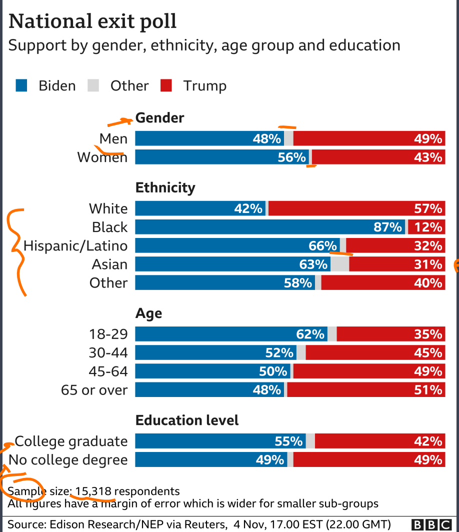
Probability and Statistics: Lecture-40

Monsoon-2020

by Dr. Pawan Kumar (IIIT, Hyderabad)
on November 16, 2020

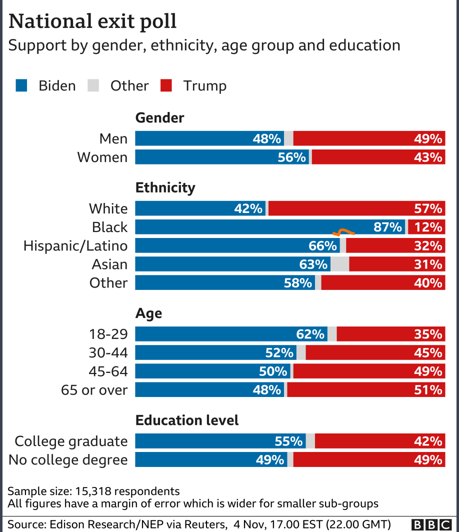
» Motivation for Statistical Inference...

» Motivation for Statistical Inference...



← 54%

» Motivation for Statistical Inference...



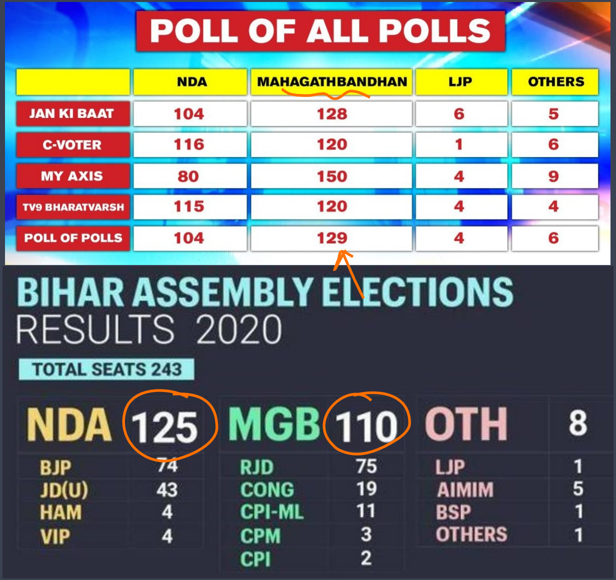
* On the left, US exit poll results
* Poll on Trump Vs Biden
* Sample size of 15,318
* Error margin shown in grey
* Draw conclusions from the sample data
* Will inference fail? How much it can fail?
* How confident we are of this?

» Motivation for Statistical Inference...

» Motivation for Statistical Inference...

POLL OF ALL POLLS

	NDA	MAHAGATHBANDHAN	LJP	OTHERS
JAN KI BAAT	104	128	6	5
C-VOTER	116	120	1	6
MY AXIS	80	150	4	9
TV9 BHARATVARSH	115	120	4	4
POLL OF POLLS	104	129	4	6



POLL OF ALL POLLS

	NDA	MAHAGATHBANDHAN	LJP	OTHERS
JAN KI BAAT	104	128	6	5
C-VOTER	116	120	1	6
MY AXIS	80	150	4	9
TV9 BHARATVARSH	115	120	4	4
POLL OF POLLS	104	129	4	6

BIHAR ASSEMBLY ELECTIONS
RESULTS 2020
TOTAL SEATS 243

NDA	125	MGB	110	OTH	8
BJP	74	RJD	75	LJP	1
JD(U)	43	CONG	19	AIMIM	5
HAM	4	CPI-ML	11	BSP	1
VIP	4	CPM	3	OTHERS	1
		CPI	2		

- On the left, poll of polls showing clear majority for MAHAGATHBANDHAN
- After election, NDA has full majority
- How do we estimate such errors?

» Definition of Statistical Inference...

» Definition of Statistical Inference...

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used
* we need to work with real data

Definition of Statistical Inference

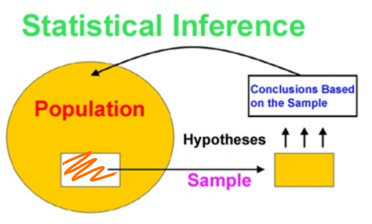
Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used
* we need to work with real data
* distribution of the data may not be known

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

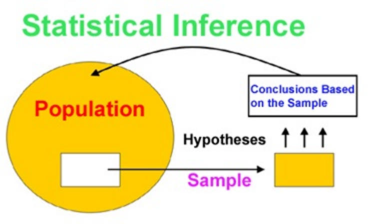
* knowledge of probability is used
* we need to work with real data
* distribution of the data may not be known



Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used
* we need to work with real data
* distribution of the data may not be known

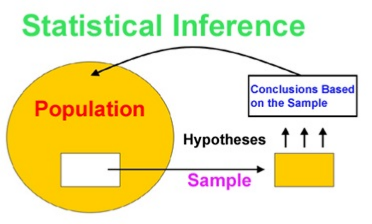


* Two types: Frequentist and Bayesian

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used
* we need to work with real data
* distribution of the data may not be known



* Two types: Frequentist and Bayesian

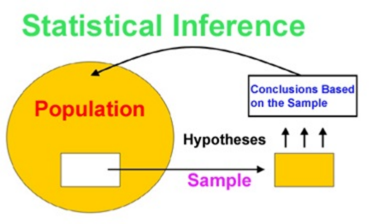
Statistical Inference Problem

To determine an unknown quantity,

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used
* we need to work with real data
* distribution of the data may not be known



* Two types: Frequentist and Bayesian

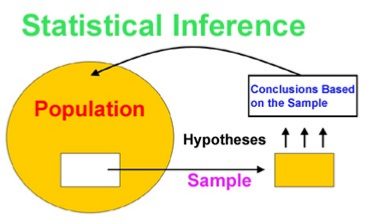
Statistical Inference Problem

To determine an unknown quantity, get some data,

Definition of Statistical Inference

Statistical inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

* knowledge of probability is used
* we need to work with real data
* distribution of the data may not be known



* Two types: Frequentist and Bayesian

Statistical Inference Problem

To determine an unknown quantity, get some data, and then estimate the required quantity using this data.

» Frequentist or Classical Inference...

Recall: A statistical inference problem is to estimate an unknown quantity

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Here the unknown quantity is assumed to be fixed quantity and not random.

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Here the unknown quantity is assumed to be fixed quantity and not random. So, the unknown quantity θ is to be estimated by the observed data.

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Here the unknown quantity is assumed to be fixed quantity and not random. So, the unknown quantity θ is to be estimated by the observed data.

* Let θ be the percentage of people who will vote for a given candidate

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Here the unknown quantity is assumed to be fixed quantity and not random. So, the unknown quantity θ is to be estimated by the observed data.

- Let θ be the percentage of people who will vote for a given candidate
  - $\hat{\Theta} = \frac{Y}{n}$, Y is the number of people among randomly chosen ones who will vote for candidate

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Here the unknown quantity is assumed to be fixed quantity and not random. So, the unknown quantity θ is to be estimated by the observed data.

* Let θ be the percentage of people who will vote for a given candidate
  * $\hat{\Theta} = \frac{Y}{n}$, Y is the number of people among randomly chosen ones who will vote for candidate
  * Although, θ is non random, we estimate it via $\hat{\Theta}$, a random variable

Recall: A statistical inference problem is to estimate an unknown quantity

Frequentist Inference

Here the unknown quantity is assumed to be fixed quantity and not random. So, the unknown quantity θ is to be estimated by the observed data.

* Let θ be the percentage of people who will vote for a given candidate
  * $\hat{\Theta} = \frac{Y}{n}$, Y is the number of people among randomly chosen ones who will vote for candidate
  * Although, θ is non random, we estimate it via $\hat{\Theta}$, a random variable
  * Here $\hat{\Theta}$ is random variable, because it depends on random sample

What is Bayesian Inference?

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable.

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ.

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission,

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p,

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p, and sends a 0 with probability $1 - p$

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p, and sends a 0 with probability $1 - p$
* Hence, if Θ is the transmitted bit,

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p, and sends a 0 with probability $1 - p$
* Hence, if Θ is the transmitted bit, then $\Theta \sim \text{Bernoulli}(p)$

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p, and sends a 0 with probability $1 - p$
* Hence, if Θ is the transmitted bit, then $\Theta \sim \text{Bernoulli}(p)$
* Let us assume that at receiver end we get the output X

What is Bayesian Inference?

Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p, and sends a 0 with probability $1 - p$
* Hence, if Θ is the transmitted bit, then $\Theta \sim \text{Bernoulli}(p)$
* Let us assume that at receiver end we get the output X
* The problem now is to estimate Θ from the noisy output X

What is Bayesian Inference?

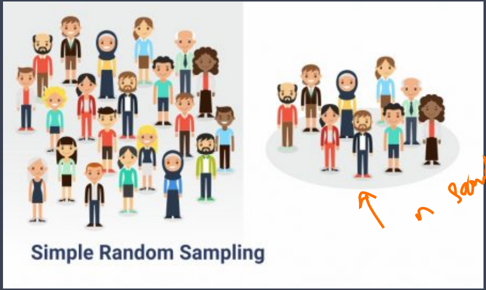
Here the unknown quantity Θ is assumed to be a random variable. Furthermore, we assume to have some initial guess about the distribution of Θ. After we observe the data, we can update the distribution of Θ using Bayes rule.

* Consider communication systems in which information is transmitted in the form of bits
* In each transmission, the transmitter sends a 1 with probability p, and sends a 0 with probability $1 - p$
* Hence, if Θ is the transmitted bit, then $\Theta \sim \text{Bernoulli}(p)$
* Let us assume that at receiver end we get the output X
* The problem now is to estimate Θ from the noisy output X
* We use the prior knowledge that $\Theta \sim \text{Bernoulli}(p)$

» What is Random Sampling? Motivation with an example...

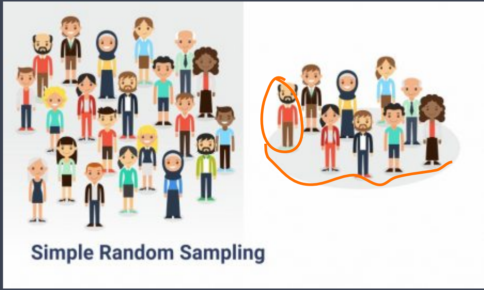
» What is Random Sampling? Motivation with an example...

» What is Random Sampling? Motivation with an example...



n sample

1. Choose a random sample of size n : $X_1, \ldots, X_n$ with replacement



1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person.

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person.

» What is Random Sampling? Motivation with an example...

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person. Again, every person in the population has the same chance of being chosen

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person. Again, every person in the population has the same chance of being chosen

* In general, X_i is the height of the ith person that is chosen uniformly and independently from the population

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person. Again, every person in the population has the same chance of being chosen

* In general, X_i is the height of the ith person that is chosen uniformly and independently from the population
* why do we do the sampling with replacement?

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person. Again, every person in the population has the same chance of being chosen

* In general, X_i is the height of the ith person that is chosen uniformly and independently from the population
* why do we do the sampling with replacement?
    * if the population is large, then the probability of choosing one person twice is extremely low

» What is Random Sampling? Motivation with an example...

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person. Again, every person in the population has the same chance of being chosen

* In general, X_i is the height of the ith person that is chosen uniformly and independently from the population
* why do we do the sampling with replacement?
  * if the population is large, then the probability of choosing one person twice is extremely low
  * big advantage of sampling with replacement is that X_i's will be independent

» What is Random Sampling? Motivation with an example...

1. Choose a random sample of size $n : X_1, \ldots, X_n$ with replacement
2. We chose a person uniformly at random from the population and let X_1 be the height of that person. Here, every person in the population has the same chance of being chosen
3. To determine the value of X_2, again we choose a person uniformly (and independently from the first person) at random and let X_2 be the height of that person. Again, every person in the population has the same chance of being chosen

* In general, X_i is the height of the ith person that is chosen uniformly and independently from the population
* why do we do the sampling with replacement?
  * if the population is large, then the probability of choosing one person twice is extremely low
  * big advantage of sampling with replacement is that X_i's will be independent
  * that is, working with independently and identically distributed makes analysis simpler

» Definition of Random Sample...

» Definition of Random Sample...

Definition of Random sample

The collection of random variables $X_1, X_2, X_3, ..., X_n$ is said to be a random sample of size n

Definition of Random sample

The collection of random variables $X_1, X_2, X_3, ..., X_n$ is said to be a random sample of size n if they are independent and identically distributed (i.i.d.), i.e.,

Definition of Random sample

The collection of random variables $X_1, X_2, X_3, \ldots, X_n$ is said to be a random sample of size n if they are independent and identically distributed (i.i.d.), i.e.,

1. $X_1, X_2, X_3, \ldots, X_n$ are independent random variables, and
2. they have the same distribution, i.e,

Definition of Random sample

The collection of random variables $X_1, X_2, X_3, ..., X_n$ is said to be a random sample of size n if they are independent and identically distributed (i.i.d.), i.e.,

1. $X_1, X_2, X_3, ..., X_n$ are independent random variables, and
2. they have the same distribution, i.e,

$$F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x), \quad \text{for all } x \in \mathbb{R}$$

» Point Estimator and Sample Mean...

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample.

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d.

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

1. The X_i's are independent (since they are i.i.d.)

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

1. The X_i's are independent (since they are i.i.d.)
2. $F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x)$ (the CDFs are same)

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

1. The X_i's are independent (since they are i.i.d.)
2. $F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x)$ (the CDFs are same)
3. $E[X_i] = E[X] = \mu < \infty$

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

1. The X_i's are independent (since they are i.i.d.)
2. $F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x)$ (the CDFs are same)
3. $E[X_i] = E[X] = \mu < \infty$
4. $0 < \mathrm{Var}(X_i) = \mathrm{Var}(X) = \sigma^2 < \infty$

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

1. The X_i's are independent (since they are i.i.d.)
2. $F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x)$ (the CDFs are same)
3. $E[X_i] = E[X] = \mu < \infty$
4. $0 < \text{Var}(X_i) = \text{Var}(X) = \sigma^2 < \infty$

Then the sample mean is defined as follows

Definition of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be random sample. That is, here $X_1, X_2, \ldots, X_n$ are i.i.d. That is, following holds true for i.i.d. random variables

1. The X_i's are independent (since they are i.i.d.)
2. $F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x)$ (the CDFs are same)
3. $E[X_i] = E[X] = \mu < \infty$
4. $0 < \text{Var}(X_i) = \text{Var}(X) = \sigma^2 < \infty$

Then the sample mean is defined as follows

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

» Recall: Properties of Sample Mean...

» Recall: Properties of Sample Mean...

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}$

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad Var(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

» Recall: Properties of Sample Mean...

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

3. Central limit theorem: The random variable

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad \text{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

3. Central limit theorem: The random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

» Recall: Properties of Sample Mean...

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

3. Central limit theorem: The random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu, \quad \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

3. Central limit theorem: The random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \quad \text{for all } x \in \mathbb{R},$$

std. normal

Properties of sample mean, $\bar{X}$

1. $E[\bar{X}] = \mu$, $\quad \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
2. Weak law of large numbers (WLLN)

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

3. Central limit theorem: The random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \quad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is standard normal CDF.

» Order Statistics and its PDF and CDF...

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$.

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, X_2, \ldots, X_n),$$

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, X_2, \ldots, X_n),$$

then $X'_{(i)}$s is called order statistics.

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, X_2, \ldots, X_n),$$

then $X'_{(i)}$s is called order statistics. The CDF and PDF of $X_{(i)}$ are given by

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, X_2, \ldots, X_n),$$

then $X'_{(i)}$s is called order statistics. The CDF and PDF of $X_{(i)}$ are given by

$$\begin{cases} f_{X_{(i)}} = \dfrac{n!}{(i-1)!(n-i)!} f_X(x)[F_X(x)]^{i-1}[1 - F_X(x)]^{n-i} \\ F_{X_{(i)}} = \displaystyle\sum_{k=i}^{n} \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k} \end{cases}$$

Order Statistics and its PDF and CDF

Let $X_1, X_2, \ldots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, X_2, \ldots, X_n),$$

then $X'_{(i)}$s is called order statistics. The CDF and PDF of $X_{(i)}$ are given by

$$f_{X_{(i)}} = \frac{n!}{(i-1)!(n-i)!} f_X(x) [F_X(x)]^{i-1} [1 - F_X(x)]^{n-i}$$

$$F_{X_{(i)}} = \sum_{k=i}^{n} \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k}$$

Also, the joint PDF of $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ is given by

Order Statistics and its PDF and CDF

Let $X_1, X_2, \dots, X_n$ be random sample from a continuous distribution with CDF $F_X(x)$. If we order the random variables from smallest to largest i.e., $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, X_2, \dots, X_n),$$

then $X'_{(i)}$s is called order statistics. The CDF and PDF of $X_{(i)}$ are given by

$$f_{X_{(i)}} = \frac{n!}{(i-1)!(n-i)!} f_X(x)[F_X(x)]^{i-1}[1 - F_X(x)]^{n-i}$$

$$F_{X_{(i)}} = \sum_{k=i}^{n} \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k}$$

Also, the joint PDF of $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ is given by

$$f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n!\, f_X(x_1, f_X(x_2) \cdots f_X(x_n)) & \text{for } x_1 \le x_2 \le \cdots \le x_n \\ 0 & \text{otherwise} \end{cases}$$

» Example of Order Statistics...

Example (Order Statistics)

Let $X_1, X_2, \ldots, X_4$ be a random variable from the Uniform(0,1) distribution,

Example (Order Statistics)

Let $X_1, X_2, \ldots, X_4$ be a random variable from the Uniform(0,1) distribution, and let $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ be the order statistics of $X_1, X_2, \ldots, X_4$.

Example (Order Statistics)

Let $X_1, X_2, \ldots, X_4$ be a random variable from the Uniform(0,1) distribution, and let $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ be the order statistics of $X_1, X_2, \ldots, X_4$.
Find the PDFs of $X_{(1)}, X_{(2)}$, and $X_{(4)}$.

$$f_{X_{(2)}}(x) = \frac{4!}{(2-1)!\,(4-2)!}\, f_X(x) \left[F_X(x)\right]^{2-1} \left[1 - F_X(x)\right]^{4-2}$$

$=$

» Answer to previous problem...

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \dots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

4. There can be many possible point estimators, which one to choose?

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

4. There can be many possible point estimators, which one to choose?
   * For example if $\theta = E[X]$, then $\hat{\Theta} = h(X_1, \ldots, X_n) = \dfrac{X_1 + \cdots + X_n}{n}$

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

4. There can be many possible point estimators, which one to choose?
   * For example if $\theta = E[X]$, then $\hat{\Theta} = h(X_1, \ldots, X_n) = \dfrac{X_1 + \cdots + X_n}{n}$
5. **Bias:** The bias of a point estimator $\hat{\Theta}$ is defined as

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

4. There can be many possible point estimators, which one to choose?
   * For example if $\theta = E[X]$, then $\hat{\Theta} = h(X_1, \ldots, X_n) = \dfrac{X_1 + \cdots + X_n}{n}$
5. **Bias:** The bias of a point estimator $\hat{\Theta}$ is defined as

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$$

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

4. There can be many possible point estimators, which one to choose?
   * For example if $\theta = E[X]$, then $\hat{\Theta} = h(X_1, \ldots, X_n) = \dfrac{X_1 + \cdots + X_n}{n}$
5. **Bias:** The bias of a point estimator $\hat{\Theta}$ is defined as

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$$

   * If bias is close to 0, then $\hat{\Theta}$ is closer to θ

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

4. There can be many possible point estimators, which one to choose?
   * For example if $\theta = E[X]$, then $\hat{\Theta} = h(X_1, \ldots, X_n) = \frac{X_1 + \cdots + X_n}{n}$
5. **Bias:** The bias of a point estimator $\hat{\Theta}$ is defined as

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$$

   * If bias is close to 0, then $\hat{\Theta}$ is closer to θ
   * We say that $\hat{\Theta}$ is an unbiased estimator for a parameter θ if

Definitions: point estimator, bias and unbiased estimators

1. Let θ be an unknown parameter to be estimated. For example, $\theta = E[X]$
2. Let $X_1, X_2, \ldots, X_n$ be a random sample using which we want to estimate θ. Here X_i's have same distribution
3. To estimate θ we define point estimator $\hat{\Theta}$ as follow
$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$
4. There can be many possible point estimators, which one to choose?
   * For example if $\theta = E[X]$, then $\hat{\Theta} = h(X_1, \ldots, X_n) = \dfrac{X_1 + \cdots + X_n}{n}$
5. **Bias:** The bias of a point estimator $\hat{\Theta}$ is defined as
$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$$
   * If bias is close to 0, then $\hat{\Theta}$ is closer to θ
   * We say that $\hat{\Theta}$ is an unbiased estimator for a parameter θ if
$$B(\hat{\Theta}) = 0, \quad \text{for all possible values of } \theta$$

» Unbiased Estimator is not Necessarily a Good Estimator...

Fact

Show that unbiased estimator is not necessarily a good estimator.

Pf : $x_1, x_2, \cdots x_n$ Random sample, $\theta = E[x_i] = E[x]$

$$\hat{\Theta} = \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

If we choose $\underbrace{\hat{\Theta}_1 = x_1}$, then $\hat{\Theta}_1$ is also an unbiased estimator of θ:

$$B(\hat{\Theta}_1) = E[\hat{\Theta}_1] - \theta = \underbrace{E[x_1]} - \theta = \theta - \theta = 0$$

Observe : $\hat{\Theta}_1$ is probably not as good as sample mean $\bar{x}$.

need other measures to ensure that estimator is "good" estimators.

. Better : $E[(\hat{\Theta}-\theta)^2]$

» Mean Squared Error...

Mean squared error

The **mean squared error (MSE)** of a point estimator $\hat{\Theta}$ denoted by $\text{MSE}(\hat{\Theta})$ is defined as

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by MSE($\hat{\Theta}$) is defined as

$$\text{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by $\text{MSE}(\hat{\Theta})$ is defined as

$$\text{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Example (Application of MSE)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $E[X_i] = \theta$, and variance $\text{Var}(X_i) = \sigma^2$.

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by $\mathrm{MSE}(\hat{\Theta})$ is defined as

$$\mathrm{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Example (Application of MSE)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $E[X_i] = \theta$, and variance $\mathrm{Var}(X_i) = \sigma^2$. For the following two estimators for θ

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by $\text{MSE}(\hat{\Theta})$ is defined as

$$\text{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Example (Application of MSE)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $E[X_i] = \theta$, and variance $\text{Var}(X_i) = \sigma^2$. For the following two estimators for θ

1. $\hat{\Theta}_1 = X_1$

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by $\mathrm{MSE}(\hat{\Theta})$ is defined as

$$\mathrm{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Example (Application of MSE)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $E[X_i] = \theta$, and variance $\mathrm{Var}(X_i) = \sigma^2$. For the following two estimators for θ

1. $\hat{\Theta}_1 = X_1$
2. $\hat{\Theta}_2 = \bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by $\text{MSE}(\hat{\Theta})$ is defined as

$$\text{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Example (Application of MSE)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $E[X_i] = \theta$, and variance $\text{Var}(X_i) = \sigma^2$. For the following two estimators for θ

1. $\hat{\Theta}_1 = X_1$
2. $\hat{\Theta}_2 = \bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

Find $\text{MSE}(\hat{\Theta}_1)$ and $\text{MSE}(\hat{\Theta}_2)$ and show that for $n > 1$

Mean squared error

The mean squared error (MSE) of a point estimator $\hat{\Theta}$ denoted by $\text{MSE}(\hat{\Theta})$ is defined as

$$\text{MSE}(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$$

Example (Application of MSE)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $E[X_i] = \theta$, and variance $\text{Var}(X_i) = \sigma^2$. For the following two estimators for θ

1. $\hat{\Theta}_1 = X_1$
2. $\hat{\Theta}_2 = \bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

Find $\text{MSE}(\hat{\Theta}_1)$ and $\text{MSE}(\hat{\Theta}_2)$ and show that for $n > 1$

$$\text{MSE}(\hat{\Theta}_1) > \text{MSE}(\hat{\Theta}_2)$$

Soln we have

$MSE[\hat{\theta}_1] = E[(\hat{\theta}_1 - \theta)^2]$

$= E[(x_1 - E[x_1])^2] = Var(x_1) = \sigma^2$

To find $MSE(\hat{\theta}_2) = E[(\hat{\theta}_2 - \theta)^2]$

$= E[(\bar{x} - \theta)^2] =$

$Var(\bar{x} - \theta) + (E[\bar{x} - \theta])^2$

$= Var(\bar{x}) + 0$

$\Rightarrow MSE[\hat{\theta}_2] = Var(\bar{x}) = \frac{\sigma^2}{n}$

As $n > 1$

$MSE[\hat{\theta}_1] > MSE[\hat{\theta}_2]$

» Answer to previous problem...

» Relationship of MSE, Variance, and Bias...

Property

If $\hat{\Theta}$ is a point estimator for θ,

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

Pf.

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

» Consistent Estimator...

Definition of Consistent Estimator

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_n, \ldots,$ be a sequence of point estimators of θ.

$$\hat{\Theta}_2 = \frac{X_1 + X_2}{2}$$

$$\hat{\Theta}_3 = \frac{X_1 + X_2 + X_3}{3}$$

Definition of Consistent Estimator

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_n, \ldots,$ be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a consistent estimator of θ, if

Definition of Consistent Estimator

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_n, \ldots,$ be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a consistent estimator of θ, if

$$\lim_{n \to \infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0, \quad \text{for all } \epsilon > 0$$

Theorem

Definition of Consistent Estimator

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_n, \ldots,$ be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a consistent estimator of θ, if

$$\lim_{n \to \infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0, \quad \text{for all } \epsilon > 0$$

Theorem

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots,$ be a sequence of point estimators of θ.

» Consistent Estimator...

Definition of Consistent Estimator

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_n, \ldots,$ be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a consistent estimator of θ, if

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0, \quad \text{for all } \epsilon > 0$$

Theorem

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots,$ be a sequence of point estimators of θ. If

$$\lim_{n\to\infty} \mathrm{MSE}(\hat{\Theta}_n) = 0$$

Definition of Consistent Estimator

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_n, \ldots,$ be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a consistent estimator of θ, if

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0, \qquad \text{for all } \epsilon > 0$$

Theorem

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots,$ be a sequence of point estimators of θ. If

$$\lim_{n\to\infty} \mathrm{MSE}(\hat{\Theta}_n) = 0$$

then $\hat{\Theta}_n$ is a consistent estimator of θ

Pf: $\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = P(|\hat{\Theta}_n - \theta|^2 \geq \epsilon^2) \leq \dfrac{E[\hat{\Theta}_n - \theta]^2}{\epsilon^2}$ [Markov]

$= \lim_{n\to\infty} \dfrac{MSE[\hat{\Theta}_n]}{\epsilon^2} = 0.$

~~$\to 0$ as $n \to \infty$~~

» Answer to previous problem...

» Definition of Sample Variance and Sample Standard Deviation...

» Definition of Sample Variance and Sample Standard Deviation...

Sample Variance and Sample Standard Deviation

sample

Let $X_1, X_2, \ldots, X_n$ be a random ~~variable~~ with mean $E[X_i] = \mu < \infty$, and variance $0 < \text{Var}(X_i) < \sigma^2 < \infty$.

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \mathrm{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \text{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}\right)$$

$\leftarrow$ check

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \mathrm{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}\right)$$

We can check that sample variance is an unbiased estimator of σ^2.

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \mathrm{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}\right)$$

We can check that sample variance is an unbiased estimator of σ^2. The sample standard deviation is defined as

$B = E[S^2 - \sigma^2] = 0$

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \text{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}\right)$$

We can check that sample variance is an unbiased estimator of σ^2. The sample standard deviation is defined as

$$S = \sqrt{S^2}$$

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \mathrm{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}\right)$$

We can check that sample variance is an unbiased estimator of σ^2. The sample standard deviation is defined as

$$S = \sqrt{S^2} \quad \approx \sigma$$

and it is usually used as an estimator for σ.

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random variable with mean $E[X_i] = \mu < \infty$, and variance $0 < \mathrm{Var}(X_i) < \sigma^2 < \infty$. The sample variance of this random sample is defined as

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}\right)$$

We can check that sample variance is an unbiased estimator of σ^2. The sample standard deviation is defined as

$$S = \sqrt{S^2}$$

and it is usually used as an estimator for σ. Also, S is an unbiased estimator of σ

Example (Sample Mean, Sample Variance, Sample Standard Deviation)

Let T be the time that is needed for a specific task in a factory to be completed.

Example (Sample Mean, Sample Variance, Sample Standard Deviation)

Let T be the time that is needed for a specific task in a factory to be completed. In order to estimate the mean and variance of T, we observe a random sample $T_1, T_2, \cdots, T_6$.

Example (Sample Mean, Sample Variance, Sample Standard Deviation)

Let T be the time that is needed for a specific task in a factory to be completed. In order to estimate the mean and variance of T, we observe a random sample $T_1, T_2, \cdots, T_6$. Thus, T_i's are i.i.d. and have the same distribution as T.

Example (Sample Mean, Sample Variance, Sample Standard Deviation)

Let T be the time that is needed for a specific task in a factory to be completed. In order to estimate the mean and variance of T, we observe a random sample $T_1, T_2, \cdots, T_6$. Thus, T_i's are i.i.d. and have the same distribution as T. We obtain the following values (in minutes):

$$18, 21, 17, 16, 24, 20.$$

Example (Sample Mean, Sample Variance, Sample Standard Deviation)

Let T be the time that is needed for a specific task in a factory to be completed. In order to estimate the mean and variance of T, we observe a random sample $T_1, T_2, \cdots, T_6$. Thus, T_i's are i.i.d. and have the same distribution as T. We obtain the following values (in minutes):

$$18, 21, 17, 16, 24, 20.$$

Find the values of the sample mean, the sample variance, and the sample standard deviation for the observed sample.

Sample mean

$$\overline{T} = \frac{T_1 + \cdots + T_6}{6} = \frac{18 + 21 + \cdots + 20}{6}$$

Sample Variance

$$S^2 = \frac{1}{6-1} \sum_{k=1}^{6} (T_k - \overline{T})^2 = \cdots$$

$$S = \sqrt{S^2}.$$

» Answer to previous problem...

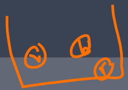
Example

I have a bag that contains 3 balls.

Example

I have a bag that contains 3 balls. Each ball is either red or blue,

Example



r : red
b : blue

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this.

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3.

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement.

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } \textit{ith} \text{ chosen ball is blue} \\ 0 & \text{if the } \textit{ith} \text{ chosen ball is red} \end{cases}$$

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$.

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$. After the experiment, we observe the values for X_i's

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$. After the experiment, we observe the values for X_i's

$$x_1 = 1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 1.$$

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$. After the experiment, we observe the values for X_i's

$$x_1 = 1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 1.$$

From above, we have 3 blue balls and 1 red ball.

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$. After the experiment, we observe the values for X_i's

$$x_1 = 1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 1.$$

From above, we have 3 blue balls and 1 red ball. Answer the following

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$. After the experiment, we observe the values for X_i's

$$x_1 = 1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 1.$$

From above, we have 3 blue balls and 1 red ball. Answer the following

1. Find the probability of the observed sample $(x_1, x_2, x_3, x_4) = (1, 0, 1, 1)$ for each possible θ

Example

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0, 1, 2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, and X_4 as follows

$$X_i = \begin{cases} 1 & \text{if the } ith \text{ chosen ball is blue} \\ 0 & \text{if the } ith \text{ chosen ball is red} \end{cases}$$

We observe here that X_i's are i.i.d. and $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$. After the experiment, we observe the values for X_i's

$$x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1.$$

From above, we have 3 blue balls and 1 red ball. Answer the following

1. Find the probability of the observed sample $(x_1, x_2, x_3, x_4) = (1, 0, 1, 1)$ for each possible θ
2. Find the value of θ that maximizes the probability of the observed sample

Sol:- Since $X_i \sim \text{Bernoulli}\left(\frac{\theta}{3}\right)$.

$$P_{X_i}(x) = \begin{cases} \theta/3 & \text{for } x=1 \quad (\text{blue}) \\ 1-\theta/3 & \text{for } x=0 \quad (\text{red}) \end{cases}$$

Since X_i's are ind., the joint PMF of X_1, X_2, X_3 and X_4 is

$$P_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)$$

$$= P_{X_1}(x_1)\, P_{X_2}(x_2)\, P_{X_3}(x_3)\, P_{X_4}(x_4)$$

This is called max. likelihood estimate of θ

$p = \frac{\theta}{3}$: prob. of picking blue.

$\Rightarrow$

$$P_{X_1X_2X_3X_4}(1,0,1,1) = \frac{\theta}{3}\left(1-\frac{\theta}{3}\right)\frac{\theta}{3}\cdot\frac{\theta}{3}$$

$$= \left(\frac{\theta}{3}\right)^3\left(1-\frac{\theta}{3}\right).$$

Joint PMF depends on θ.

Table

θ	$P_{X_1X_2X_3X_4}(1,0,1,1;\theta)$
0	0
1	0.02
2	0.09
3	0

$\underline{\theta = 2}$ max the probab.

» Answer to previous problem...

Definition of Likelihood and log likehood Function

» Likelihood and log likelihood Function...

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ.

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = P_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = \underbrace{P_{X_1, X_2, \ldots, X_n}}_{\text{PMF}}(x_1, x_2, \ldots, x_n; \theta)$$

2. If X_i's are jointly continuous,

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = P_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

2. If X_i's are jointly continuous, then the likelihood function is defined as

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = P_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

2. If X_i's are jointly continuous, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

CDF

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = P_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

2. If X_i's are jointly continuous, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

In some problems, it is easier to work with the log likelihood function given by

Definition of Likelihood and log likehood Function

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

1. If X_i's are discrete, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = P_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

2. If X_i's are jointly continuous, then the likelihood function is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

In some problems, it is easier to work with the log likelihood function given by

$$\ln L(x_1, x_2, \ldots, x_n; \theta)$$

» Example

Example (Example)

Find the likelihood function for the following random sample

» Example

Example (Example)

Find the likelihood function for the following random sample

1. $X_i \sim \text{Binomial}(3, \theta)$ and we have observed $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$

Example (Example)

Find the likelihood function for the following random sample

1. $X_i \sim \text{Binomial}(3, \theta)$ and we have observed $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$
2. $X_i \sim \text{Exponential}(\theta)$ and we have observed $(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12)$

» Answer to previous problem...

Solution Recall that Random sample: X_1, X_2, X_3, X_4 are i.i.d.

$\Rightarrow$ the joint PMF & (PDF)

= product of marginal PMFs (& PDFs respect.)

① If $X_i \sim$ Binomial$(3, \theta)$,

then $P_{X_i}(x_i; \theta) = \binom{3}{x}\theta^{x}(1-\theta)^{3-x}$

$L(x_1, x_2, x_3, x_4; \theta)$

$= P_{X_1 X_2 X_3 X_4}(x_1, x_2, x_3, x_4; \theta)$

$= P_{X_1} P_{X_2} P_{X_3} P_{X_4}$

$= \binom{3}{x_1}\binom{3}{x_2}\binom{3}{x_3}\binom{3}{x_4}\theta^{x_1+x_2+x_3+x_4}(1-\theta)^{12-(x_1+x_2+x_3+x_4)}$

Since $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$

$\Rightarrow L(1, 3, 2, 2; \theta)$

$= \binom{3}{1}\binom{3}{3}\binom{3}{2}\binom{3}{2}\theta^{8}(1-\theta)^{4}$

$= 27\,\theta^{8}(1-\theta)^{4}.$

» Answer to previous problem...

Definition of maximum likelihood estimator

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ.

Definition of maximum likelihood estimator

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Given that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, a maximum likelihood estimate of θ, shown by $\hat{\theta}_{ML}$ is a value of θ that maximizes the likelihood function

Definition of maximum likelihood estimator

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Given that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, a maximum likelihood estimate of θ, shown by $\hat{\theta}_{ML}$ is a value of θ that maximizes the likelihood function

$$L(x_1, x_2, \ldots, x_n; \theta)$$

Definition of maximum likelihood estimator

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Given that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, a maximum likelihood estimate of θ, shown by $\hat{\theta}_{ML}$ is a value of θ that maximizes the likelihood function

$$L(x_1, x_2, \ldots, x_n; \theta) \quad : \quad f(\theta)$$

A maximum likelihood estimator (MLE) of the parameter θ, shown by $\hat{\theta}_{ML}$ is a random variable $\hat{\theta}_{ML} = \hat{\theta}_{ML}(X_1, X_2, \ldots, X_n)$ whose value when $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is given by $\hat{\theta}_{ML}$.

» Example of Maximum Likelihood Estimator...

Example

For the following examples, find the maximum likelihood estimator (MLE) of θ :

Example

For the following examples, find the maximum likelihood estimator (MLE) of θ :

1. $X_i \sim \text{Binomial}(m, \theta)$, and we have observed $X_1, X_2, \ldots, X_n$

Example

For the following examples, find the maximum likelihood estimator (MLE) of θ :

1. $X_i \sim \text{Binomial}(m, \theta)$, and we have observed $X_1, X_2, \ldots, X_n$ $(1, 3, 2, 2)$
2. $X_i \sim \text{Exponential}(\theta)$ and we have observed $X_1, X_2, \ldots, X_n$ $(1.23, 3.32, \ldots)$

$$L(\underbrace{1, 3, 2, 2}; \theta) = 27\,\theta^8 (1-\theta)^4$$

$$\frac{dL}{d\theta} = 27\left[8\theta^7(1-\theta)^4 - 4\theta^8(1-\theta)^3\right] = 0$$

$$\Rightarrow \hat{\theta}_{ML} = \frac{2}{3}$$

» Answer to previous problem...

» Answer to previous problem...

» Example of Maximum Likelihood Estimators...

Example (Example of maximum likelihood estimator)

Suppose that we have observed the random sample $X_1, X_2, X_3, ..., X_n$, where $X_i \sim N(\theta_1, \theta_2)$ so

$$f_{X_i}(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\dfrac{(x_i - \theta_1)^2}{2\theta_2}}$$

Example (Example of maximum likelihood estimator)

Suppose that we have observed the random sample $X_1, X_2, X_3, ..., X_n$, where $X_i \sim N(\theta_1, \theta_2)$ so

$$f_{X_i}(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2}}$$

Find the maximum likelihood estimators for θ_1 and θ_2.

Better to use log likelihood.

① Find $L(\cdot)$ ② Take $\ln L(\ )$

③ maximize $\ln L(\ )$

» Answer to previous problem...

» Answer to previous problem...

Asymptotic Properties of MLEs

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ.

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ.

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e.,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e.,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable.

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely, the random variable

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely, the random variable

(CLT)

$$\frac{\hat{\Theta}_{ML} - \theta}{\sqrt{\mathrm{Var}(\hat{\Theta}_{ML})}}$$

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e., $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely, the random variable

$$\frac{\hat{\Theta}_{ML} - \theta}{\sqrt{\mathrm{Var}(\hat{\Theta}_{ML})}}$$

   converges in distribution to $N(0, 1)$.

Example

Show the following:

» Solved Example 1 ...

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable.

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable. Show that

$$\hat{\Theta}_2 = \hat{\Theta}_1 + W$$

   is also an unbiased estimator for θ

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable. Show that

$$\hat{\Theta}_2 = \hat{\Theta}_1 + W$$

   is also an unbiased estimator for θ
2. Let $\hat{\Theta}_1$ be an estimator for θ such that $E[\hat{\Theta}_1] = a\theta + b$, where $a \neq 0$.

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable. Show that

$$\hat{\Theta}_2 = \hat{\Theta}_1 + W$$

is also an unbiased estimator for θ

2. Let $\hat{\Theta}_1$ be an estimator for θ such that $E[\hat{\Theta}_1] = a\theta + b$, where $a \neq 0$. Show that

$$\hat{\Theta}_2 = \frac{\hat{\Theta}_1 - b}{a}$$

is an unbiased estimator for θ

Exercise