Review of Statistics II

Hüseyin Taştan[1]

[1]Department of Economics
Yildiz Technical University

April 17, 2010

- Estimation
- Properties of Estimators
- Finite and asymptotic properties
- Estimation methods
- Hypothesis testing

Review of Statistical Inference Theory

- Estimation
- Properties of Estimators
- Finite and asymptotic properties
- Estimation methods
- Hypothesis testing

Review of Statistical Inference Theory

- Estimation
- Properties of Estimators
- Finite and asymptotic properties
- Estimation methods
- Hypothesis testing

Review of Statistical Inference Theory

- Estimation
- Properties of Estimators
- Finite and asymptotic properties
- Estimation methods
- Hypothesis testing

Review of Statistical Inference Theory

- Estimation
- Properties of Estimators
- Finite and asymptotic properties
- Estimation methods
- Hypothesis testing

Point Estimation

- Purpose: To make inferences about population parameters using the sample information.
- For example suppose that we want to estimate the average disposable income of households in Istanbul. To make such an inference we need to know the sampling distribution of statistics e.g., sample means.
- Inference is based on the sample statistics.
- Two types of estimation: Point vs Interval

Point Estimation

- Purpose: To make inferences about population parameters using the sample information.
- For example suppose that we want to estimate the average disposable income of households in Istanbul. To make such an inference we need to know the sampling distribution of statistics e.g., sample means.
- Inference is based on the sample statistics.
- Two types of estimation: Point vs Interval

Point Estimation

- Purpose: To make inferences about population parameters using the sample information.
- For example suppose that we want to estimate the average disposable income of households in Istanbul. To make such an inference we need to know the sampling distribution of statistics e.g., sample means.
- Inference is based on the sample statistics.
- Two types of estimation: Point vs Interval

Point Estimation

- Purpose: To make inferences about population parameters using the sample information.
- For example suppose that we want to estimate the average disposable income of households in Istanbul. To make such an inference we need to know the sampling distribution of statistics e.g., sample means.
- Inference is based on the sample statistics.
- Two types of estimation: Point vs Interval

Point Estimation

- An estimator for a population parameter is simply a function of sample information. Since it is a function of random variables an estimator itself is a random variable.
- The distribution of an estimator is called sampling distribution.
- A certain realization of an estimator based on a sample observation is called an estimate.
- Suppose we want to estimate the mean income of households in Istanbul. For this purpose we can use sample mean which is an estimator for the population mean.
- As we select a new sample it is almost certain that we will obtain a different estimate.
- But on the average we will obtain a value close to the true population mean.

Point Estimation

- An estimator for a population parameter is simply a function of sample information. Since it is a function of random variables an estimator itself is a random variable.
- The distribution of an estimator is called sampling distribution.
- A certain realization of an estimator based on a sample observation is called an estimate.
- Suppose we want to estimate the mean income of households in Istanbul. For this purpose we can use sample mean which is an estimator for the population mean.
- As we select a new sample it is almost certain that we will obtain a different estimate.
- But on the average we will obtain a value close to the true population mean.

Point Estimation

- An estimator for a population parameter is simply a function of sample information. Since it is a function of random variables an estimator itself is a random variable.
- The distribution of an estimator is called sampling distribution.
- A certain realization of an estimator based on a sample observation is called an estimate.
- Suppose we want to estimate the mean income of households in Istanbul. For this purpose we can use sample mean which is an estimator for the population mean.
- As we select a new sample it is almost certain that we will obtain a different estimate.
- But on the average we will obtain a value close to the true population mean.

Point Estimation

- An estimator for a population parameter is simply a function of sample information. Since it is a function of random variables an estimator itself is a random variable.
- The distribution of an estimator is called sampling distribution.
- A certain realization of an estimator based on a sample observation is called an estimate.
- Suppose we want to estimate the mean income of households in Istanbul. For this purpose we can use sample mean which is an estimator for the population mean.
- As we select a new sample it is almost certain that we will obtain a different estimate.
- But on the average we will obtain a value close to the true population mean.

Point Estimation

- An estimator for a population parameter is simply a function of sample information. Since it is a function of random variables an estimator itself is a random variable.
- The distribution of an estimator is called sampling distribution.
- A certain realization of an estimator based on a sample observation is called an estimate.
- Suppose we want to estimate the mean income of households in Istanbul. For this purpose we can use sample mean which is an estimator for the population mean.
- As we select a new sample it is almost certain that we will obtain a different estimate.
- But on the average we will obtain a value close to the true population mean.

Point Estimation

- An estimator for a population parameter is simply a function of sample information. Since it is a function of random variables an estimator itself is a random variable.
- The distribution of an estimator is called sampling distribution.
- A certain realization of an estimator based on a sample observation is called an estimate.
- Suppose we want to estimate the mean income of households in Istanbul. For this purpose we can use sample mean which is an estimator for the population mean.
- As we select a new sample it is almost certain that we will obtain a different estimate.
- But on the average we will obtain a value close to the true population mean.

Some parameters and estimators

Population parameter	Estimator	Estimate
Mean (μ)	$\overline{X}$	$\bar{x}$
Variance (σ^2)	s_X^2	s_x^2
Standard Deviation (σ)	s_X	s_x
Success rate (p)	$\hat{p}_X$	$\hat{p}_x$

Properties of Estimators

- How do we decide which estimators to use for inference? We require estimators to have certain properties.
- Finite sample properties: these are valid for any sample size n, unbiasedness and efficiency
- Asymptotic properties: only valid for large samples, e.g., consistency, asymptotic efficiency, asymptotic normality

Properties of Estimators

- How do we decide which estimators to use for inference? We require estimators to have certain properties.
- Finite sample properties: these are valid for any sample size n, unbiasedness and efficiency
- Asymptotic properties: only valid for large samples, e.g., consistency, asymptotic efficiency, asymptotic normality

Properties of Estimators

- How do we decide which estimators to use for inference? We require estimators to have certain properties.
- Finite sample properties: these are valid for any sample size n, unbiasedness and efficiency
- Asymptotic properties: only valid for large samples, e.g., consistency, asymptotic efficiency, asymptotic normality

Unbiasedness

θ: Unknown population parameter

$\hat{\theta}$: a point estimator for θ

Definition: If the mean of the sampling distribution of $\hat{\theta}$ is equal to the population parameter θ, i.e.,

$$E(\hat{\theta}) = \theta$$

then we say that $\hat{\theta}$ is an *unbiased estimator* of θ.

Unbiasedness

What does it mean?

- Let us make an hypothetical experiment. Suppose we draw many samples from the population of interest and calculate the estimator for each sample.
- If our estimator is unbiased then the arithmetic mean of our calculated estimates will be equal to the population parameter.
- One can show that the following point estimators are unbiased:

$$E(\overline{X}) = \mu, \quad E(s_X^2) = \sigma^2, \quad E(\hat{p}_X) = p$$

- In other words, sample mean is an unbiased estimator of population mean, sample variance is an unbiased estimator of population variance.

What does it mean?

- Let us make an hypothetical experiment. Suppose we draw many samples from the population of interest and calculate the estimator for each sample.
- If our estimator is unbiased then the arithmetic mean of our calculated estimates will be equal to the population parameter.
- One can show that the following point estimators are unbiased:

$$E(\overline{X}) = \mu, \quad E(s_X^2) = \sigma^2, \quad E(\hat{p}_X) = p$$

- In other words, sample mean is an unbiased estimator of population mean, sample variance is an unbiased estimator of population variance.

Unbiasedness

What does it mean?

- Let us make an hypothetical experiment. Suppose we draw many samples from the population of interest and calculate the estimator for each sample.
- If our estimator is unbiased then the arithmetic mean of our calculated estimates will be equal to the population parameter.
- One can show that the following point estimators are unbiased:

$$E(\overline{X}) = \mu, \quad E(s_X^2) = \sigma^2, \quad E(\hat{p}_X) = p$$

- In other words, sample mean is an unbiased estimator of population mean, sample variance is an unbiased estimator of population variance.

Unbiasedness

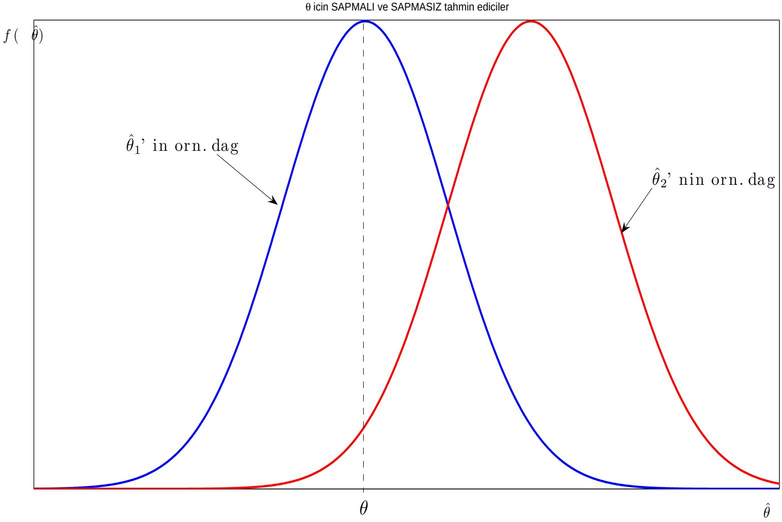
What does it mean?

- Let us make an hypothetical experiment. Suppose we draw many samples from the population of interest and calculate the estimator for each sample.
- If our estimator is unbiased then the arithmetic mean of our calculated estimates will be equal to the population parameter.
- One can show that the following point estimators are unbiased:

$$E(\overline{X}) = \mu, \quad E(s_X^2) = \sigma^2, \quad E(\hat{p}_X) = p$$

- In other words, sample mean is an unbiased estimator of population mean, sample variance is an unbiased estimator of population variance.

Unbiasedness



Unbiasedness

- If an estimator is not unbiased then it is called biased:

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

- For an unbiased estimator we have $Bias(\hat{\theta}) = 0$.
- Unbiasedness does not mean that the value of the estimator is equal to the true parameter value. It just means that on average we would obtain the true parameter.

Unbiasedness

- If an estimator is not unbiased then it is called biased:

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

- For an unbiased estimator we have $Bias(\hat{\theta}) = 0$.
- Unbiasedness does not mean that the value of the estimator is equal to the true parameter value. It just means that on average we would obtain the true parameter.

Unbiasedness

- If an estimator is not unbiased then it is called biased:

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

- For an unbiased estimator we have $Bias(\hat{\theta}) = 0$.
- Unbiasedness does not mean that the value of the estimator is equal to the true parameter value. It just means that on average we would obtain the true parameter.

Efficiency

- Unbiasedness is not sufficient to obtain good estimators
- In general one can define many unbiased estimators for a parameter.
- We can choose from these unbiased estimators based on their variability, i.e., their variance.
- We would prefer an estimator with smaller variability.
- Definition: let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. If

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

then $\hat{\theta}_1$ is a more efficient than $\hat{\theta}_2$.

Efficiency

- Unbiasedness is not sufficient to obtain good estimators
- In general one can define many unbiased estimators for a parameter.
- We can choose from these unbiased estimators based on their variability, i.e., their variance.
- We would prefer an estimator with smaller variability.
- Definition: let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. If

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

then $\hat{\theta}_1$ is a more efficient than $\hat{\theta}_2$.

Efficiency

- Unbiasedness is not sufficient to obtain good estimators
- In general one can define many unbiased estimators for a parameter.
- We can choose from these unbiased estimators based on their variability, i.e., their variance.
- We would prefer an estimator with smaller variability.
- Definition: let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. If

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

then $\hat{\theta}_1$ is a more efficient than $\hat{\theta}_2$.

Efficiency

- Unbiasedness is not sufficient to obtain good estimators
- In general one can define many unbiased estimators for a parameter.
- We can choose from these unbiased estimators based on their variability, i.e., their variance.
- We would prefer an estimator with smaller variability.
- Definition: let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. If

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

then $\hat{\theta}_1$ is a more efficient than $\hat{\theta}_2$.

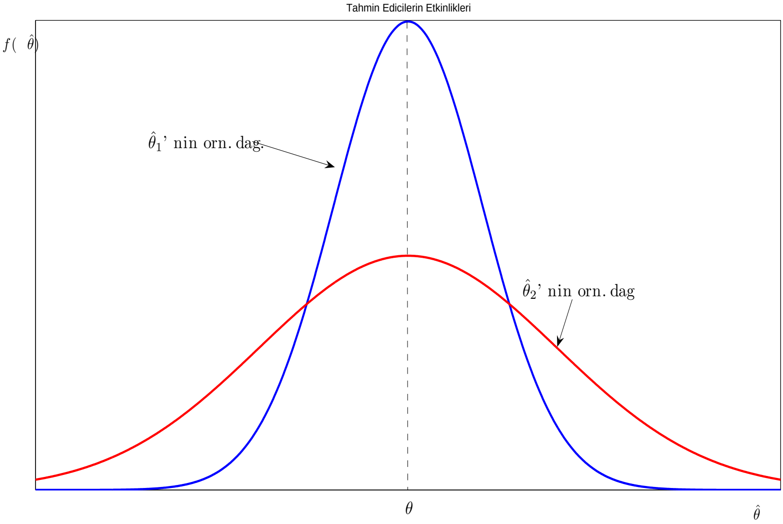
Efficiency

- Unbiasedness is not sufficient to obtain good estimators
- In general one can define many unbiased estimators for a parameter.
- We can choose from these unbiased estimators based on their variability, i.e., their variance.
- We would prefer an estimator with smaller variability.
- Definition: let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. If

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

then $\hat{\theta}_1$ is a more efficient than $\hat{\theta}_2$.

Efficiency



Tahmin Edicilerin Etkinlikleri
$f(\hat{\theta})$
$\hat{\theta}_1$' nin orn. dag.
$\hat{\theta}_2$' nin orn. dag
θ
$\hat{\theta}$

Efficiency - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a population with mean μ and variance σ^2 . Suppose also that there are two estimators for the population mean: $\hat{\theta}_1 = X_1$ and $\hat{\theta}_2 = 10^{-1} \sum X_i$. Are these estimators unbiased? Which one is more efficient?
- Answer: For unbiasedness we must have $E(\hat{\theta}_1) = \mu$. Since $E(\hat{\theta}_1) = E(X_1) = \mu$, $\hat{\theta}_1$ is an unbiased estimator. Similarly, $E(\hat{\theta}_2) = E(10^{-1} \sum X_i) = \mu$, $\hat{\theta}_2$ is also an unbiased estimator.
- For efficiency we need to calculate their variances.

$$Var(\hat{\theta}_1) = Var(X_1) = \sigma^2 \quad Var(\hat{\theta}_2) = Var(10^{-1} \sum X_i) = 0.01\sigma^2$$

- Obviously $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ thus $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$.
- Relative efficiency $= \frac{Var(X_1)}{Var(10^{-1} \sum X_i)} = \frac{\sigma^2}{0.01\sigma^2} = 10$

Efficiency - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a population with mean μ and variance σ^2 . Suppose also that there are two estimators for the population mean: $\hat{\theta}_1 = X_1$ and $\hat{\theta}_2 = 10^{-1} \sum X_i$. Are these estimators unbiased? Which one is more efficient?
- Answer: For unbiasedness we must have $E(\hat{\theta}_1) = \mu$. Since $E(\hat{\theta}_1) = E(X_1) = \mu$, $\hat{\theta}_1$ is an unbiased estimator. Similarly, $E(\hat{\theta}_2) = E(10^{-1} \sum X_i) = \mu$, $\hat{\theta}_2$ is also an unbiased estimator.
- For efficiency we need to calculate their variances.

$$Var(\hat{\theta}_1) = Var(X_1) = \sigma^2 \quad Var(\hat{\theta}_2) = Var(10^{-1} \sum X_i) = 0.01\sigma^2$$

- Obviously $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ thus $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$.
- Relative efficiency $= \frac{Var(X_1)}{Var(10^{-1} \sum X_i)} = \frac{\sigma^2}{0.01\sigma^2} = 10$

Efficiency - Example

- Let $X_1, X_2, \dots, X_{10}$ be a random sample from a population with mean μ and variance σ^2 . Suppose also that there are two estimators for the population mean: $\hat{\theta}_1 = X_1$ and $\hat{\theta}_2 = 10^{-1} \sum X_i$. Are these estimators unbiased? Which one is more efficient?
- Answer: For unbiasedness we must have $E(\hat{\theta}_1) = \mu$. Since $E(\hat{\theta}_1) = E(X_1) = \mu$, $\hat{\theta}_1$ is an unbiased estimator. Similarly, $E(\hat{\theta}_2) = E(10^{-1} \sum X_i) = \mu$, $\hat{\theta}_2$ is also an unbiased estimator.
- For efficiency we need to calculate their variances.

$$Var(\hat{\theta}_1) = Var(X_1) = \sigma^2 \quad Var(\hat{\theta}_2) = Var(10^{-1} \sum X_i) = 0.01\sigma^2$$

- Obviously $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ thus $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$.
- Relative efficiency $= \frac{Var(X_1)}{Var(10^{-1} \sum X_i)} = \frac{\sigma^2}{0.01\sigma^2} = 10$

Efficiency - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a population with mean μ and variance σ^2 . Suppose also that there are two estimators for the population mean: $\hat{\theta}_1 = X_1$ and $\hat{\theta}_2 = 10^{-1} \sum X_i$. Are these estimators unbiased? Which one is more efficient?
- Answer: For unbiasedness we must have $E(\hat{\theta}_1) = \mu$. Since $E(\hat{\theta}_1) = E(X_1) = \mu$, $\hat{\theta}_1$ is an unbiased estimator. Similarly, $E(\hat{\theta}_2) = E(10^{-1} \sum X_i) = \mu$, $\hat{\theta}_2$ is also an unbiased estimator.
- For efficiency we need to calculate their variances.

$$Var(\hat{\theta}_1) = Var(X_1) = \sigma^2 \quad Var(\hat{\theta}_2) = Var(10^{-1} \sum X_i) = 0.01\sigma^2$$

- Obviously $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ thus $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$.
- Relative efficiency $= \frac{Var(X_1)}{Var(10^{-1} \sum X_i)} = \frac{\sigma^2}{0.01\sigma^2} = 10$

Efficiency - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a population with mean μ and variance σ^2 . Suppose also that there are two estimators for the population mean: $\hat{\theta}_1 = X_1$ and $\hat{\theta}_2 = 10^{-1} \sum X_i$. Are these estimators unbiased? Which one is more efficient?
- Answer: For unbiasedness we must have $E(\hat{\theta}_1) = \mu$. Since $E(\hat{\theta}_1) = E(X_1) = \mu$, $\hat{\theta}_1$ is an unbiased estimator. Similarly, $E(\hat{\theta}_2) = E(10^{-1} \sum X_i) = \mu$, $\hat{\theta}_2$ is also an unbiased estimator.
- For efficiency we need to calculate their variances.

$$Var(\hat{\theta}_1) = Var(X_1) = \sigma^2 \quad Var(\hat{\theta}_2) = Var(10^{-1} \sum X_i) = 0.01\sigma^2$$

- Obviously $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ thus $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$.
- Relative efficiency $= \frac{Var(X_1)}{Var(10^{-1} \sum X_i)} = \frac{\sigma^2}{0.01\sigma^2} = 10$

Asymptotic Properties

- In the previous example our first estimator was defined as the value of the first observation. Note that in this case we would not use the full information contained in the sample.
- As we have shown this estimator is unbiased but inefficient. Not matter how large the sample is this variance will not get smaller.
- In most cases we would like our estimators to have good properties as we collect more information, that is as the sample size gets bigger.
- For example, as n becomes large the variance of $\overline{X}$ gets smaller so that it approaches the true value μ at certain rate.
- On the contrary an estimator such as X_1 does not change as n approaches infinity.
- We can eliminate stupid estimators such X_1 by looking at their asymptotic properties.

Asymptotic Properties

- In the previous example our first estimator was defined as the value of the first observation. Note that in this case we would not use the full information contained in the sample.
- As we have shown this estimator is unbiased but inefficient. Not matter how large the sample is this variance will not get smaller.
- In most cases we would like our estimators to have good properties as we collect more information, that is as the sample size gets bigger.
- For example, as n becomes large the variance of $\overline{X}$ gets smaller so that it approaches the true value μ at certain rate.
- On the contrary an estimator such as X_1 does not change as n approaches infinity.
- We can eliminate stupid estimators such X_1 by looking at their asymptotic properties.

Asymptotic Properties

- In the previous example our first estimator was defined as the value of the first observation. Note that in this case we would not use the full information contained in the sample.
- As we have shown this estimator is unbiased but inefficient. Not matter how large the sample is this variance will not get smaller.
- In most cases we would like our estimators to have good properties as we collect more information, that is as the sample size gets bigger.
- For example, as n becomes large the variance of $\overline{X}$ gets smaller so that it approaches the true value μ at certain rate.
- On the contrary an estimator such as X_1 does not change as n approaches infinity.
- We can eliminate stupid estimators such X_1 by looking at their asymptotic properties.

Asymptotic Properties

- In the previous example our first estimator was defined as the value of the first observation. Note that in this case we would not use the full information contained in the sample.
- As we have shown this estimator is unbiased but inefficient. Not matter how large the sample is this variance will not get smaller.
- In most cases we would like our estimators to have good properties as we collect more information, that is as the sample size gets bigger.
- For example, as n becomes large the variance of $\overline{X}$ gets smaller so that it approaches the true value μ at certain rate.
- On the contrary an estimator such as X_1 does not change as n approaches infinity.
- We can eliminate stupid estimators such X_1 by looking at their asymptotic properties.

Asymptotic Properties

- In the previous example our first estimator was defined as the value of the first observation. Note that in this case we would not use the full information contained in the sample.
- As we have shown this estimator is unbiased but inefficient. Not matter how large the sample is this variance will not get smaller.
- In most cases we would like our estimators to have good properties as we collect more information, that is as the sample size gets bigger.
- For example, as n becomes large the variance of $\overline{X}$ gets smaller so that it approaches the true value μ at certain rate.
- On the contrary an estimator such as X_1 does not change as n approaches infinity.
- We can eliminate stupid estimators such X_1 by looking at their asymptotic properties.

Asymptotic Properties

- In the previous example our first estimator was defined as the value of the first observation. Note that in this case we would not use the full information contained in the sample.
- As we have shown this estimator is unbiased but inefficient. Not matter how large the sample is this variance will not get smaller.
- In most cases we would like our estimators to have good properties as we collect more information, that is as the sample size gets bigger.
- For example, as n becomes large the variance of $\overline{X}$ gets smaller so that it approaches the true value μ at certain rate.
- On the contrary an estimator such as X_1 does not change as n approaches infinity.
- We can eliminate stupid estimators such X_1 by looking at their asymptotic properties.

Asymptotic Properties

- Another reason for analyzing asymptotic properties is that in most cases we may not know the finite sample properties of estimators. For several estimators in statistics it is hard to verify the finite sample distributions.
- In this case, it would not be possible to compare estimators based on unbiasedness and efficiency.
- However, it may be relatively easy to obtain the behavior of the estimator as n gets bigger and bigger.
- Asymptotic Properties: consistency, asymptotic efficiency

Asymptotic Properties

- Another reason for analyzing asymptotic properties is that in most cases we may not know the finite sample properties of estimators. For several estimators in statistics it is hard to verify the finite sample distributions.
- In this case, it would not be possible to compare estimators based on unbiasedness and efficiency.
- However, it may be relatively easy to obtain the behavior of the estimator as n gets bigger and bigger.
- Asymptotic Properties: consistency, asymptotic efficiency

Asymptotic Properties

- Another reason for analyzing asymptotic properties is that in most cases we may not know the finite sample properties of estimators. For several estimators in statistics it is hard to verify the finite sample distributions.
- In this case, it would not be possible to compare estimators based on unbiasedness and efficiency.
- However, it may be relatively easy to obtain the behavior of the estimator as n gets bigger and bigger.
- Asymptotic Properties: consistency, asymptotic efficiency

Asymptotic Properties

- Another reason for analyzing asymptotic properties is that in most cases we may not know the finite sample properties of estimators. For several estimators in statistics it is hard to verify the finite sample distributions.
- In this case, it would not be possible to compare estimators based on unbiasedness and efficiency.
- However, it may be relatively easy to obtain the behavior of the estimator as n gets bigger and bigger.
- Asymptotic Properties: consistency, asymptotic efficiency

Consistency

- For a small arbitrary number $\epsilon > 0$

$$\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| < \epsilon\right] = 1 \quad \text{or} \quad \lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| > \epsilon\right] = 0$$

  then we say that $\hat{\theta}_n$ is a consistent estimator for θ

- When this condition is satisfied θ is the probability limit of $\hat{\theta}_n$ and written more compactly as:

$$\text{plim}(\hat{\theta}_n) = \theta$$

- This means: as the sample size gets bigger the sampling distribution of $\hat{\theta}_n$ becomes more concentrated around the true parameter value θ. In the limit it becomes degenerate at the true value.
- The probability of getting away from the true value becomes smaller.

Consistency

- For a small arbitrary number $\epsilon > 0$

$$\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| < \epsilon\right] = 1 \quad \text{or} \quad \lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| > \epsilon\right] = 0$$

  then we say that $\hat{\theta}_n$ is a consistent estimator for θ
- When this condition is satisfied θ is the probability limit of $\hat{\theta}_n$ and written more compactly as:

$$\text{plim}(\hat{\theta}_n) = \theta$$

- This means: as the sample size gets bigger the sampling distribution of $\hat{\theta}_n$ becomes more concentrated around the true parameter value θ. In the limit it becomes degenerate at the true value.
- The probability of getting away from the true value becomes smaller.

Consistency

- For a small arbitrary number $\epsilon > 0$

$$\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| < \epsilon\right] = 1 \quad \text{or} \quad \lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| > \epsilon\right] = 0$$

  then we say that $\hat{\theta}_n$ is a consistent estimator for θ
- When this condition is satisfied θ is the probability limit of $\hat{\theta}_n$ and written more compactly as:

$$\text{plim}(\hat{\theta}_n) = \theta$$

- This means: as the sample size gets bigger the sampling distribution of $\hat{\theta}_n$ becomes more concentrated around the true parameter value θ. In the limit it becomes degenerate at the true value.
- The probability of getting away from the true value becomes smaller.

Consistency

- For a small arbitrary number $\epsilon > 0$

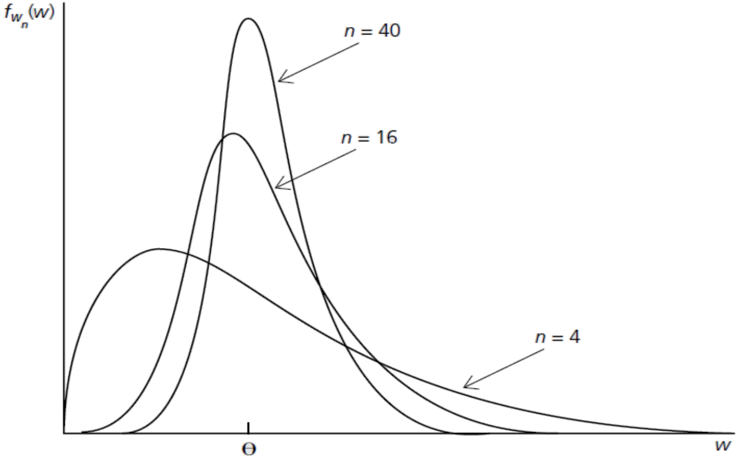
$$\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| < \epsilon\right] = 1 \quad \text{or} \quad \lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| > \epsilon\right] = 0$$

  then we say that $\hat{\theta}_n$ is a consistent estimator for θ
- When this condition is satisfied θ is the probability limit of $\hat{\theta}_n$ and written more compactly as:

$$\text{plim}(\hat{\theta}_n) = \theta$$

- This means: as the sample size gets bigger the sampling distribution of $\hat{\theta}_n$ becomes more concentrated around the true parameter value θ. In the limit it becomes degenerate at the true value.
- The probability of getting away from the true value becomes smaller.

Consistency - Sampling Distributions of a Consistent Estimator



Consistency

- Law of Large Numbers: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with μ and variance σ^2. According to LLN:

$$\text{plim}(\overline{X}_n) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

- Since $Var(\overline{X}_n)$ converges to 0 in the limit sample mean approaches the true value μ. Collecting more information leads to smaller and smaller variability.
- In general unbiased estimators with variances converging to zero as n gets bigger are consistent.
- An inconsistent estimator cannot provide information on population parameters.
- An unbiased estimator can be inconsistent and vice versa.

Consistency

- Law of Large Numbers: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with μ and variance σ^2. According to LLN:

$$\text{plim}(\overline{X}_n) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

- Since $Var(\overline{X}_n)$ converges to 0 in the limit sample mean approaches the true value μ. Collecting more information leads to smaller and smaller variability.
- In general unbiased estimators with variances converging to zero as n gets bigger are consistent.
- An inconsistent estimator cannot provide information on population parameters.
- An unbiased estimator can be inconsistent and vice versa.

Consistency

- Law of Large Numbers: Let $X_1, X_2, \dots, X_n$ be a random sample from a population with μ and variance σ^2. According to LLN:

$$\text{plim}(\overline{X}_n) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

- Since $Var(\overline{X}_n)$ converges to 0 in the limit sample mean approaches the true value μ. Collecting more information leads to smaller and smaller variability.
- In general unbiased estimators with variances converging to zero as n gets bigger are consistent.
- An inconsistent estimator cannot provide information on population parameters.
- An unbiased estimator can be inconsistent and vice versa.

Consistency

- Law of Large Numbers: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with μ and variance σ^2. According to LLN:

$$\text{plim}(\overline{X}_n) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

- Since $Var(\overline{X}_n)$ converges to 0 in the limit sample mean approaches the true value μ. Collecting more information leads to smaller and smaller variability.
- In general unbiased estimators with variances converging to zero as n gets bigger are consistent.
- An inconsistent estimator cannot provide information on population parameters.
- An unbiased estimator can be inconsistent and vice versa.

Consistency

- Law of Large Numbers: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with μ and variance σ^2. According to LLN:

$$\text{plim}(\overline{X}_n) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

- Since $Var(\overline{X}_n)$ converges to 0 in the limit sample mean approaches the true value μ. Collecting more information leads to smaller and smaller variability.
- In general unbiased estimators with variances converging to zero as n gets bigger are consistent.
- An inconsistent estimator cannot provide information on population parameters.
- An unbiased estimator can be inconsistent and vice versa.

Consistency

- The following two conditions are sufficient for an estimator to be consistent:

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

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

- According to the first condition as the sample size increases expected value becomes closer to the true value. In other words as n goes to infinity bias converges to zero.
- According to the second condition, as the sample size increases the variability of the sampling distribution around the true value becomes smaller and smaller and equals 0 in the limit.

Consistency

- The following two conditions are sufficient for an estimator to be consistent:

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

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

- According to the first condition as the sample size increases expected value becomes closer to the true value. In other words as n goes to infinity bias converges to zero.
- According to the second condition, as the sample size increases the variability of the sampling distribution around the true value becomes smaller and smaller and equals 0 in the limit.

Consistency

- The following two conditions are sufficient for an estimator to be consistent:

$$1. \qquad \lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

$$2. \qquad \lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

- According to the first condition as the sample size increases expected value becomes closer to the true value. In other words as n goes to infinity bias converges to zero.
- According to the second condition, as the sample size increases the variability of the sampling distribution around the true value becomes smaller and smaller and equals 0 in the limit.

Consistency - Example

- Besides the sample mean $\overline{X}_n$ the following estimators are defined based on a random sample of size n

  1. $$\hat{\mu}_n^1 = \frac{1}{n+1}\sum_{i=1}^{n} X_t,$$
  2. $$\hat{\mu}_n^2 = \frac{1.02}{n}\sum_{i=1}^{n} X_t,$$
  3. $$\hat{\mu}_n^3 = 0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t$$

- Determine if these estimators are biased and/or consistent.

Consistency - Example

- Besides the sample mean $\overline{X}_n$ the following estimators are defined based on a random sample of size n
  1. $$\hat{\mu}_n^1 = \frac{1}{n+1}\sum_{i=1}^{n} X_t,$$
  2. $$\hat{\mu}_n^2 = \frac{1.02}{n}\sum_{i=1}^{n} X_t,$$
  3. $$\hat{\mu}_n^3 = 0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t$$
- Determine if these estimators are biased and/or consistent.

Consistency - Example

- Besides the sample mean $\overline{X}_n$ the following estimators are defined based on a random sample of size n

1. $$\hat{\mu}_n^1 = \frac{1}{n+1}\sum_{i=1}^{n} X_t,$$

2. $$\hat{\mu}_n^2 = \frac{1.02}{n}\sum_{i=1}^{n} X_t,$$

3. $$\hat{\mu}_n^3 = 0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t$$

- Determine if these estimators are biased and/or consistent.

Consistency - Example

- Besides the sample mean $\overline{X}_n$ the following estimators are defined based on a random sample of size n

  1. $$\hat{\mu}_n^1 = \frac{1}{n+1}\sum_{i=1}^{n} X_t,$$

  2. $$\hat{\mu}_n^2 = \frac{1.02}{n}\sum_{i=1}^{n} X_t,$$

  3. $$\hat{\mu}_n^3 = 0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t$$

- Determine if these estimators are biased and/or consistent.

Consistency - Example

- Besides the sample mean $\overline{X}_n$ the following estimators are defined based on a random sample of size n

  1. $$\hat{\mu}_n^1 = \frac{1}{n+1}\sum_{i=1}^{n} X_t,$$

  2. $$\hat{\mu}_n^2 = \frac{1.02}{n}\sum_{i=1}^{n} X_t,$$

  3. $$\hat{\mu}_n^3 = 0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t$$

- Determine if these estimators are biased and/or consistent.

Consistency - Example

- The first estimator is biased but consistent:

$$E(\hat{\mu}_n^1) = E\left(\frac{1}{n+1}\sum_{i=1}^{n} X_t\right) = \frac{n}{n+1}\mu \Longrightarrow \text{biased}$$

$$\begin{aligned} Var(\hat{\mu}_n^1) &= Var\left(\frac{1}{n+1}\sum_{i=1}^{n} X_t\right) = \frac{n^2}{(n+1)^2} Var(\overline{X}_n) \\ &= \frac{n^2}{(n+1)^2}\frac{\sigma^2}{n} = \frac{n}{(n+1)^2}\sigma^2 \end{aligned}$$

-

$$n \longrightarrow \infty, \quad \frac{n}{n+1}\mu \longrightarrow \mu, \quad \text{ve} \quad \frac{n}{(n+1)^2}\sigma^2 \longrightarrow 0 \Longrightarrow \text{consistent}$$

Note that $\hat{\mu}_n^1 = \frac{n}{n+1}\overline{X}_n$

Consistency - Example

- The first estimator is biased but consistent:

$$E(\hat{\mu}_n^1) = E\left(\frac{1}{n+1}\sum_{i=1}^{n} X_t\right) = \frac{n}{n+1}\mu \Longrightarrow \text{biased}$$

$$\begin{aligned} Var(\hat{\mu}_n^1) &= Var\left(\frac{1}{n+1}\sum_{i=1}^{n} X_t\right) = \frac{n^2}{(n+1)^2} Var(\overline{X}_n) \\ &= \frac{n^2}{(n+1)^2}\frac{\sigma^2}{n} = \frac{n}{(n+1)^2}\sigma^2 \end{aligned}$$

- $$n \longrightarrow \infty, \quad \frac{n}{n+1}\mu \longrightarrow \mu, \quad \text{ve} \quad \frac{n}{(n+1)^2}\sigma^2 \longrightarrow 0 \Longrightarrow \text{consistent}$$

Note that $\hat{\mu}_n^1 = \frac{n}{n+1}\overline{X}_n$

- Second estimator is biased and inconsistent:

$$E(\hat{\mu}_n^2) = E\left(\frac{1.01}{n}\sum_{i=1}^{n} X_t\right) = 1.01E(\overline{X}_n) = 1.01\mu \Longrightarrow \text{biased}$$

$$n \longrightarrow \infty, \quad 1.01\overline{X}_n \longrightarrow 1.01\mu \Longrightarrow \text{inconsistent}$$

- Third estimator is unbiased and inconsistent:

$$E(\hat{\mu}_n^3) = 0.01E(X_1) + \frac{0.99}{n-1}E\left(\sum_{i=2}^{n} X_t\right) = 0.01\mu + \frac{0.99}{n-1}(n-1)\mu$$

$$= \mu \Longrightarrow \text{unbiased}$$

$$n \longrightarrow \infty, \quad E\left(0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t\right) \longrightarrow 0.01X_1 + 0.99\mu$$

$$\Longrightarrow \text{inconsistent}$$

- Second estimator is biased and inconsistent:

$$E(\hat{\mu}_n^2) = E\left(\frac{1.01}{n}\sum_{i=1}^{n} X_t\right) = 1.01E(\overline{X}_n) = 1.01\mu \Longrightarrow \text{biased}$$

$$n \longrightarrow \infty, \quad 1.01\overline{X}_n \longrightarrow 1.01\mu \Longrightarrow \text{inconsistent}$$

- Third estimator is unbiased and inconsistent:

$$E(\hat{\mu}_n^3) = 0.01E(X_1) + \frac{0.99}{n-1}E\left(\sum_{i=2}^{n} X_t\right) = 0.01\mu + \frac{0.99}{n-1}(n-1)\mu$$

$$= \mu \Longrightarrow \text{unbiased}$$

$$n \longrightarrow \infty, \quad E\left(0.01X_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} X_t\right) \longrightarrow 0.01X_1 + 0.99\mu$$

$$\Longrightarrow \text{inconsistent}$$

- Recall the definition of probability limit:

$$\text{plim}(Y_n) = \alpha$$

- The most important property of plim is the following: for any function $g(Y_n)$ of Y_n

$$\text{plim} g((Y_n)) = g(\text{plim}(Y_n)) = g(\alpha)$$

- For example let $\overline{X}_n > 0$, $g(\overline{X}_n) = \ln(\overline{X}_n)$ .We know that expectation operator does not apply to nonlinear functions:

$$E(\ln(\overline{X}_n)) \neq \ln(E(\overline{X}_n))$$

- But using plim we can write

$$\text{plim}(\ln(\overline{X}_n)) = \ln(\text{plim}(\overline{X}_n)) = \ln(\mu)$$

Consistency - Properties of PLIM

- Recall the definition of probability limit:

$$\text{plim}(Y_n) = \alpha$$

- The most important property of plim is the following: for any function $g(Y_n)$ of Y_n

$$\text{plim}g((Y_n)) = g(\text{plim}(Y_n)) = g(\alpha)$$

- For example let $\overline{X}_n > 0$, $g(\overline{X}_n) = \ln(\overline{X}_n)$ .We know that expectation operator does not apply to nonlinear functions:

$$E(\ln(\overline{X}_n)) \neq \ln(E(\overline{X}_n))$$

- But using plim we can write

$$\text{plim}(\ln(\overline{X}_n)) = \ln(\text{plim}(\overline{X}_n)) = \ln(\mu)$$

- Recall the definition of probability limit:

$$\text{plim}(Y_n) = \alpha$$

- The most important property of plim is the following: for any function $g(Y_n)$ of Y_n

$$\text{plim}g((Y_n)) = g(\text{plim}(Y_n)) = g(\alpha)$$

- For example let $\overline{X}_n > 0$, $g(\overline{X}_n) = \ln(\overline{X}_n)$ .We know that expectation operator does not apply to nonlinear functions:

$$E(\ln(\overline{X}_n)) \neq \ln(E(\overline{X}_n))$$

- But using plim we can write

$$\text{plim}(\ln(\overline{X}_n)) = \ln(\text{plim}(\overline{X}_n)) = \ln(\mu)$$

Consistency - Properties of PLIM

- Recall the definition of probability limit:

$$\text{plim}(Y_n) = \alpha$$

- The most important property of plim is the following: for any function $g(Y_n)$ of Y_n

$$\text{plim}g((Y_n)) = g(\text{plim}(Y_n)) = g(\alpha)$$

- For example let $\overline{X}_n > 0$, $g(\overline{X}_n) = \ln(\overline{X}_n)$ .We know that expectation operator does not apply to nonlinear functions:

$$E(\ln(\overline{X}_n)) \neq \ln(E(\overline{X}_n))$$

- But using plim we can write

$$\text{plim}(\ln(\overline{X}_n)) = \ln(\text{plim}(\overline{X}_n)) = \ln(\mu)$$

Properties of PLIM

- Let Y_n and W_n be two estimators with plim$(Y_n) = \alpha$ and plim$(W_n) = \beta$.
- plim$(Y_n + W_n) =$ plim$(Y_n) +$ plim$(W_n) = \alpha + \beta$
- plim$(Y_n W_n) =$ plim(Y_n)plim$(W_n) = \alpha\beta$
- plim$(Y_n / W_n) =$ plim$(Y_n)/$plim$(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \dots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \dots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with plim$(Y_n) = \alpha$ and plim$(W_n) = \beta$.
- plim$(Y_n + W_n) =$ plim$(Y_n) +$ plim$(W_n) = \alpha + \beta$
- plim$(Y_n W_n) =$ plim(Y_n)plim$(W_n) = \alpha\beta$
- plim$(Y_n / W_n) =$ plim$(Y_n)/$plim$(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with plim$(Y_n) = \alpha$ and plim$(W_n) = \beta$.
- plim$(Y_n + W_n) =$ plim$(Y_n) +$ plim$(W_n) = \alpha + \beta$
- plim$(Y_n W_n) =$ plim(Y_n)plim$(W_n) = \alpha\beta$
- plim$(Y_n / W_n) =$ plim$(Y_n)/$plim$(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with plim$(Y_n) = \alpha$ and plim$(W_n) = \beta$.
- plim$(Y_n + W_n) =$ plim$(Y_n) +$ plim$(W_n) = \alpha + \beta$
- plim$(Y_n W_n) =$ plim(Y_n)plim$(W_n) = \alpha\beta$
- plim$(Y_n/W_n) =$ plim$(Y_n)/$plim$(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with $\text{plim}(Y_n) = \alpha$ and $\text{plim}(W_n) = \beta$.
- $\text{plim}(Y_n + W_n) = \text{plim}(Y_n) + \text{plim}(W_n) = \alpha + \beta$
- $\text{plim}(Y_n W_n) = \text{plim}(Y_n)\text{plim}(W_n) = \alpha\beta$
- $\text{plim}(Y_n / W_n) = \text{plim}(Y_n)/\text{plim}(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with plim$(Y_n) = \alpha$ and plim$(W_n) = \beta$.
- plim$(Y_n + W_n) =$ plim$(Y_n) +$ plim$(W_n) = \alpha + \beta$
- plim$(Y_n W_n) =$ plim(Y_n)plim$(W_n) = \alpha\beta$
- plim$(Y_n / W_n) =$ plim$(Y_n)/$plim$(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with $\text{plim}(Y_n) = \alpha$ and $\text{plim}(W_n) = \beta$.
- $\text{plim}(Y_n + W_n) = \text{plim}(Y_n) + \text{plim}(W_n) = \alpha + \beta$
- $\text{plim}(Y_n W_n) = \text{plim}(Y_n)\text{plim}(W_n) = \alpha\beta$
- $\text{plim}(Y_n / W_n) = \text{plim}(Y_n)/\text{plim}(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with $\text{plim}(Y_n) = \alpha$ and $\text{plim}(W_n) = \beta$.
- $\text{plim}(Y_n + W_n) = \text{plim}(Y_n) + \text{plim}(W_n) = \alpha + \beta$
- $\text{plim}(Y_n W_n) = \text{plim}(Y_n)\text{plim}(W_n) = \alpha\beta$
- $\text{plim}(Y_n / W_n) = \text{plim}(Y_n)/\text{plim}(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \ldots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \ldots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Properties of PLIM

- Let Y_n and W_n be two estimators with $\text{plim}(Y_n) = \alpha$ and $\text{plim}(W_n) = \beta$.
- $\text{plim}(Y_n + W_n) = \text{plim}(Y_n) + \text{plim}(W_n) = \alpha + \beta$
- $\text{plim}(Y_n W_n) = \text{plim}(Y_n)\text{plim}(W_n) = \alpha\beta$
- $\text{plim}(Y_n / W_n) = \text{plim}(Y_n)/\text{plim}(W_n) = \alpha/\beta$
- Using these properties we can define new consistent estimators.
- For example let $Y_1, Y_2, \dots, Y_n$ be observations on annual income for high school graduates. Let the population mean be μ_Y for the high school graduates.
- Similarly let $Z_1, Z_2, \dots, Z_n$ be a random sample for annual earnings of university graduates and let the population mean for this group be μ_Z
- Suppose we want to estimate the percentage difference in earnings $\alpha = (\mu_Z - \mu_Y)/\mu_Y$ between the two groups.
- Since $\overline{Y_n}$ is consistent for μ_Y and $\overline{Z_n}$ is consistent for μ_Z, $(\overline{Z_n} - \overline{Y_n})/\overline{Y_n}$ is a consistent estimator for α.

Asymptotic Normality

- Consistency does not say anything about the distribution around the true value. In most cases we want to determine the distribution of the estimator in the limit.
- Most estimators we encounter in the social sciences have normal limit distributions. In other words their sampling distributions converge to normal as the sample size approaches infinity.
- This is called *asymptotic normality*.
- Let $\{Z_n: \ n = 1, 2, \ldots, n\}$ be sequence of random variables. Also let $\Phi(z)$ be cdf of standard normal distribution. If for every z the following condition is satisfied

$$n \longrightarrow \infty \quad P(Z \leq z) \longrightarrow \Phi(z)$$

  then we say that Z_n has an asymptotic normal distribution.
- More compactly:

$$Z_n \sim N(0, 1) \quad \text{or} \quad Z_n \longrightarrow N(0, 1)$$

Asymptotic Normality

- Consistency does not say anything about the distribution around the true value. In most cases we want to determine the distribution of the estimator in the limit.
- Most estimators we encounter in the social sciences have normal limit distributions. In other words their sampling distributions converge to normal as the sample size approaches infinity.
- This is called *asymptotic normality*.
- Let $\{Z_n : \ n = 1, 2, \ldots, n\}$ be sequence of random variables. Also let $\Phi(z)$ be cdf of standard normal distribution. If for every z the following condition is satisfied

$$n \longrightarrow \infty \quad P(Z \leq z) \longrightarrow \Phi(z)$$

then we say that Z_n has an asymptotic normal distribution.
- More compactly:

$$Z_n \sim N(0, 1) \quad \text{or} \quad Z_n \longrightarrow N(0, 1)$$

Asymptotic Normality

- Consistency does not say anything about the distribution around the true value. In most cases we want to determine the distribution of the estimator in the limit.
- Most estimators we encounter in the social sciences have normal limit distributions. In other words their sampling distributions converge to normal as the sample size approaches infinity.
- This is called *asymptotic normality*.
- Let $\{Z_n : \ n = 1, 2, \ldots, n\}$ be sequence of random variables. Also let $\Phi(z)$ be cdf of standard normal distribution. If for every z the following condition is satisfied

$$n \longrightarrow \infty \quad P(Z \leq z) \longrightarrow \Phi(z)$$

  then we say that Z_n has an asymptotic normal distribution.
- More compactly:

$$Z_n \sim N(0, 1) \quad \text{or} \quad Z_n \longrightarrow N(0, 1)$$

Asymptotic Normality

- Consistency does not say anything about the distribution around the true value. In most cases we want to determine the distribution of the estimator in the limit.
- Most estimators we encounter in the social sciences have normal limit distributions. In other words their sampling distributions converge to normal as the sample size approaches infinity.
- This is called *asymptotic normality*.
- Let $\{Z_n : \ n = 1, 2, \ldots, n\}$ be sequence of random variables. Also let $\Phi(z)$ be cdf of standard normal distribution. If for every z the following condition is satisfied

$$n \longrightarrow \infty \quad P(Z \leq z) \longrightarrow \Phi(z)$$

  then we say that Z_n has an asymptotic normal distribution.
- More compactly:

$$Z_n \sim N(0,1) \quad \text{or} \quad Z_n \longrightarrow N(0,1)$$

Asymptotic Normality

- Consistency does not say anything about the distribution around the true value. In most cases we want to determine the distribution of the estimator in the limit.
- Most estimators we encounter in the social sciences have normal limit distributions. In other words their sampling distributions converge to normal as the sample size approaches infinity.
- This is called *asymptotic normality*.
- Let $\{Z_n : \; n = 1, 2, \ldots, n\}$ be sequence of random variables. Also let $\Phi(z)$ be cdf of standard normal distribution. If for every z the following condition is satisfied

$$n \longrightarrow \infty \quad P(Z \leq z) \longrightarrow \Phi(z)$$

  then we say that Z_n has an asymptotic normal distribution.
- More compactly:

$$Z_n \sim N(0,1) \quad \text{or} \quad Z_n \longrightarrow N(0,1)$$

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Estimation Methods

- General ways of obtaining point estimators for unknown population parameters
- There are several estimation procedures available.
- Classical estimation methods are
- Method of Moments
- Maximum Likelihood
- Least Squares
- Other methods: minimum chi-squared estimators, generalized method of moments, simulation based estimators, etc.

Method of Moments - MoM

- One of the oldest estimation methods
- Relies on matching population moments to sample moments
- Suppose we have k population parameters which we want to estimate: $\theta_1, \theta_1, \ldots, \theta_k$
- MoM estimators are found by solving the following system

$$\begin{aligned} E(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ E(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \vdots &= \vdots \\ E(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k \end{aligned}$$

- Note that this system has k equations in k unknowns.

Method of Moments - MoM

- One of the oldest estimation methods
- Relies on matching population moments to sample moments
- Suppose we have k population parameters which we want to estimate: $\theta_1, \theta_1, \ldots, \theta_k$
- MoM estimators are found by solving the following system

$$
\begin{aligned}
E(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
E(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
\vdots &= \vdots \\
E(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k
\end{aligned}
$$

- Note that this system has k equations in k unknowns.

Method of Moments - MoM

- One of the oldest estimation methods
- Relies on matching population moments to sample moments
- Suppose we have k population parameters which we want to estimate: $\theta_1, \theta_1, \ldots, \theta_k$
- MoM estimators are found by solving the following system

$$
\begin{aligned}
E(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
E(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
\vdots &= \vdots \\
E(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k
\end{aligned}
$$

- Note that this system has k equations in k unknowns.

Method of Moments - MoM

- One of the oldest estimation methods
- Relies on matching population moments to sample moments
- Suppose we have k population parameters which we want to estimate: $\theta_1, \theta_1, \ldots, \theta_k$
- MoM estimators are found by solving the following system

$$\begin{aligned} E(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ E(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \vdots &= \vdots \\ E(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k \end{aligned}$$

- Note that this system has k equations in k unknowns.

Method of Moments - MoM

- One of the oldest estimation methods
- Relies on matching population moments to sample moments
- Suppose we have k population parameters which we want to estimate: $\theta_1, \theta_1, \ldots, \theta_k$
- MoM estimators are found by solving the following system

$$\begin{aligned} E(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ E(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \vdots &= \vdots \\ E(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k \end{aligned}$$

- Note that this system has k equations in k unknowns.

MoM

Population and sample moments

k	Population Moments	Sample Moments
1	$\mu_1 = E(X)$	$\hat{\mu}_1 = n^{-1} \sum X_i$
2	$\mu_2 = E(X^2)$	$\hat{\mu}_2 = n^{-1} \sum X_i^2$
3	$\mu_3 = E(X^3)$	$\hat{\mu}_3 = n^{-1} \sum X_i^3$
4	$\mu_4 = E(X^4)$	$\hat{\mu}_4 = n^{-1} \sum X_i^4$
$\vdots$	$\vdots$	$\vdots$
k	$\mu_k = E(X^k)$	$\hat{\mu}_k = n^{-1} \sum X_i^k$

MoM - Example

- Using a random sample of size n drawn from $N(\mu, \sigma^2)$ find the MoM estimators for μ and σ^2.
- Note that there are two unknown parameters. Thus we need two moment conditions.
- Using the first two moments we have

$$\begin{aligned} E(X) &= \mu = \overline{X} \\ E(X^2) &= \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{aligned}$$

- Solving this system we obtain

$$\begin{aligned} \hat{\mu}_{mom} &= \overline{X} \\ \hat{\sigma}^2_{mom} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

- Using a random sample of size n drawn from $N(\mu, \sigma^2)$ find the MoM estimators for μ and σ^2.
- Note that there are two unknown parameters. Thus we need two moment conditions.
- Using the first two moments we have

$$
\begin{aligned}
E(X) &= \mu = \overline{X} \\
E(X^2) &= \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2
\end{aligned}
$$

- Solving this system we obtain

$$
\begin{aligned}
\hat{\mu}_{mom} &= \overline{X} \\
\hat{\sigma}^2_{mom} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2
\end{aligned}
$$

MoM - Example

- Using a random sample of size n drawn from $N(\mu, \sigma^2)$ find the MoM estimators for μ and σ^2.
- Note that there are two unknown parameters. Thus we need two moment conditions.
- Using the first two moments we have

$$\begin{aligned} E(X) &= \mu = \overline{X} \\ E(X^2) &= \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{aligned}$$

- Solving this system we obtain

$$\begin{aligned} \hat{\mu}_{mom} &= \overline{X} \\ \hat{\sigma}^2_{mom} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

MoM - Example

- Using a random sample of size n drawn from $N(\mu, \sigma^2)$ find the MoM estimators for μ and σ^2.
- Note that there are two unknown parameters. Thus we need two moment conditions.
- Using the first two moments we have

$$\begin{aligned} E(X) &= \mu = \overline{X} \\ E(X^2) &= \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{aligned}$$

- Solving this system we obtain

$$\begin{aligned} \hat{\mu}_{mom} &= \overline{X} \\ \hat{\sigma}^2_{mom} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

Maximum Likelihood Estimation - MLE

- Frequently used in practice to estimate fully specified parametric models
- Good asymptotic properties, consistent and asymptotically most efficient
- The distribution of the population (pdf and cdf) has to be known. Strong distributional assumptions may be a weakness.
- MLE finds the estimators by maximizing the likelihood function which is a different interpretation of the joint density of the sample.

Maximum Likelihood Estimation - MLE

- Frequently used in practice to estimate fully specified parametric models
- Good asymptotic properties, consistent and asymptotically most efficient
- The distribution of the population (pdf and cdf) has to be known. Strong distributional assumptions may be a weakness.
- MLE finds the estimators by maximizing the likelihood function which is a different interpretation of the joint density of the sample.

Maximum Likelihood Estimation - MLE

- Frequently used in practice to estimate fully specified parametric models
- Good asymptotic properties, consistent and asymptotically most efficient
- The distribution of the population (pdf and cdf) has to be known. Strong distributional assumptions may be a weakness.
- MLE finds the estimators by maximizing the likelihood function which is a different interpretation of the joint density of the sample.

Maximum Likelihood Estimation - MLE

- Frequently used in practice to estimate fully specified parametric models
- Good asymptotic properties, consistent and asymptotically most efficient
- The distribution of the population (pdf and cdf) has to be known. Strong distributional assumptions may be a weakness.
- MLE finds the estimators by maximizing the likelihood function which is a different interpretation of the joint density of the sample.

Maximum Likelihood Estimation - MLE

- Let $f(x;\theta)$ represent the distribution (in most cases pdf). Also let $X_1, X_2, \ldots, X_n$ be a random sample of n observations and let $x_1, x_2, \ldots, x_n$ be realizations of these random variables.
- Since we have a random sample the joint pdf can be written as follows

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n; \theta) &= f_1(x_1;\theta) \cdot f_2(x_2;\theta) \cdot, \ldots, \cdot f_n(x_n;\theta) \\
&= \prod_{i=1}^{n} f(x_i;\theta), \quad i = 1, 2, \ldots, n
\end{aligned}
$$

- Now suppose that we observe the random sample but we do not know the population parameters. Likelihood function is simply a function of θ given the realization of the random sample $\mathbf{x}$:

$$L(\theta \mid x_1, x_2, \ldots, x_n) = L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta), \quad i = 1, 2, \ldots, n$$

Maximum Likelihood Estimation - MLE

- Let $f(x;\theta)$ represent the distribution (in most cases pdf). Also let $X_1, X_2, \ldots, X_n$ be a random sample of n observations and let $x_1, x_2, \ldots, x_n$ be realizations of these random variables.
- Since we have a random sample the joint pdf can be written as follows

$$\begin{aligned} f(x_1, x_2, \ldots, x_n; \theta) &= f_1(x_1;\theta) \cdot f_2(x_2;\theta) \cdot, \ldots, \cdot f_n(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta), \quad i = 1, 2, \ldots, n \end{aligned}$$

- Now suppose that we observe the random sample but we do not know the population parameters. Likelihood function is simply a function of θ given the realization of the random sample $\mathbf{x}$:

$$L(\theta \mid x_1, x_2, \ldots, x_n) = L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta), \quad i = 1, 2, \ldots, n$$

Maximum Likelihood Estimation - MLE

- Let $f(x;\theta)$ represent the distribution (in most cases pdf). Also let $X_1, X_2, \ldots, X_n$ be a random sample of n observations and let $x_1, x_2, \ldots, x_n$ be realizations of these random variables.
- Since we have a random sample the joint pdf can be written as follows

$$\begin{aligned} f(x_1, x_2, \ldots, x_n; \theta) &= f_1(x_1;\theta) \cdot f_2(x_2;\theta) \cdot, \ldots, \cdot f_n(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta), \quad i = 1, 2, \ldots, n \end{aligned}$$

- Now suppose that we observe the random sample but we do not know the population parameters. Likelihood function is simply a function of θ given the realization of the random sample $\mathbf{x}$:

$$L(\theta \mid x_1, x_2, \ldots, x_n) = L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta), \quad i = 1, 2, \ldots, n$$

Maximum Likelihood Estimation - MLE

- MLE finds estimators by maximizing the likelihood function, i.e.,

$$\max_{\theta} L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta)$$

- Let $\hat{\theta}_{mle}$ be the ML estimator of θ then

$$\hat{\theta}_{mle} = \arg\max_{\theta} L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta)$$

- Usually instead of solving the maximization problem defined above it is much easier to work with the loglikelihood function:

$$\max_{\theta} \ln L(\theta \mid \mathbf{x}) = \ln\left(\prod_{i=1}^{n} f(x_i;\theta)\right) = \sum_{i=1}^{n} \ln\left(f(x_i;\theta)\right)$$

Maximum Likelihood Estimation - MLE

- MLE finds estimators by maximizing the likelihood function, i.e.,

$$\max_{\theta} L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta)$$

- Let $\hat{\theta}_{mle}$ be the ML estimator of θ then

$$\hat{\theta}_{mle} = \arg\max_{\theta} L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta)$$

- Usually instead of solving the maximization problem defined above it is much easier to work with the loglikelihood function:

$$\max_{\theta} \ln L(\theta \mid \mathbf{x}) = \ln \left(\prod_{i=1}^{n} f(x_i; \theta) \right) = \sum_{i=1}^{n} \ln \left(f(x_i; \theta) \right)$$

Maximum Likelihood Estimation - MLE

- MLE finds estimators by maximizing the likelihood function, i.e.,

$$\max_{\theta} L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta)$$

- Let $\hat{\theta}_{mle}$ be the ML estimator of θ then

$$\hat{\theta}_{mle} = \arg\max_{\theta} L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta)$$

- Usually instead of solving the maximization problem defined above it is much easier to work with the loglikelihood function:

$$\max_{\theta} \ln L(\theta \mid \mathbf{x}) = \ln \left(\prod_{i=1}^{n} f(x_i; \theta) \right) = \sum_{i=1}^{n} \ln \left(f(x_i; \theta) \right)$$

- MLE of θ:

$$\hat{\theta}_{mle} = \arg\max_{\theta} \ln L(\theta \mid \mathbf{x})$$

- or

$$\hat{\theta}_{mle} = \arg\max_{\theta} \sum_{i=1}^{n} \ln\left(f(x_i;\theta)\right)$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial\theta} \ln L(\theta \mid \mathbf{x}) = 0, \quad \frac{\partial^2}{\partial\theta^2} \ln L(\theta \mid \mathbf{x}) < 0$$

- Hessian matrix must be negative definite.

Maximum Likelihood Estimation - MLE

- MLE of θ:

$$\hat{\theta}_{mle} = \text{arg}\max_{\theta} \ln L(\theta \mid \mathbf{x})$$

- or

$$\hat{\theta}_{mle} = \text{arg}\max_{\theta} \sum_{i=1}^{n} \ln\left(f(x_i;\theta)\right)$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial\theta} \ln L(\theta \mid \mathbf{x}) = 0, \quad \frac{\partial^2}{\partial\theta^2} \ln L(\theta \mid \mathbf{x}) < 0$$

- Hessian matrix must be negative definite.

Maximum Likelihood Estimation - MLE

- MLE of θ:

$$\hat{\theta}_{mle} = \text{arg}\max_{\theta} \ln L(\theta \mid \mathbf{x})$$

- or

$$\hat{\theta}_{mle} = \text{arg}\max_{\theta} \sum_{i=1}^{n} \ln\left(f(x_i; \theta)\right)$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial \theta} \ln L(\theta \mid \mathbf{x}) = 0, \quad \frac{\partial^2}{\partial \theta^2} \ln L(\theta \mid \mathbf{x}) < 0$$

- Hessian matrix must be negative definite.

Maximum Likelihood Estimation - MLE

- MLE of θ:

$$\hat{\theta}_{mle} = \text{arg} \max_{\theta} \ln L(\theta \mid \mathbf{x})$$

- or

$$\hat{\theta}_{mle} = \text{arg} \max_{\theta} \sum_{i=1}^{n} \ln \left(f(x_i; \theta) \right)$$

- Necessary and sufficient conditions:

$$\frac{\partial}{\partial \theta} \ln L(\theta \mid \mathbf{x}) = 0, \quad \frac{\partial^2}{\partial \theta^2} \ln L(\theta \mid \mathbf{x}) < 0$$

- Hessian matrix must be negative definite.

MLE - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample drawn from a Bernoulli distribution with success probability p $(0 < p < 1)$. Also let $x_1, x_2, \ldots, x_{10}$ be the observed values of this random sample.
- Let $y = x_1 + x_2 + \ldots + x_{10}$ denote total number of successes in the random sample. Then the likelihood function is given by:

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^y (1-p)^{n-y}$$

- Suppose that out of 10 trials we observed 6 successes, i.e., $y = x_1 + x_2 + \ldots + x_{10} = 6$.
- The likelihood is

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^6 (1-p)^4$$

MLE - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample drawn from a Bernoulli distribution with success probability p $(0 < p < 1)$. Also let $x_1, x_2, \ldots, x_{10}$ be the observed values of this random sample.
- Let $y = x_1 + x_2 + \ldots + x_{10}$ denote total number of successes in the random sample. Then the likelihood function is given by:

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^y (1-p)^{n-y}$$

- Suppose that out of 10 trials we observed 6 successes, i.e., $y = x_1 + x_2 + \ldots + x_{10} = 6$.
- The likelihood is

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^6 (1-p)^4$$

MLE - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample drawn from a Bernoulli distribution with success probability p $(0 < p < 1)$. Also let $x_1, x_2, \ldots, x_{10}$ be the observed values of this random sample.
- Let $y = x_1 + x_2 + \ldots + x_{10}$ denote total number of successes in the random sample. Then the likelihood function is given by:
$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^y (1-p)^{n-y}$$
- Suppose that out of 10 trials we observed 6 successes, i.e., $y = x_1 + x_2 + \ldots + x_{10} = 6$.
- The likelihood is
$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^6 (1-p)^4$$

MLE - Example

- Let $X_1, X_2, \ldots, X_{10}$ be a random sample drawn from a Bernoulli distribution with success probability p $(0 < p < 1)$. Also let $x_1, x_2, \ldots, x_{10}$ be the observed values of this random sample.
- Let $y = x_1 + x_2 + \ldots + x_{10}$ denote total number of successes in the random sample. Then the likelihood function is given by:

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^y (1-p)^{n-y}$$

- Suppose that out of 10 trials we observed 6 successes, i.e., $y = x_1 + x_2 + \ldots + x_{10} = 6$.
- The likelihood is

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^6 (1-p)^4$$

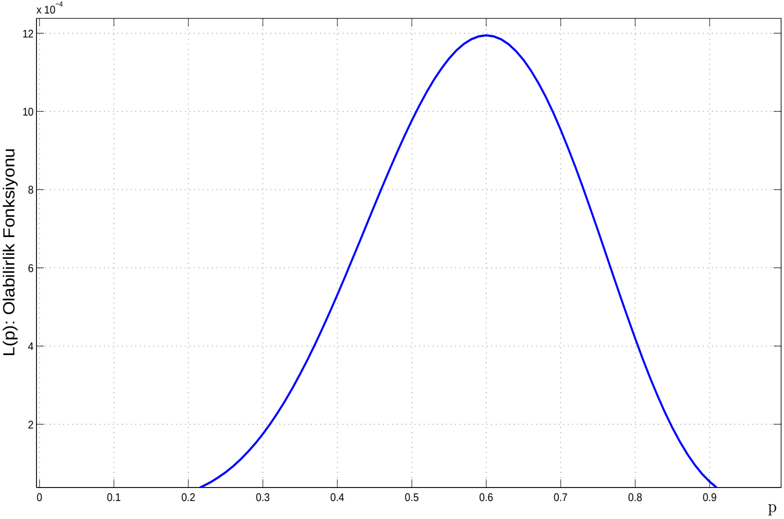
MLE - Example

$$L(p \mid x_1, x_2, \ldots, x_{10}) = p^6(1-p)^4$$

The likelihood values for the interval $0.1, 0.2, \ldots, 0.9$ are given in the following table:

p	$L(p \mid \mathbf{x}) = p^6(1-p)^4$
0.1	0.00000066
0.2	0.00002621
0.3	0.00017503
0.4	0.00053084
0.5	0.00097656
0.6	0.00119439
0.7	0.00095296
0.8	0.00041943
0.9	0.00005314

MLE - Likelihood Function



MLE - Example

- Suppose that $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is random sample drawn from a Poisson distribution. Find the MLE of λ.
- The probability distribution of $X \sim Poisson(\lambda)$ is given by:

$$f(x_i; \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \quad x_i = 1, 2, 3, \dots, \quad i = 1, 2, \dots, n$$

- The loglikelihood function is

$$\begin{aligned}
\ln L(\lambda \mid \mathbf{x}) &= \ln\left[\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] = \ln\left[e^{-n\lambda}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}\right] \\
&= -n\lambda + \ln(\lambda)\sum_{i=1}^{n} x_i - \ln\left[\prod_{i=1}^{n} x_i!\right] \\
&= -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]
\end{aligned}$$

MLE - Example

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is random sample drawn from a Poisson distribution. Find the MLE of λ.
- The probability distribution of $X \sim Poisson(\lambda)$ is given by:

$$f(x_i; \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \quad x_i = 1, 2, 3, \ldots, \quad i = 1, 2, \ldots, n$$

- The loglikelihood function is

$$\begin{aligned}
\ln L(\lambda \mid \mathbf{x}) &= \ln\left[\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] = \ln\left[e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}\right] \\
&= -n\lambda + \ln(\lambda) \sum_{i=1}^{n} x_i - \ln\left[\prod_{i=1}^{n} x_i!\right] \\
&= -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]
\end{aligned}$$

MLE - Example

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is random sample drawn from a Poisson distribution. Find the MLE of λ.
- The probability distribution of $X \sim Poisson(\lambda)$ is given by:

$$f(x_i; \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \quad x_i = 1, 2, 3, \ldots, \quad i = 1, 2, \ldots, n$$

- The loglikelihood function is

$$\begin{aligned}
\ln L(\lambda \mid \mathbf{x}) &= \ln\left[\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] = \ln\left[e^{-n\lambda}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}\right] \\
&= -n\lambda + \ln(\lambda)\sum_{i=1}^{n} x_i - \ln\left[\prod_{i=1}^{n} x_i!\right] \\
&= -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]
\end{aligned}$$

MLE Example continued

- $$\ln L(\lambda \mid \mathbf{x}) = -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]$$
- FOC w.r.t. λ:
$$\frac{\partial}{\partial\lambda}\ln L(\lambda \mid \mathbf{x}) = -n + \frac{n}{\lambda}\bar{x} = 0, \quad \Longrightarrow \lambda = \bar{x}$$
- SOC:
$$\frac{\partial^2}{\partial\lambda^2}\ln L(\lambda \mid \mathbf{x}) = -\frac{n\bar{x}}{\lambda^2} < 0, \quad \text{for each } \lambda$$
- Thus the MLE of λ is
$$\hat{\lambda}_{mle} = \overline{X}$$

MLE Example continued

- $$\ln L(\lambda \mid \mathbf{x}) = -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]$$
- FOC w.r.t. λ:
$$\frac{\partial}{\partial\lambda}\ln L(\lambda \mid \mathbf{x}) = -n + \frac{n}{\lambda}\bar{x} = 0, \quad \Longrightarrow \lambda = \bar{x}$$
- SOC:
$$\frac{\partial^2}{\partial\lambda^2}\ln L(\lambda \mid \mathbf{x}) = -\frac{n\bar{x}}{\lambda^2} < 0, \quad \text{for each } \lambda$$
- Thus the MLE of λ is
$$\hat{\lambda}_{mle} = \overline{X}$$

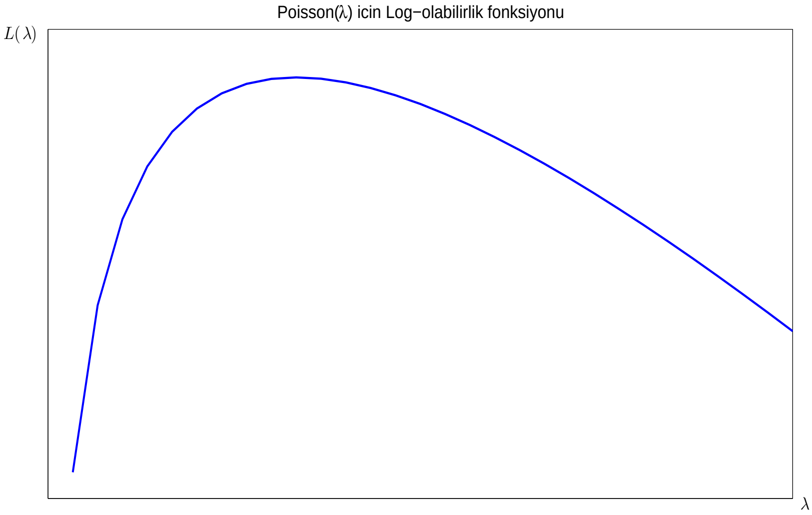
MLE Example continued

- $$\ln L(\lambda \mid \mathbf{x}) = -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]$$
- FOC w.r.t. λ:
$$\frac{\partial}{\partial\lambda}\ln L(\lambda \mid \mathbf{x}) = -n + \frac{n}{\lambda}\bar{x} = 0, \quad \Longrightarrow \lambda = \bar{x}$$
- SOC:
$$\frac{\partial^2}{\partial\lambda^2}\ln L(\lambda \mid \mathbf{x}) = -\frac{n\bar{x}}{\lambda^2} < 0, \quad \text{for each } \lambda$$
- Thus the MLE of λ is
$$\hat{\lambda}_{mle} = \overline{X}$$

MLE Example continued

- $$\ln L(\lambda \mid \mathbf{x}) = -n\lambda + n\ln(\lambda)\bar{x} - \ln\left[\prod_{i=1}^{n} x_i!\right]$$
- FOC w.r.t. λ:
$$\frac{\partial}{\partial\lambda}\ln L(\lambda \mid \mathbf{x}) = -n + \frac{n}{\lambda}\bar{x} = 0, \quad \Longrightarrow \lambda = \bar{x}$$
- SOC:
$$\frac{\partial^2}{\partial\lambda^2}\ln L(\lambda \mid \mathbf{x}) = -\frac{n\bar{x}}{\lambda^2} < 0, \quad \text{for each } \lambda$$
- Thus the MLE of λ is
$$\hat{\lambda}_{mle} = \overline{X}$$

MLE Example continued



MLE - Another Example

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent the observed values of a random sample drawn from a normal distribution with mean μ and variance σ^2. Find the MLE of μ and σ^2.
- $X \sim N(\mu, \sigma^2)$ the pdf is:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right), \quad -\infty < x_i < \infty,$$

- We first need to find the joint pdf and the loglikelihood function using the independence property.

MLE - Another Example

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent the observed values of a random sample drawn from a normal distribution with mean μ and variance σ^2. Find the MLE of μ and σ^2.
- $X \sim N(\mu, \sigma^2)$ the pdf is:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right), \quad -\infty < x_i < \infty, \quad i$$

- We first need to find the joint pdf and the loglikelihood function using the independence property.

MLE - Another Example

- Suppose that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent the observed values of a random sample drawn from a normal distribution with mean μ and variance σ^2. Find the MLE of μ and σ^2.
- $X \sim N(\mu, \sigma^2)$ the pdf is:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right), \quad -\infty < x_i < \infty,$$

- We first need to find the joint pdf and the loglikelihood function using the independence property.

- Likelihood function is

$$
\begin{aligned}
L(\mu, \sigma^2 \mid \mathbf{x}) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)
\end{aligned}
$$

- Loglikelihood function is

$$
\ln L(\mu, \sigma^2 \mid \mathbf{x}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2
$$

MLE Example continued

- Likelihood function is

$$
\begin{aligned}
L(\mu, \sigma^2 \mid \mathbf{x}) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)
\end{aligned}
$$

- Loglikelihood function is

$$
\ln L(\mu, \sigma^2 \mid \mathbf{x}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2
$$

MLE Example continued

- FOC:

$$\begin{aligned}
\frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\
\frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0
\end{aligned}$$

- By soling this system we obtain

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

MLE Example continued

- FOC:

$$\begin{aligned}
\frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\
\frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0
\end{aligned}$$

- By soling this system we obtain

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

Example continued

- SOC

$$
\begin{aligned}
\frac{\partial^2}{\partial \mu^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= -\frac{n}{\sigma^2} \\
\frac{\partial^2}{\partial \mu \partial \sigma^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= -\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu) \\
\frac{\partial^2}{\partial (\sigma^2)^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} (x_i - \mu)^2 \\
\frac{\partial^2}{\partial \sigma^2 \partial \mu} \ln L(\mu, \sigma^2 \mid \mathbf{x}) &= -\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu)
\end{aligned}
$$

Example continued

- Evaluating the Hessian at the MLE we obtain

$$H\big|_{\hat{\mu}_{mle},\hat{\sigma}^2_{mle}} = \begin{bmatrix} \frac{\partial^2}{\partial\mu^2}\ln L(\mu,\sigma^2 \mid \mathbf{x}) & \frac{\partial^2}{\partial\mu\partial\sigma^2}\ln L(\mu,\sigma^2 \mid \mathbf{x}) \\ \frac{\partial^2}{\partial\sigma^2\partial\mu}\ln L(\mu,\sigma^2 \mid \mathbf{x}) & \frac{\partial^2}{\partial(\sigma^2)^2}\ln L(\mu,\sigma^2 \mid \mathbf{x}) \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2_{mle}} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4_{mle}} \end{bmatrix}$$

- Since this matrix is negative definite SOC will be satisfied.

Example continued

- Evaluating the Hessian at the MLE we obtain

$$H\big|_{\hat{\mu}_{mle},\hat{\sigma}^2_{mle}} = \begin{bmatrix} \frac{\partial^2}{\partial \mu^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) & \frac{\partial^2}{\partial \mu \partial \sigma^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) \\ \frac{\partial^2}{\partial \sigma^2 \partial \mu} \ln L(\mu, \sigma^2 \mid \mathbf{x}) & \frac{\partial^2}{\partial (\sigma^2)^2} \ln L(\mu, \sigma^2 \mid \mathbf{x}) \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2_{mle}} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4_{mle}} \end{bmatrix}$$

- Since this matrix is negative definite SOC will be satisfied.

Properties of MLE

- Invariance: Let $\hat{\theta}_{mle}$ be the MLE of θ. Also let $\gamma = g(\theta)$ be a function of θ then the MLE of γ is $\hat{\gamma}_{mle} = g(\hat{\theta}_{mle})$.
- Consistency: $\hat{\theta}_{mle}$ is consistent.
- Asymptotic Normality:

$$n \longrightarrow \infty, \qquad \sqrt{n}(\hat{\theta}_{mle} - \theta) \longrightarrow N\left(0, \sigma^2_{\hat{\theta}}\right)$$

where

$$\sigma^2_{\theta} = \frac{1}{I(\theta)}, \quad I(\theta) = E_{\theta}\left[\frac{\partial}{\partial\theta} \ln L(\theta \mid \mathbf{x})\right]^2$$

$I(\theta)$ is called *Fisher information* matrix.

Properties of MLE

- Invariance: Let $\hat{\theta}_{mle}$ be the MLE of θ. Also let $\gamma = g(\theta)$ be a function of θ then the MLE of γ is $\hat{\gamma}_{mle} = g(\hat{\theta}_{mle})$.
- Consistency: $\hat{\theta}_{mle}$ is consistent.
- Asymptotic Normality:

$$n \longrightarrow \infty, \qquad \sqrt{n}(\hat{\theta}_{mle} - \theta) \longrightarrow N\left(0, \sigma^2_{\hat{\theta}}\right)$$

where

$$\sigma^2_{\theta} = \frac{1}{I(\theta)}, \quad I(\theta) = E_{\theta}\left[\frac{\partial}{\partial\theta} \ln L(\theta \mid \mathbf{x})\right]^2$$

$I(\theta)$ is called *Fisher information* matrix.

Properties of MLE

- Invariance: Let $\hat{\theta}_{mle}$ be the MLE of θ. Also let $\gamma = g(\theta)$ be a function of θ then the MLE of γ is $\hat{\gamma}_{mle} = g(\hat{\theta}_{mle})$.
- Consistency: $\hat{\theta}_{mle}$ is consistent.
- Asymptotic Normality:

$$n \longrightarrow \infty, \qquad \sqrt{n}(\hat{\theta}_{mle} - \theta) \longrightarrow N\left(0, \sigma^2_{\hat{\theta}}\right)$$

where

$$\sigma^2_{\theta} = \frac{1}{I(\theta)}, \quad I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta} \ln L(\theta \mid \mathbf{x})\right]^2$$

$I(\theta)$ is called *Fisher information* matrix.

Hypothesis Tests

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

Hypothesis Tests

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

Hypothesis Tests

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

Hypothesis Tests

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

Hypothesis Tests

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

- Definitions
- Let the unknown population parameter be θ
- **Null Hypothesis**: this is the hypothesis that we believe to be true until evidence suggests the otherwise. Contradiction of the alternative hypothesis. Denoted by the symbol H_0
- **Alternative Hypothesis**: is the hypothesis we wish to support. If the null hypothesis is false, the alternative hypothesis must be true. Denoted by the symbol H_1 or H_a
- H_0 specifies hypothesized values for one or more population parameters.
- For example, we might wish to test the null hypothesis that a population mean is equal to 10, hoping to show, in fact that the mean exceeds 10.
- $H_0 : \mu = 10$ vs $H_1 : \mu \geq 10$

Hypothesis Tests

DECISION: After specifying the null and alternative hypothesis and collecting the sample data we need to make a decision about the null hypothesis using a decision rule.

- Two decisions: **Fail to reject** ("accept") the null hypothesis or **reject** the null hypothesis in favor of the alternative.
- We need to develop a decision rule using the sample.
- This decision rule is generally based on the sampling distributions of test statistics.
- We cannot know for certain whether a null is true or false
- In other words, we might make mistakes in the decision process: accept the null when it is false vs. reject the null when it is true.

DECISION: After specifying the null and alternative hypothesis and collecting the sample data we need to make a decision about the null hypothesis using a decision rule.

- Two decisions: **Fail to reject** ("accept") the null hypothesis or **reject** the null hypothesis in favor of the alternative.
- We need to develop a decision rule using the sample.
- This decision rule is generally based on the sampling distributions of test statistics.
- We cannot know for certain whether a null is true or false
- In other words, we might make mistakes in the decision process: accept the null when it is false vs. reject the null when it is true.

DECISION: After specifying the null and alternative hypothesis and collecting the sample data we need to make a decision about the null hypothesis using a decision rule.

- Two decisions: **Fail to reject** ("accept") the null hypothesis or **reject** the null hypothesis in favor of the alternative.
- We need to develop a decision rule using the sample.
- This decision rule is generally based on the sampling distributions of test statistics.
- We cannot know for certain whether a null is true or false
- In other words, we might make mistakes in the decision process: accept the null when it is false vs. reject the null when it is true.

Hypothesis Tests

DECISION: After specifying the null and alternative hypothesis and collecting the sample data we need to make a decision about the null hypothesis using a decision rule.

- Two decisions: **Fail to reject** ("accept") the null hypothesis or **reject** the null hypothesis in favor of the alternative.
- We need to develop a decision rule using the sample.
- This decision rule is generally based on the sampling distributions of test statistics.
- We cannot know for certain whether a null is true or false
- In other words, we might make mistakes in the decision process: accept the null when it is false vs. reject the null when it is true.

Hypothesis Tests

DECISION: After specifying the null and alternative hypothesis and collecting the sample data we need to make a decision about the null hypothesis using a decision rule.

- Two decisions: **Fail to reject** ("accept") the null hypothesis or **reject** the null hypothesis in favor of the alternative.
- We need to develop a decision rule using the sample.
- This decision rule is generally based on the sampling distributions of test statistics.
- We cannot know for certain whether a null is true or false
- In other words, we might make mistakes in the decision process: accept the null when it is false vs. reject the null when it is true.

Hypothesis Tests

- TYPE I ERROR: rejecting a null hypothesis when it is in fact TRUE
- The probability of Type I Error is denoted α and called the significance level.
- TYPE II ERROR: accepting a null hypothesis when it is in fact FALSE
- The probability of Type II Error is denoted β
- Power of test: probability of not conducting a Type II Error, denoted $(1 - \beta)$

Hypothesis Tests

- TYPE I ERROR: rejecting a null hypothesis when it is in fact TRUE
- The probability of Type I Error is denoted α and called the significance level.
- TYPE II ERROR: accepting a null hypothesis when it is in fact FALSE
- The probability of Type II Error is denoted β
- Power of test: probability of not conducting a Type II Error, denoted $(1-\beta)$

Hypothesis Tests

- TYPE I ERROR: rejecting a null hypothesis when it is in fact TRUE
- The probability of Type I Error is denoted α and called the significance level.
- TYPE II ERROR: accepting a null hypothesis when it is in fact FALSE
- The probability of Type II Error is denoted β
- Power of test: probability of not conducting a Type II Error, denoted $(1 - \beta)$

Hypothesis Tests

- TYPE I ERROR: rejecting a null hypothesis when it is in fact TRUE
- The probability of Type I Error is denoted α and called the significance level.
- TYPE II ERROR: accepting a null hypothesis when it is in fact FALSE
- The probability of Type II Error is denoted β
- Power of test: probability of not conducting a Type II Error, denoted $(1 - \beta)$

- TYPE I ERROR: rejecting a null hypothesis when it is in fact TRUE
- The probability of Type I Error is denoted α and called the significance level.
- TYPE II ERROR: accepting a null hypothesis when it is in fact FALSE
- The probability of Type II Error is denoted β
- Power of test: probability of not conducting a Type II Error, denoted $(1-\beta)$

DECISION		
	H_0 TRUE	H_0 FALSE
ACCEPT H_0	Correct Decision probability$= 1 - \alpha$	Type II Error probability$= \beta$
REJECT H_0	Type I Error probability$= \alpha$	Correct Decision probability$= 1 - \beta$

$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$: Significance Level
$\beta = P(\text{fail to reject } H_0 \mid H_0 \text{ is false})$: probability of Type II Error
$1 - \beta = P(\text{reject } H_0 \mid H_0 \text{ is false})$: Power of test

Hypothesis Tests

- Trade-off: we cannot make both α and β as small as we want
- As α gets smaller, that is, as the probability of Type I error decreases, the probability of conducting Type II error increases, i.e., β gets bigger
- $\alpha \downarrow \beta \uparrow (1-\beta) \downarrow$
- What to do?
- Fix α at a small level, e.g. 0.05 and let β be determined
- To decrease both α and β we need to collect more data

Hypothesis Tests

- Trade-off: we cannot make both α and β as small as we want
- As α gets smaller, that is, as the probability of Type I error decreases, the probability of conducting Type II error increases, i.e., β gets bigger
- $\alpha \downarrow \beta \uparrow (1-\beta) \downarrow$
- What to do?
- Fix α at a small level, e.g. 0.05 and let β be determined
- To decrease both α and β we need to collect more data

- Trade-off: we cannot make both α and β as small as we want
- As α gets smaller, that is, as the probability of Type I error decreases, the probability of conducting Type II error increases, i.e., β gets bigger
- $\alpha \downarrow \beta \uparrow (1-\beta) \downarrow$
- What to do?
- Fix α at a small level, e.g. 0.05 and let β be determined
- To decrease both α and β we need to collect more data

Hypothesis Tests

- Trade-off: we cannot make both α and β as small as we want
- As α gets smaller, that is, as the probability of Type I error decreases, the probability of conducting Type II error increases, i.e., β gets bigger
- $\alpha \downarrow \beta \uparrow (1 - \beta) \downarrow$
- What to do?
- Fix α at a small level, e.g. 0.05 and let β be determined
- To decrease both α and β we need to collect more data

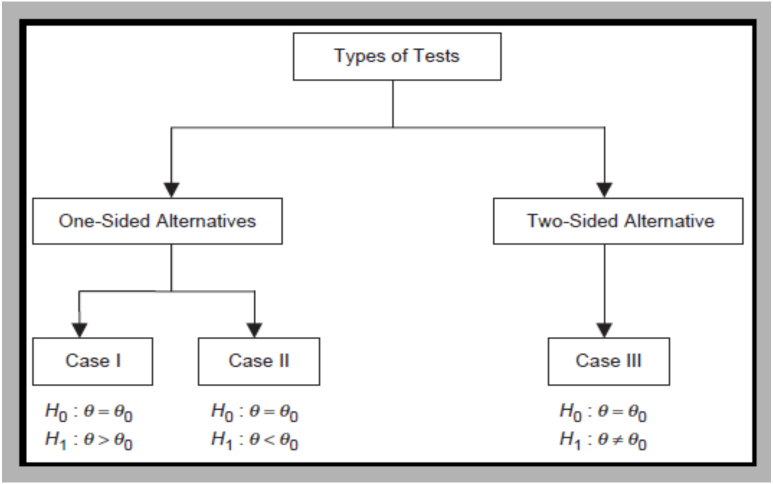
Hypothesis Tests

- Trade-off: we cannot make both α and β as small as we want
- As α gets smaller, that is, as the probability of Type I error decreases, the probability of conducting Type II error increases, i.e., β gets bigger
- $\alpha \downarrow \beta \uparrow (1-\beta) \downarrow$
- What to do?
- Fix α at a small level, e.g. 0.05 and let β be determined
- To decrease both α and β we need to collect more data

Hypothesis Tests

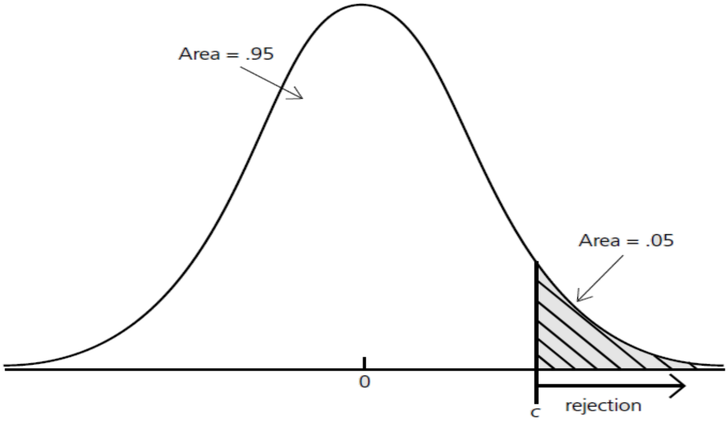
- Trade-off: we cannot make both α and β as small as we want
- As α gets smaller, that is, as the probability of Type I error decreases, the probability of conducting Type II error increases, i.e., β gets bigger
- $\alpha \downarrow \beta \uparrow (1 - \beta) \downarrow$
- What to do?
- Fix α at a small level, e.g. 0.05 and let β be determined
- To decrease both α and β we need to collect more data

Types of tests



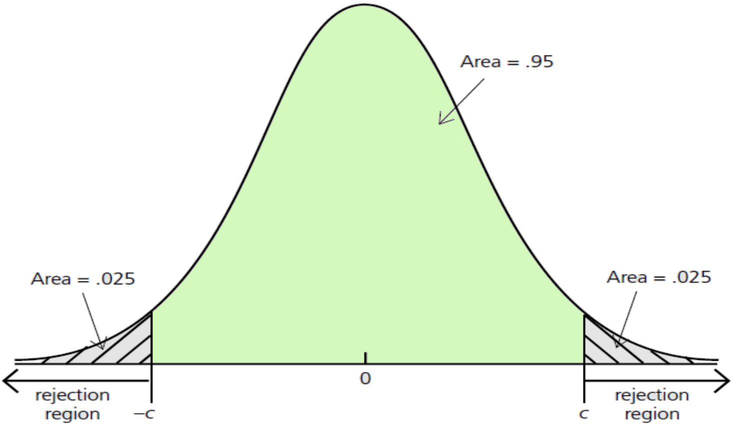
One-sided Alternative - Right Tail Test

Rejection region for a 5% significance level test against the one-sided alternative $\mu > \mu_0$.



Two-sided Alternative

Rejection region for a 5% significance level test against the two-sided alternative H_1: $\mu \neq \mu_0$.



p-value

- Test results may depend on the significance level chosen by the researcher.
- Instead of choosing a significance level prior to the testing we may ask the following question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?
- This is called the p-value
- A small p-value indicates strong evidence against H_0
- If $p - value < \alpha$ H_0 can be rejected at significance level α
- For example, suppose that a large random sample $(n > 120)$ is drawn from a $X \sim N(\mu, \sigma^2)$ population and we want to test $H_0 : \mu = 0$ against one-sided alternative $H_1 : \mu > 0$

p-value

- Test results may depend on the significance level chosen by the researcher.
- Instead of choosing a significance level prior to the testing we may ask the following question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?
- This is called the p-value
- A small p-value indicates strong evidence against H_0
- If $p - value < \alpha$ H_0 can be rejected at significance level α
- For example, suppose that a large random sample $(n > 120)$ is drawn from a $X \sim N(\mu, \sigma^2)$ population and we want to test $H_0 : \mu = 0$ against one-sided alternative $H_1 : \mu > 0$

p-value

- Test results may depend on the significance level chosen by the researcher.
- Instead of choosing a significance level prior to the testing we may ask the following question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?
- This is called the p-value
- A small p-value indicates strong evidence against H_0
- If $p - value < \alpha$ H_0 can be rejected at significance level α
- For example, suppose that a large random sample $(n > 120)$ is drawn from a $X \sim N(\mu, \sigma^2)$ population and we want to test $H_0 : \mu = 0$ against one-sided alternative $H_1 : \mu > 0$

p-value

- Test results may depend on the significance level chosen by the researcher.
- Instead of choosing a significance level prior to the testing we may ask the following question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?
- This is called the p-value
- A small p-value indicates strong evidence against H_0
- If $p - value < \alpha$ H_0 can be rejected at significance level α
- For example, suppose that a large random sample $(n > 120)$ is drawn from a $X \sim N(\mu, \sigma^2)$ population and we want to test $H_0 : \mu = 0$ against one-sided alternative $H_1 : \mu > 0$

p-value

- Test results may depend on the significance level chosen by the researcher.
- Instead of choosing a significance level prior to the testing we may ask the following question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?
- This is called the p-value
- A small p-value indicates strong evidence against H_0
- If $p - value < \alpha$ H_0 can be rejected at significance level α
- For example, suppose that a large random sample $(n > 120)$ is drawn from a $X \sim N(\mu, \sigma^2)$ population and we want to test $H_0 : \mu = 0$ against one-sided alternative $H_1 : \mu > 0$

p-value

- Test results may depend on the significance level chosen by the researcher.
- Instead of choosing a significance level prior to the testing we may ask the following question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?
- This is called the p-value
- A small p-value indicates strong evidence against H_0
- If $p - value < \alpha$ H_0 can be rejected at significance level α
- For example, suppose that a large random sample $(n > 120)$ is drawn from a $X \sim N(\mu, \sigma^2)$ population and we want to test $H_0 : \mu = 0$ against one-sided alternative $H_1 : \mu > 0$

p-value

- Our test statistic is
$$T = \frac{\overline{X}}{s/\sqrt{n}}$$
where s is the sample standard deviation.
- Suppose that the value of T for out sample is $t = 1.52$
- What is the p-value? i.e., the largest significance level at which we would fail to reject H_0
- ... or the smallest significance level at which we would reject H_0
- Since our sample is large enough so that the distribution of the test statistic is standard normal we can write

$$p - value = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) = 0.065$$

p-value

- Our test statistic is

$$T = \frac{\overline{X}}{s/\sqrt{n}}$$

where s is the sample standard deviation.

- Suppose that the value of T for out sample is $t = 1.52$
- What is the p-value? i.e., the largest significance level at which we would fail to reject H_0
- ... or the smallest significance level at which we would reject H_0
- Since our sample is large enough so that the distribution of the test statistic is standard normal we can write

$$p - value = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) = 0.065$$

- Our test statistic is

$$T = \frac{\overline{X}}{s/\sqrt{n}}$$

where s is the sample standard deviation.

- Suppose that the value of T for out sample is $t = 1.52$
- What is the p-value? i.e., the largest significance level at which we would fail to reject H_0
- ... or the smallest significance level at which we would reject H_0
- Since our sample is large enough so that the distribution of the test statistic is standard normal we can write

$$p - value = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) = 0.065$$

- Our test statistic is
$$T = \frac{\overline{X}}{s/\sqrt{n}}$$
where s is the sample standard deviation.
- Suppose that the value of T for out sample is $t = 1.52$
- What is the p-value? i.e., the largest significance level at which we would fail to reject H_0
- ... or the smallest significance level at which we would reject H_0
- Since our sample is large enough so that the distribution of the test statistic is standard normal we can write

$$p - value = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) = 0.065$$

p-value

- Our test statistic is
$$T = \frac{\overline{X}}{s/\sqrt{n}}$$
where s is the sample standard deviation.
- Suppose that the value of T for out sample is $t = 1.52$
- What is the p-value? i.e., the largest significance level at which we would fail to reject H_0
- ... or the smallest significance level at which we would reject H_0
- Since our sample is large enough so that the distribution of the test statistic is standard normal we can write

$$p - value = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) = 0.065$$

p-value

$$p - value = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) = 0.065$$

The p-value when $t = 1.52$ for the one-sided alternative $\mu > \mu_0$.