bar{X}_n)\big) = \mathrm{E}\big(a\bar{X}_n + b\big) = a\,\mathrm{E}(\bar{X}_n) + b = a\mu + b$
- $\mathrm{Var}\big(g(\bar{X}_n)\big) = \mathrm{Var}\big(a\bar{X}_n + b\big) = a^2\,\mathrm{Var}(\bar{X}_n) = \dfrac{a^2\sigma^2}{n}$

The Delta Method

Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Central Limit Theorem says

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

Let $g(t) = at + b$ be a linear function

- $\mathrm{E}\big(g(\bar{X}_n)\big) = \mathrm{E}\big(a\bar{X}_n + b\big) = a\,\mathrm{E}(\bar{X}_n) + b = a\mu + b$
- $\mathrm{Var}\big(g(\bar{X}_n)\big) = \mathrm{Var}\big(a\bar{X}_n + b\big) = a^2\,\mathrm{Var}(\bar{X}_n) = \dfrac{a^2\sigma^2}{n}$

Central Limit Theorem result for $g(\bar{X}_n)$

$$\frac{\big(g(\bar{X}_n) - g(\mu)\big)}{a\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

The Delta Method

Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Central Limit Theorem says

$$\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

Let $g(t) = at + b$ be a linear function

- $\mathrm{E}\big(g(\bar{X}_n)\big) = \mathrm{E}\big(a\bar{X}_n + b\big) = a\,\mathrm{E}(\bar{X}_n) + b = a\mu + b$
- $\mathrm{Var}\big(g(\bar{X}_n)\big) = \mathrm{Var}\big(a\bar{X}_n + b\big) = a^2\,\mathrm{Var}(\bar{X}_n) = \dfrac{a^2\sigma^2}{n}$

Central Limit Theorem result for $g(\bar{X}_n)$

$$\frac{\big(g(\bar{X}_n) - g(\mu)\big)}{a\sigma/\sqrt{n}} \to Z \qquad Z \sim \mathcal{N}(0,1)$$

Can write as

$$\sqrt{n}\big(g(\bar{X}_n) - g(\mu)\big) \to a\sigma Z$$

The Delta Method

Now, suppose $g(t)$ is a nonlinear function and that $g'(\mu)$ exists

The Delta Method

Now, suppose $g(t)$ is a nonlinear function and that $g'(\mu)$ exists

If $\bar{X}_n$ is a consistent estimator of μ, then $\bar{X}_n$ close to μ for large n

The Delta Method

Now, suppose $g(t)$ is a nonlinear function and that $g'(\mu)$ exists

If $\bar{X}_n$ is a consistent estimator of μ, then $\bar{X}_n$ close to μ for large n

Can approximate $g(t)$ with a linear function

$$g(t) \approx g(\mu) + g'(\mu)(t - \mu)$$

The Delta Method

Now, suppose $g(t)$ is a nonlinear function and that $g'(\mu)$ exists

If $\bar{X}_n$ is a consistent estimator of μ, then $\bar{X}_n$ close to μ for large n

Can approximate $g(t)$ with a linear function

$$g(t) \approx g(\mu) + g'(\mu)(t - \mu)$$

$$\sqrt{n}\big(g(\bar{X}_n) - g(\mu)\big) \to g'(\mu)\sigma Z \sim \mathcal{N}\big(0, \sigma^2[g'(\mu)]^2\big)$$

Example: Estimating Binomial Variance

Let $X_n \sim \text{binomial}(n, p)$ (sum of n independent Bernoulli(p) trials)

Example: Estimating Binomial Variance

Let $X_n \sim \text{binomial}(n, p)$ (sum of n independent Bernoulli(p) trials)

MLE $\frac{1}{n}X_n$ consistent, unbiased estimator of p

Example: Estimating Binomial Variance

Let $X_n \sim \text{binomial}(n, p)$ (sum of n independent Bernoulli(p) trials)

MLE $\frac{1}{n}X_n$ consistent, unbiased estimator of p

Let $g(t) = t(1-t)$

$$\implies g\left(\frac{X_n}{n}\right) = \frac{X_n}{n}\left(1 - \frac{X_n}{n}\right) = \frac{X_n(n - X_n)}{n^2}$$

an estimator for the variance $p(1-p)$

Example: Estimating Binomial Variance

Let $X_n \sim \text{binomial}(n, p)$ (sum of n independent Bernoulli(p) trials)

MLE $\frac{1}{n}X_n$ consistent, unbiased estimator of p

Let $g(t) = t(1-t)$

$$\implies g\left(\frac{X_n}{n}\right) = \frac{X_n}{n}\left(1 - \frac{X_n}{n}\right) = \frac{X_n(n - X_n)}{n^2}$$

an estimator for the variance $p(1-p)$

Since $g'(p) = 1 - 2p$

$$\sqrt{n}\left(g\left(\frac{X_n}{n}\right) - p(1-p)\right) \to \mathcal{N}\left(0, p(1-p)(1-2p)^2\right)$$

Outline

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Let $\hat{\theta}$ be an estimator of θ and suppose that there exists a function $T = q(\hat{\theta}, \theta)$ whose distribution is known and does not depend on θ

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Let $\hat{\theta}$ be an estimator of θ and suppose that there exists a function $T = q(\hat{\theta}, \theta)$ whose distribution is known and does not depend on θ

Let $\alpha_1 + \alpha_2 = \alpha < 1$ an let a and b be constants such that

$$\alpha_1 = \mathrm{P}(T < a) \qquad \alpha_2 = \mathrm{P}(T > b) \qquad (\text{often } \alpha_1 = \alpha_2)$$

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Let $\hat{\theta}$ be an estimator of θ and suppose that there exists a function $T = q(\hat{\theta}, \theta)$ whose distribution is known and does not depend on θ

Let $\alpha_1 + \alpha_2 = \alpha < 1$ an let a and b be constants such that

$$\alpha_1 = \mathrm{P}(T < a) \qquad \alpha_2 = \mathrm{P}(T > b) \qquad \text{(often } \alpha_1 = \alpha_2\text{)}$$

Then

$$\mathrm{P}(a \leq T \leq b) = \mathrm{P}(T \leq b) - \mathrm{P}(T < a) = (1 - \alpha_2) - \alpha_1 = 1 - \alpha$$

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Let $\hat{\theta}$ be an estimator of θ and suppose that there exists a function $T = q(\hat{\theta}, \theta)$ whose distribution is known and does not depend on θ

Let $\alpha_1 + \alpha_2 = \alpha < 1$ an let a and b be constants such that

$$\alpha_1 = \mathsf{P}(T < a) \qquad \alpha_2 = \mathsf{P}(T > b) \qquad (\text{often } \alpha_1 = \alpha_2)$$

Then

$$\mathsf{P}(a \leq T \leq b) = \mathsf{P}(T \leq b) - \mathsf{P}(T < a) = (1 - \alpha_2) - \alpha_1 = 1 - \alpha$$

Isolate θ to find random variables I and S such that

$$\mathsf{P}(I \leq \theta \leq S) = 1 - \alpha$$

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Let $\hat{\theta}$ be an estimator of θ and suppose that there exists a function $T = q(\hat{\theta}, \theta)$ whose distribution is known and does not depend on θ

Let $\alpha_1 + \alpha_2 = \alpha < 1$ an let a and b be constants such that

$$\alpha_1 = \mathrm{P}(T < a) \qquad \alpha_2 = \mathrm{P}(T > b) \qquad (\text{often } \alpha_1 = \alpha_2)$$

Then

$$\mathrm{P}(a \leq T \leq b) = \mathrm{P}(T \leq b) - \mathrm{P}(T < a) = (1 - \alpha_2) - \alpha_1 = 1 - \alpha$$

Isolate θ to find random variables I and S such that

$$\mathrm{P}(I \leq \theta \leq S) = 1 - \alpha$$

$[I, S]$ is called a $100(1 - \alpha)\%$ *confidence interval* for θ

Interval Estimators

Instead of estimating the parameter θ with only a number $\hat{\theta}$, consider a random interval that covers θ with high probability

Let $\hat{\theta}$ be an estimator of θ and suppose that there exists a function $T = q(\hat{\theta}, \theta)$ whose distribution is known and does not depend on θ

Let $\alpha_1 + \alpha_2 = \alpha < 1$ an let a and b be constants such that

$$\alpha_1 = \mathrm{P}(T < a) \qquad \alpha_2 = \mathrm{P}(T > b) \qquad (\text{often } \alpha_1 = \alpha_2)$$

Then

$$\mathrm{P}(a \leq T \leq b) = \mathrm{P}(T \leq b) - \mathrm{P}(T < a) = (1 - \alpha_2) - \alpha_1 = 1 - \alpha$$

Isolate θ to find random variables I and S such that

$$\mathrm{P}(I \leq \theta \leq S) = 1 - \alpha$$

$[I, S]$ is called a $100(1-\alpha)\%$ *confidence interval* for θ

The quantity $1 - \alpha$ is called the *confidence coefficient*

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, 1)$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, 1)$
- $\{\Phi(a) = 0.025,\ \Phi(b) = 0.975\} \Rightarrow \{a = z_{0.025},\ b = z_{0.975}\}$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, 1)$
- $\{\Phi(a) = 0.025,\ \Phi(b) = 0.975\} \Rightarrow \{a = z_{0.025},\ b = z_{0.975}\}$
- $\mathrm{P}(-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96) = 0.95$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, 1)$
- $\{\Phi(a) = 0.025,\ \Phi(b) = 0.975\} \Rightarrow \{a = z_{0.025},\ b = z_{0.975}\}$
- $\mathrm{P}(-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96) = 0.95$
- Isolate the parameter:

$$\tfrac{-1.96}{\sqrt{n}} \leq \bar{X} - \mu \leq \tfrac{1.96}{\sqrt{n}}$$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, 1)$
- $\{\Phi(a) = 0.025,\ \Phi(b) = 0.975\} \Rightarrow \{a = z_{0.025},\ b = z_{0.975}\}$
- $\mathrm{P}(-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96) = 0.95$
- Isolate the parameter:

$$\tfrac{-1.96}{\sqrt{n}} \leq \bar{X} - \mu \leq \tfrac{1.96}{\sqrt{n}}$$

$$-\bar{X} - \tfrac{1.96}{\sqrt{n}} \leq -\mu \leq -\bar{X} + \tfrac{1.96}{\sqrt{n}}$$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, 1)$
- $\{\Phi(a) = 0.025,\ \Phi(b) = 0.975\} \Rightarrow \{a = z_{0.025},\ b = z_{0.975}\}$
- $\mathrm{P}(-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96) = 0.95$
- Isolate the parameter:

$$\frac{-1.96}{\sqrt{n}} \leq \bar{X} - \mu \leq \frac{1.96}{\sqrt{n}}$$

$$-\bar{X} - \frac{1.96}{\sqrt{n}} \leq -\mu \leq -\bar{X} + \frac{1.96}{\sqrt{n}}$$

$$\bar{X} - \frac{1.96}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.96}{\sqrt{n}}$$

Example: Normal Confidence Interval, σ^2 Known

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Find a $(1-0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $Z = \sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0,1)$
- $\{\Phi(a) = 0.025,\ \Phi(b) = 0.975\} \Rightarrow \{a = z_{0.025},\ b = z_{0.975}\}$
- $\mathrm{P}(-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96) = 0.95$
- Isolate the parameter:

$$\tfrac{-1.96}{\sqrt{n}} \leq \bar{X} - \mu \leq \tfrac{1.96}{\sqrt{n}}$$

$$-\bar{X} - \tfrac{1.96}{\sqrt{n}} \leq -\mu \leq -\bar{X} + \tfrac{1.96}{\sqrt{n}}$$

$$\bar{X} - \tfrac{1.96}{\sqrt{n}} \leq \mu \leq \bar{X} + \tfrac{1.96}{\sqrt{n}}$$

- $\left[\dfrac{\bar{X} - 1.96}{\sqrt{n}},\ \dfrac{\bar{X} + 1.96}{\sqrt{n}}\right]$ is a 95% confidence interval for μ

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\Longrightarrow \alpha_1 = \alpha_2 = 0.025$
- $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1-0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\Longrightarrow \alpha_1 = \alpha_2 = 0.025$
- $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\mathrm{P}(t_{0.025,n-1} \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{0.975,n-1}) = 0.95$

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\mathrm{P}(t_{0.025,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0.975,n-1}) = 0.95$
- Isolate the parameter:

$$\frac{S}{\sqrt{n}} t_{0.025,n-1} \leq \bar{X} - \mu \leq \frac{S}{\sqrt{n}} t_{0.975,n-1}$$

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\mathrm{P}(t_{0.025,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0.975,n-1}) = 0.95$
- Isolate the parameter:

$$\frac{S}{\sqrt{n}} t_{0.025,n-1} \leq \bar{X} - \mu \leq \frac{S}{\sqrt{n}} t_{0.975,n-1}$$

$$-\bar{X} + \frac{S}{\sqrt{n}} t_{0.025,n-1} \leq -\mu \leq -\bar{X} + \frac{S}{\sqrt{n}} t_{0.975,n-1}$$

Example: Normal Confidence Interval, Variance Unknown

Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Find a $(1 - 0.05)100\%$ symmetric confidence interval for the mean μ

- $\alpha = 0.05$, symmetric $\implies \alpha_1 = \alpha_2 = 0.025$
- $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\mathrm{P}(t_{0.025,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0.975,n-1}) = 0.95$
- Isolate the parameter:

$$\frac{S}{\sqrt{n}} t_{0.025,n-1} \leq \bar{X} - \mu \leq \frac{S}{\sqrt{n}} t_{0.975,n-1}$$

$$-\bar{X} + \frac{S}{\sqrt{n}} t_{0.025,n-1} \leq -\mu \leq -\bar{X} + \frac{S}{\sqrt{n}} t_{0.975,n-1}$$

$$\bar{X} + \frac{S}{\sqrt{n}} t_{0.025,n-1} \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{0.975,n-1}$$

Confidence Interval Interpretation

The parameter θ is unknown but fixed

Confidence Interval Interpretation

The parameter θ is unknown but fixed   It is *NOT* random

Confidence Interval Interpretation

The parameter θ is unknown but fixed   It is *NOT* random

The confidence interval $[I, S]$ is a random interval that covers θ with probability $(1 - \alpha)$

Confidence Interval Interpretation

The parameter θ is unknown but fixed   It is *NOT* random

The confidence interval $[I, S]$ is a random interval that covers θ with probability $(1-\alpha)$

Let A be the event $I \leq \theta$ and let B be the event $S \geq \theta$

$$P(A \cap B) = (1-\alpha)$$

Confidence Interval Interpretation

The parameter θ is unknown but fixed It is *NOT* random

The confidence interval $[I, S]$ is a random interval that covers θ with probability $(1-\alpha)$

Let A be the event $I \leq \theta$ and let B be the event $S \geq \theta$

$$P(A \cap B) = (1-\alpha)$$

Saying that $\theta \in [I, S]$ with probability $(1-\alpha)$ is incorrect because it is I and S that are random (and not θ)

Confidence Interval Interpretation

The parameter θ is unknown but fixed   It is *NOT* random

The confidence interval $[I, S]$ is a random interval that covers θ with probability $(1 - \alpha)$

Let A be the event $I \leq \theta$ and let B be the event $S \geq \theta$

$$P(A \cap B) = (1 - \alpha)$$

Saying that $\theta \in [I, S]$ with probability $(1 - \alpha)$ is incorrect because it is I and S that are random (and not θ)

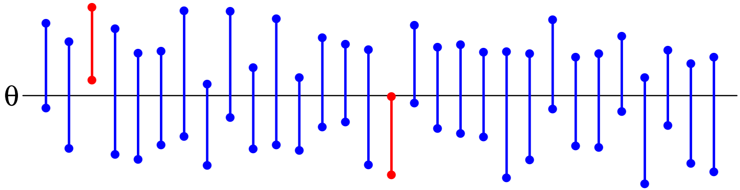


Figure: 30 repetitions of the confidence interval calculation

Confidence Intervals: Remarks

How can we ...

- ... increase the probability that the confidence interval covers θ?

Confidence Intervals: Remarks

How can we ...

- ... increase the probability that the confidence interval covers θ?
  - Make α smaller, then $(1 - \alpha)$ gets bigger making the confidence interval larger

Confidence Intervals: Remarks

How can we ...

- ... increase the probability that the confidence interval covers θ?
  - Make α smaller, then $(1-\alpha)$ gets bigger making the confidence interval larger
- ... make the confidence interval shorter?

Confidence Intervals: Remarks

How can we ...

- ... increase the probability that the confidence interval covers θ?
  - Make α smaller, then $(1 - \alpha)$ gets bigger making the confidence interval larger
- ... make the confidence interval shorter?
  - Confidence intervals for the mean are always of the form

$$\bar{x} \pm \frac{c}{\sqrt{n}}$$

Confidence Intervals: Remarks

How can we ...

- ... increase the probability that the confidence interval covers θ?
  - Make α smaller, then $(1-\alpha)$ gets bigger making the confidence interval larger
- ... make the confidence interval shorter?
  - Confidence intervals for the mean are always of the form

$$\bar{x} \pm \frac{c}{\sqrt{n}}$$

  - If n increases, then the confidence interval gets shorter

Confidence Intervals: Remarks

How can we ...

- ... increase the probability that the confidence interval covers θ?
  - Make α smaller, then $(1-\alpha)$ gets bigger making the confidence interval larger
- ... make the confidence interval shorter?
  - Confidence intervals for the mean are always of the form

$$\bar{x} \pm \frac{c}{\sqrt{n}}$$

  - If n increases, then the confidence interval gets shorter

One-sided confidence intervals

- Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$

Confidence Intervals: Remarks

How can we ...

- ...increase the probability that the confidence interval covers θ?
  - Make α smaller, then $(1-\alpha)$ gets bigger making the confidence interval larger
- ...make the confidence interval shorter?
  - Confidence intervals for the mean are always of the form

$$\bar{x} \pm \frac{c}{\sqrt{n}}$$

  - If n increases, then the confidence interval gets shorter

One-sided confidence intervals

- Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1^2)$
- A confidence interval of the form $\big(-\infty, \bar{X} + z_{(1-\alpha)}/\sqrt{n}\,\big]$ is a one-sided confidence interval for μ

Confidence Intervals for Maximum Likelihood Estimators

Often construct confidence intervals based on maximum likelihood estimators

Confidence Intervals for Maximum Likelihood Estimators

Often construct confidence intervals based on maximum likelihood estimators

Theorem Let $\hat{\theta}_{ML}$ the maximum likelihood estimator of θ, then for large n,

$$\hat{\theta}_{ML} \dot{\sim} \mathcal{N}\left(\theta,\ J(\hat{\theta}_{ML})^{-1}\right)$$

Confidence Intervals for Maximum Likelihood Estimators

Often construct confidence intervals based on maximum likelihood estimators

Theorem Let $\hat{\theta}_{ML}$ the maximum likelihood estimator of θ, then for large n,

$$\hat{\theta}_{ML} \dot{\sim} \mathcal{N}\left(\theta,\, J(\hat{\theta}_{ML})^{-1}\right)$$

$J(\theta) = -\frac{d^2}{d\theta^2}\ell(\theta)$ is the *observed information* for θ

Confidence Intervals for Maximum Likelihood Estimators

Often construct confidence intervals based on maximum likelihood estimators

Theorem Let $\hat{\theta}_{ML}$ the maximum likelihood estimator of θ, then for large n,

$$\hat{\theta}_{ML} \dot{\sim} \mathcal{N}\left(\theta,\ J(\hat{\theta}_{ML})^{-1}\right)$$

$J(\theta) = -\frac{d^2}{d\theta^2}\ell(\theta)$ is the *observed information* for θ

Can make a symmetric $(1-\alpha)100\%$ confidence interval for θ

$$\left[\hat{\theta}_{ML} - z_{1-\frac{\alpha}{2}} J(\hat{\theta}_{ML})^{-\frac{1}{2}},\ \ \hat{\theta}_{ML} + z_{1-\frac{\alpha}{2}} J(\hat{\theta}_{ML})^{-\frac{1}{2}}\right]$$

COMPUTATIONAL FINANCE & RISK MANAGEMENT

UNIVERSITY *of* WASHINGTON

Department of Applied Mathematics

http://computational-finance.uw.edu