

# Introduction to Detection and Estimation

# Detection and Estimation: Intro

- The Detection and Estimation theory is a branch of the statistical signal processing that deal with the decision making and the extraction of relevant information from noisy data.
- Many electronic signal processing systems are designed to decide when an event of interest occurs and then extract more information about that event.
- Detection and Estimation theory can be found at the core of those systems.

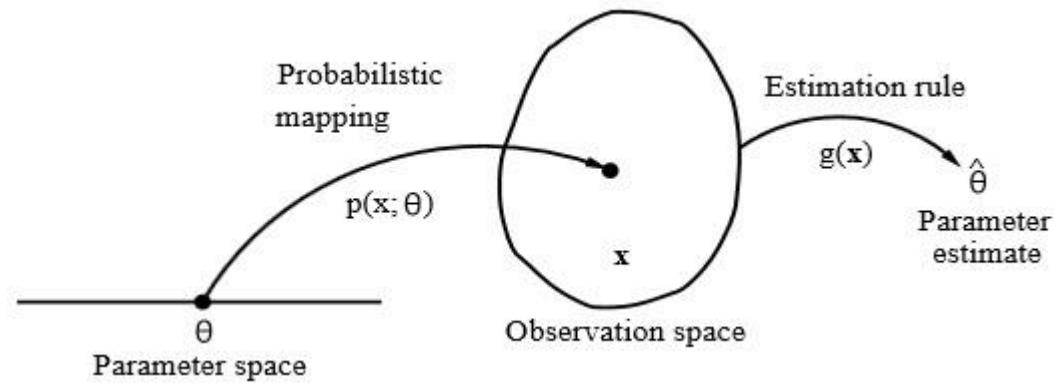
# Applications of Detection

- **Biomedicine:** presence of cardiac arrhythmia need to be detected from ECG.
- **Control System:** occurrence of an abrupt change in the system need to be detected.
- **Communication Systems:** detection of transmitted data in BPSK
- **Image Processing:** detection of an object.
- **Radar Systems:** the occurrence of an air-bourne target (e.g., an aircraft, a missile) is to be detected
- **Seismology:** detection of presence of underground oil
- **Sonar Systems:** the presence of a submarine is to be detected
- **Speech Processing:** presence of different events (such as phonemes or words) is to be detected in speech signal in context of speech recognition application

# Applications of Estimation

- **Biomedicine:** the heart rate of a fetus has to be estimated from sonography during pregnancy.
- **Control System:** the position of a powerboat for the corrective navigation system has to be estimated.
- **Communication Systems:** the carrier frequency of a signal needs to be estimated for demodulation of the baseband signal in the presence of degradation noise.
- **Image Processing:** object's position and orientation estimation
- **Radar Systems:** the delay of the received pulse echo is to be estimated to determine the location of the target
- **Seismology:** the distance of oil deposit has to be estimated from noisy sound reflection due to different densities of oil and rock layers.
- **Sonar Systems:** the delay of the received signal from each of sensors is to be estimated to locate a submarine
- **Speech Processing:** the parameters of the speech production model have to be estimated

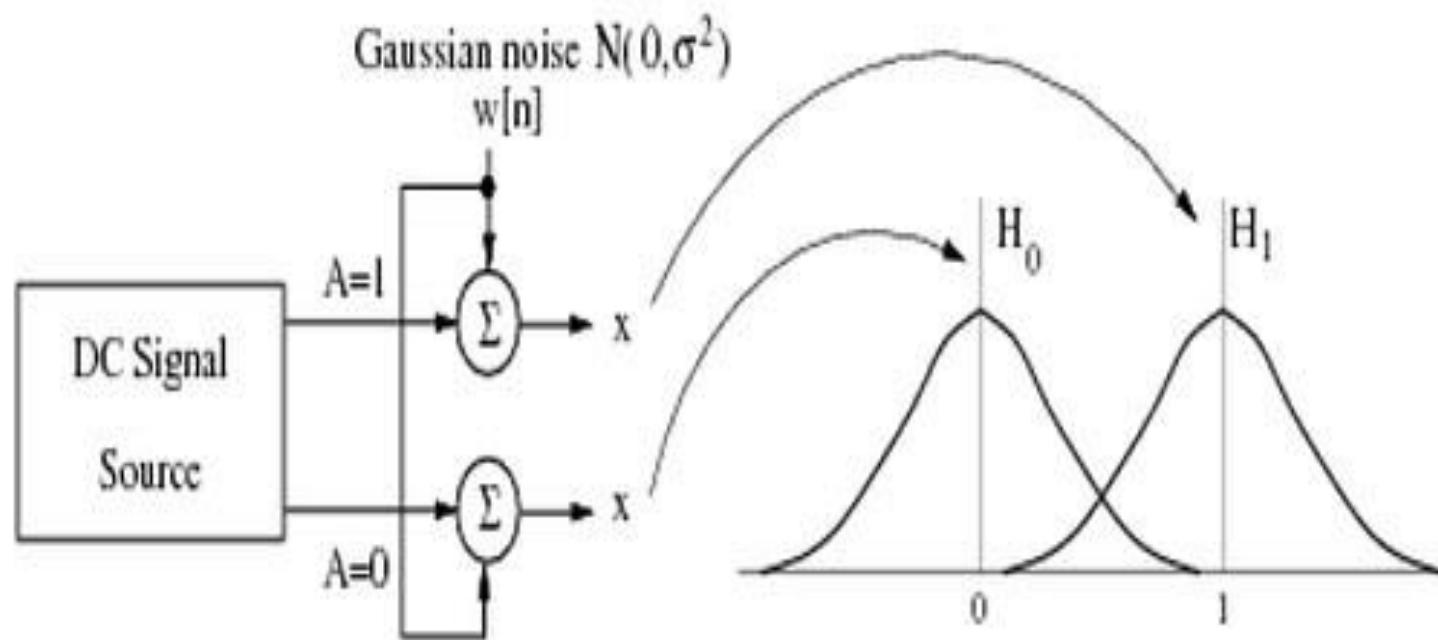
# Formulation of the Estimation Problem



# Classification of Estimation Approaches

	Unknown param.	Probabilistic assumption	Other requirement	Estimator kind (salient property)	Practical utility
Non-random	Completely known PDF	Sufficient statistics	Minimum variance unbiased (optimal)	Low	
	First-two moments only	Large data; no statistics	Maximum likelihood (asymptotically optimal)	Very high	
	Known signal model; no PDF		Best linear unbiased (suboptimal in general)	Moderate	
Random	Known joint and prior PDFs	Conjugate prior; quadratic cost	Minimum mean square error (optimal)	High	
		Hit-or-Miss cost	Maximum a posteriori	High	
		Uniform prior; Hit-or-Miss cost	Bayesian maximum likelihood	Low	
	First-two moments only		Linear minimum mean square error; Wiener filter (suboptimal in general)	Very high	

# Formulation of the Detection Problem



# Hierarchy of Detection Problems

Conditions	Applications
<b>Level 1: Known signals in noise</b>	1. Synchronous digital communication 2. Pattern recognition
<b>Level 2: Signals with unknown parameters in noise</b>	1. Digital communication system without phase reference 2. Digital communication over slowly fading channels 3. Conventional pulse radar and sonar, target detection
<b>Level 3: Random signals in noise</b>	1. Digital communication over scatter link 2. Passive sonar 3. Radio astronomy (detection of noise sources)

# Foundations

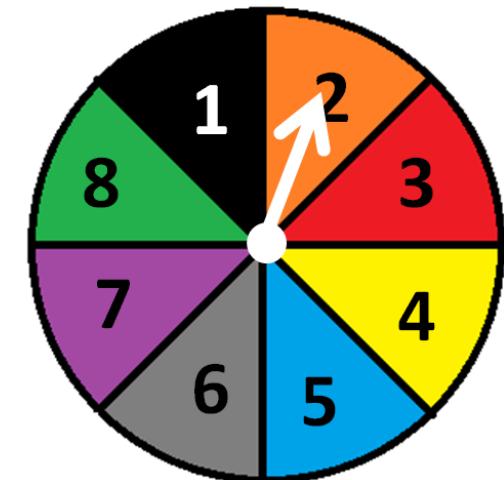
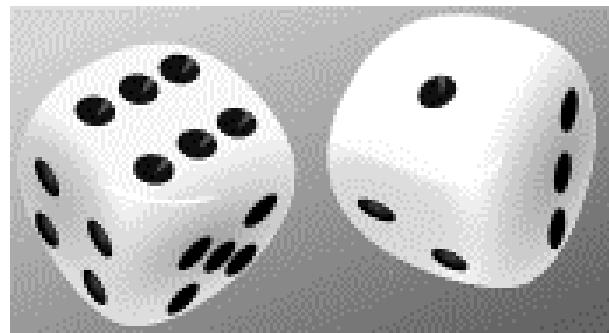
Srimanta Mandal

# Outline

- Probability Basic
- Independent, Dependent Events
- Conditional Probability
- Bayes Theorem
- Random Variable
- Expectation, variance
- Gaussian Distribution
- Correlation

# Probability

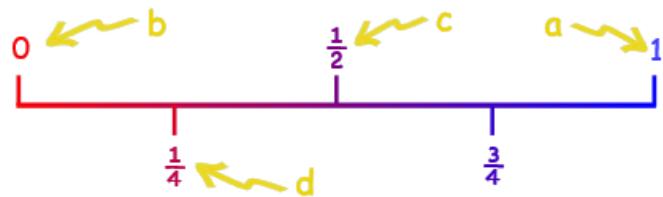
- How **likely** something is to happen.
- Many events can't be predicted with total certainty. The best we can say is how **likely** they are to happen, using the idea of probability.
- Tossing coin, Rolling dice, Rotate a spinner, etc.



# Probability Line

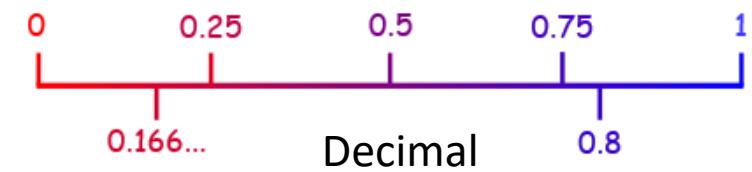
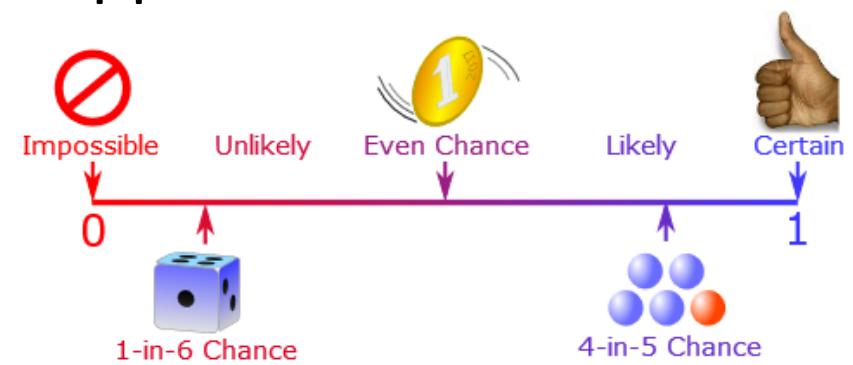
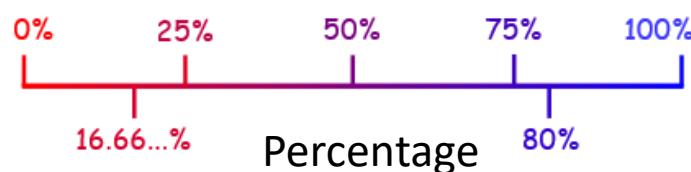
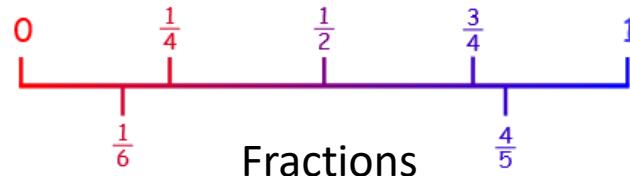
- Probability is the chance that something will happen.

- Impossible = 0
- Certain = 1



- a) The sun will rise tomorrow
- b) I will not have to know probability for this course
- c) If I flip a coin it will land heads up
- d) Choosing a red ball from a bag with 1 red ball and 3 green balls

- Representations:



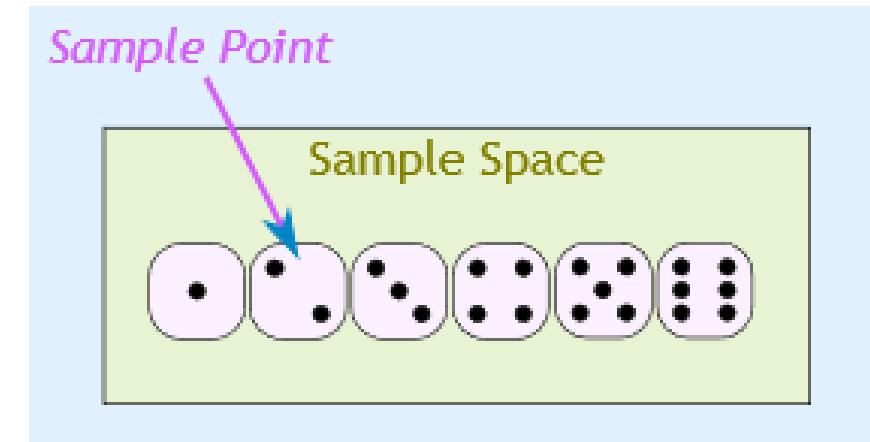
# Classical Definition

- Probability is defined as the ratio of the number of desired outcomes (what you want to happen) to the number of total possible outcomes (what could possibly happen).

$$\text{Probability} = \frac{\text{Number of Desired Outcomes}}{\text{Total Number of Possible Outcomes}}$$

# Random Experiment & Sample Space

- Any process of observation is referred to as an experiment
- The results of an observation are called the outcomes of the experiment
- An experiment is called a random experiment if its outcome cannot be predicted.
- The set of all possible outcomes of a random experiment is called the sample space (denoted by  $S$ ).
- An element in  $S$  is called a sample point.
- Each outcome of a random experiment corresponds to a sample point.



# Events

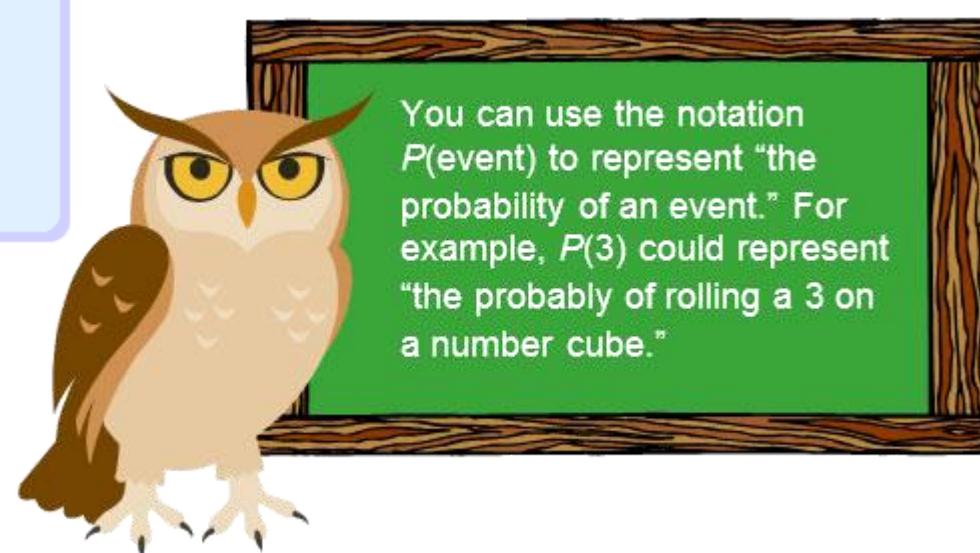
- Life is full of random events!
- Event means one (simple) or more outcomes (compound).

## Example Events:

- Getting a Tail when tossing a coin is an event
- Rolling a "5" is an event.

An event can include several outcomes:

- Choosing a "King" from a deck of cards (any of the 4 Kings) **is also** an event
- Rolling an "even number" (2, 4 or 6) is an event



You can use the notation  $P(\text{event})$  to represent "the probability of an event." For example,  $P(3)$  could represent "the probability of rolling a 3 on a number cube."

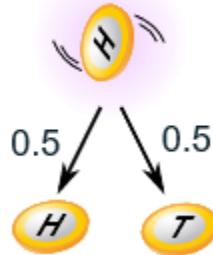
# Independent Events

- Events can be "Independent", meaning each event is **not affected** by any other events.
- A coin does not "know" that it came up heads before ... each toss of a coin is a perfect isolated thing.

Example: You toss a coin three times and it comes up "Heads" each time ... what is the chance that the next toss will also be a "Head"?

The chance is simply  $1/2$ , or 50%, just like ANY OTHER toss of the coin.

# Independent Events (contd.)



Example: Probability of 3 Heads in a Row

F

# Independent Events (contd.)

Example: your boss (to be fair) randomly assigns everyone an extra 2 hours work on weekend evenings between 4 and midnight.

What are the chances you get Saturday between 6 and 8?

# Summary

Probability is defined as the ratio of the number of desired outcomes (what you want to happen) to the number of total possible outcomes (what could possibly happen).

$$\text{Probability} = \frac{\text{Number of Desired Outcomes}}{\text{Total Number of Possible Outcomes}}$$

A **simple event** is when one event occurs. A **compound event** is when two or more events occur in a sequence. If any event does not affect the outcomes of the other events, then the events are **independent events**.

To determine the probability of independent events, multiply the probabilities of each event occurring by itself.

# Dependent Events

- Some events can be "dependent" ... which means they **can be affected by previous events**.

## Example: Drawing 2 Cards from a Deck

After taking one card from the deck there are **less cards** available, so the probabilities change!

**Let's look at the chances of getting a King.**

For the 1st card the chance of drawing a King is 4 out of 52

But for the 2nd card:

- If the 1st card was a King, then the 2nd card is **less** likely to be a King, as only 3 of the 51 cards left are Kings.
- If the 1st card was **not** a King, then the 2nd card is slightly **more** likely to be a King, as 4 of the 51 cards left are King.

- With** Replacement: the events are **Independent** (the chances don't change)
- Without** Replacement: the events are **Dependent** (the chances change)

# Independent vs Dependent Events

## Independent Events

Two or more events that occur in a sequence. If the outcome of any event **does not** affect the possible outcomes of the other event(s), then the events are independent.

## Dependent Events

Two or more events that occur in a sequence. If the outcome of any event **changes** the possible outcomes of the other event(s), then the events are dependent.

# Practice

1. A bag of letter tiles contains the following tiles.



What is the probability that Bradley will draw a vowel, replace the tile, and then draw another tile with a vowel?

# Solution

$$P(\text{vowel}) = \frac{\text{Number of Vowels}}{\text{Total Number of Tiles}} = \frac{5}{10} = \frac{5 \div 5}{10 \div 5} = \frac{1}{2}$$

$$P(\text{vowel then vowel}) = P(\text{vowel}) \times P(\text{vowel}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

# Tree Diagram

- When we have Dependent Events it helps to make a "Tree Diagram".

## Example: Marbles in a Bag

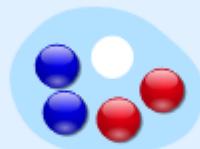
2 blue and 3 red marbles are in a bag.

What are the chances of getting a blue marble?

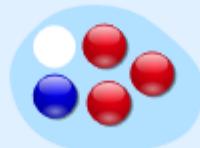
The chance is **2 in 5**

**But after taking one out** the chances change!

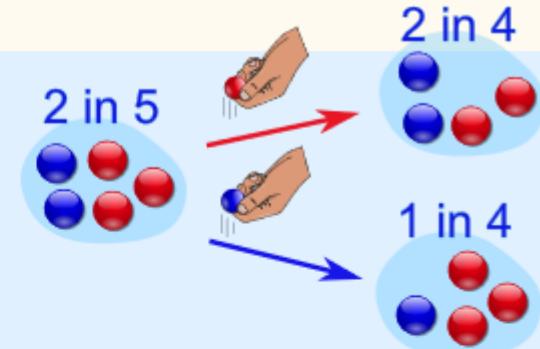
So the next time:



if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**

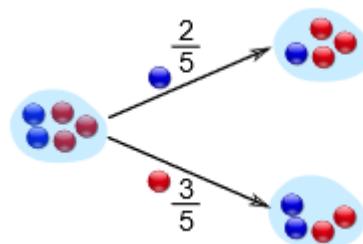


if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**

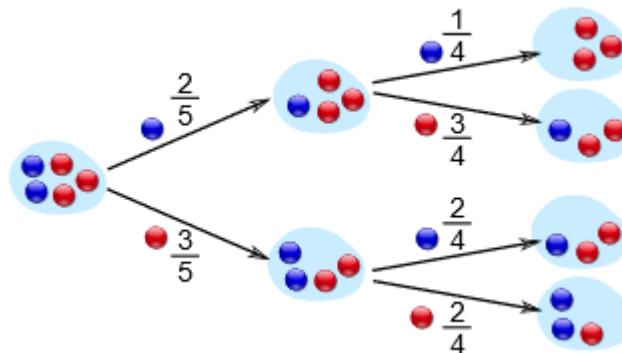


# Tree Diagram (contd.)

- There is a  $2/5$  chance of pulling out a Blue marble, and a  $3/5$  chance for Red:

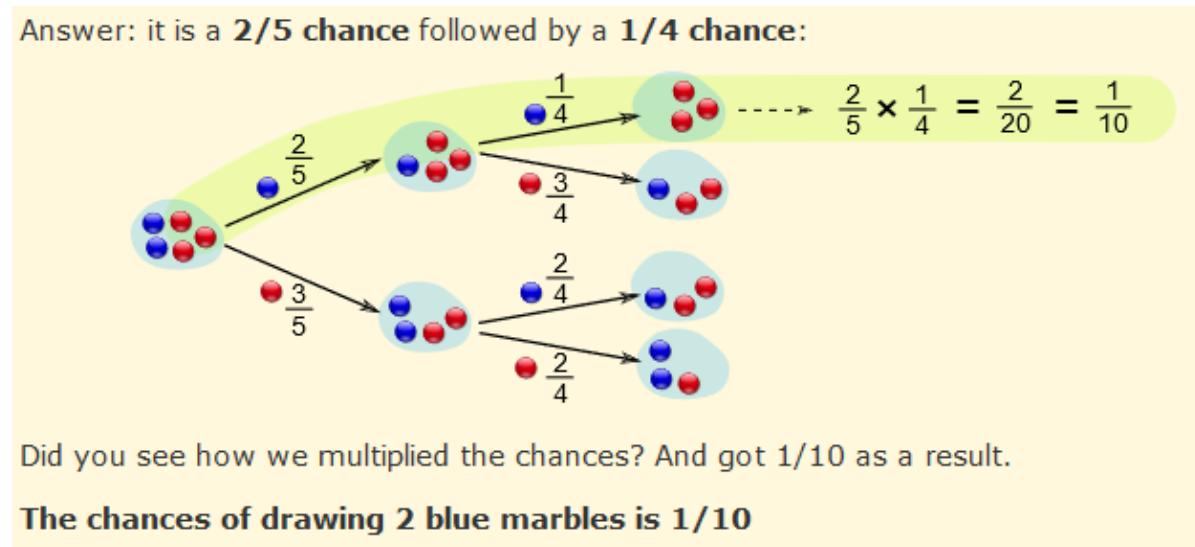


- We can go one step further and see what happens when we pick a second marble:



# Tree Diagram (contd.)

- If a blue marble was selected first there is now a  $1/4$  chance of getting a blue marble and a  $3/4$  chance of getting a red marble.
- If a red marble was selected first there is now a  $2/4$  chance of getting a blue marble and a  $2/4$  chance of getting a red marble.
- Now we can answer questions like "**What are the chances of drawing 2 blue marbles?**"



# Notations

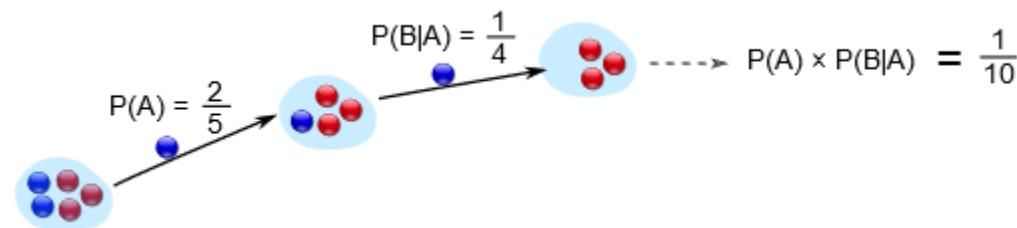
- $P(A)$  means "Probability Of Event A"
- In our marbles example Event A is "get a Blue Marble first" with a probability of  $2/5$ :  $P(A) = 2/5$
- And Event B is "get a Blue Marble second" ... but for that we have 2 choices:
  - If we got a **Blue Marble first** the chance is now  **$1/4$**
  - If we got a **Red Marble first** the chance is now  **$2/4$**

# Conditional Probability

- $P(B|A)$  means "Event B **given** Event A"
- In other words, event A has already happened, now what is the chance of event B?
- $P(B|A)$  is also called the "Conditional Probability" of B given A.
- And in our case:  $P(B|A) = 1/4$

# Conditional Probability (contd.)

- So the probability of getting **2 blue marbles** is:



- We write it as

$$P(\text{Event } A \text{ and Event } B) = P(\text{Event } A) \times P(\text{Event } B | \text{Event } A)$$

*"Probability Of"*  
↓      ↓  
 $P(\text{Event } A \text{ and Event } B)$  =  $P(\text{Event } A) \times P(\text{Event } B | \text{Event } A)$   
↑      ↑  
*"Given"*  
↓  
*Event A    Event B*

*"Probability of event A and event B equals  
the probability of event A times the probability of event B given event A"*

# Example

## Example: Drawing 2 Kings from a Deck

**Event A** is drawing a King first, and **Event B** is drawing a King second.

For the first card the chance of drawing a King is 4 out of 52 (there are 4 Kings in a deck of 52 cards):

$$P(A) = 4/52$$

But after removing a King from the deck the probability of the 2nd card drawn is **less** likely to be a King (only 3 of the 51 cards left are Kings):

$$P(B|A) = 3/51$$

And so:

$$\mathbf{P(A \text{ and } B) = P(A) \times P(B|A) = (4/52) \times (3/51) = 12/2652 = 1/221}$$

So the chance of getting 2 Kings is 1 in 221, or about 0.5%

# Conditional Probability (contd.)

- Start with:  $P(A \text{ and } B) = P(A) \times P(B|A)$
- Swap sides:  $P(A) \times P(B|A) = P(A \text{ and } B)$
- Divide by  $P(A)$ :  $P(B|A) = P(A \text{ and } B) / P(A)$

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$


*"The probability of event **B given A** equals  
the probability of event **A and B** divided by the probability of event **A**"*

You want to determine the probability that two events will happen. To do so, you need to determine the probability of the first event happening. Then, you need to determine that if the first event happens, what is the probability that the second event will happen.

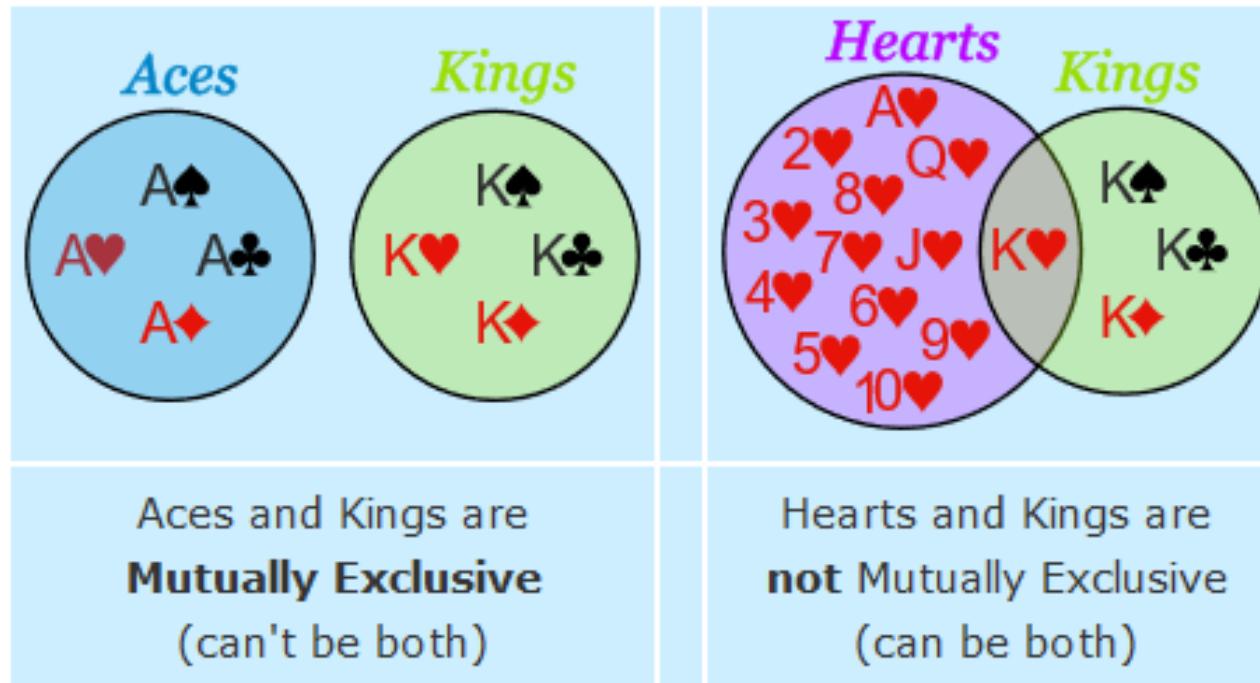
# Example

- 70% of your friends like Chocolate, and 35% like Chocolate AND like Strawberry.
- What percent of those who like Chocolate also like Strawberry?

# Mutually Exclusive Events

- **Mutually Exclusive:** can't happen at the same time.
- Examples:
  - Turning left and turning right are Mutually Exclusive (you can't do both at the same time)
  - Tossing a coin: Heads and Tails are Mutually Exclusive
  - Cards: Kings and Aces are Mutually Exclusive
- What is **not** Mutually Exclusive:
  - Turning left and scratching your head can happen at the same time
  - Kings and Hearts, because we can have a King of Hearts!

# Example



# Mutually Exclusive

- When two events (call them "A" and "B") are Mutually Exclusive it is **impossible** for them to happen together:  $P(A \text{ and } B) = 0$
- "*The probability of A and B together equals 0 (impossible)*"

Example: King AND Queen

A card cannot be a King AND a Queen at the same time!

- The probability of a King **and** a Queen is **0** (Impossible)

- The probability of A **or** B is the sum of the individual probabilities:  
$$P(A \text{ or } B) = P(A) + P(B)$$
- "*The probability of A or B equals the probability of A plus the probability of B*"

# Example

## Example: King OR Queen

In a Deck of 52 Cards:

- the probability of a King is  $1/13$ , so  $P(\text{King})=1/13$
- the probability of a Queen is also  $1/13$ , so  $P(\text{Queen})=1/13$

When we combine those two Events:

- The probability of a King **or** a Queen is  $(1/13) + (1/13) = 2/13$

Which is written like this:

$$P(\text{King or Queen}) = (1/13) + (1/13) = 2/13$$

So, we have:

- $P(\text{King and Queen}) = 0$
- $P(\text{King or Queen}) = (1/13) + (1/13) = 2/13$

# Special Notation

- Instead of "and" you will often see the symbol  $\cap$  (which is the "Intersection" symbol used in Venn Diagrams)
- Instead of "or" you will often see the symbol  $\cup$  (the "Union" symbol)
- So we can also write:
- $P(\text{King} \cap \text{Queen}) = 0$
- $P(\text{King} \cup \text{Queen}) = (1/13) + (1/13) = 2/13$

# Example

## Example: Scoring Goals

If the probability of:

- scoring no goals (Event "A") is **20%**
- scoring exactly 1 goal (Event "B") is **15%**



Then:

- The probability of scoring no goals **and** 1 goal is **0** (Impossible)
- The probability of scoring no goals **or** 1 goal is  $20\% + 15\% = \mathbf{35\%}$

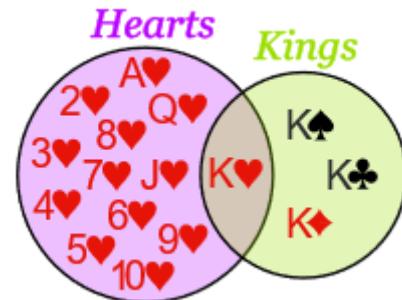
Which is written:

$$P(A \cap B) = 0$$

$$P(A \cup B) = 20\% + 15\% = 35\%$$

# Not Mutually Exclusive

- Example: Hearts and Kings



- Hearts **and** Kings together is only the King of Hearts:

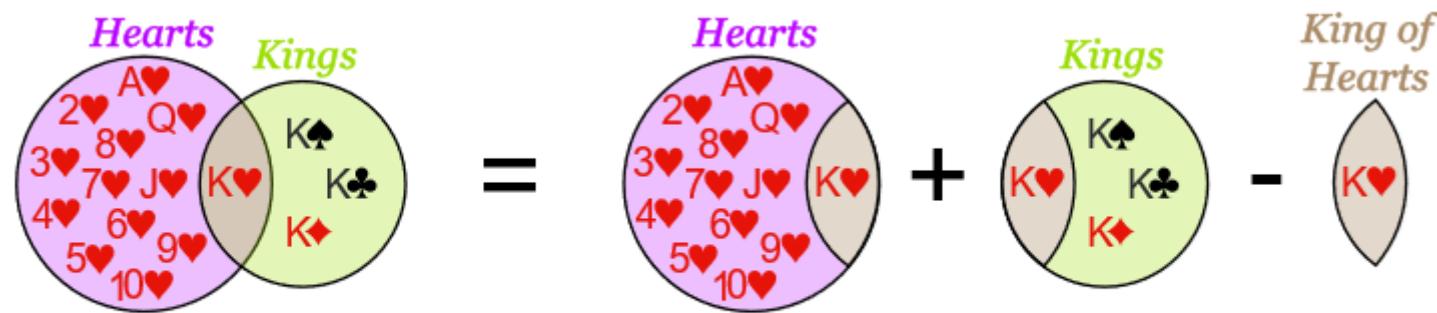
But Hearts **or** Kings is:

- all the Hearts (13 of them)
- all the Kings (4 of them)

**But that counts the King of Hearts twice!**

# Correction

- We correct our answer, by subtracting the extra "and" part:



$$16 \text{ Cards} = 13 \text{ Hearts} + 4 \text{ Kings} - \text{the 1 extra King of Hearts}$$

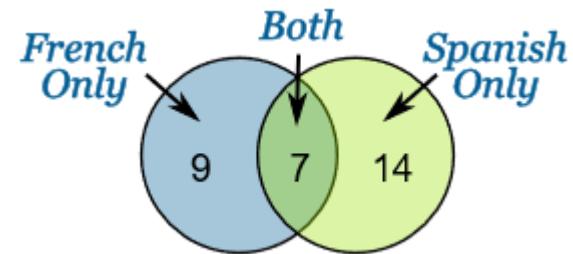
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
- "The probability of **A or B** equals the probability of **A plus the probability of B minus the probability of A and B**"
- Here is the **same formula**, but using **U** and **∩**:
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

# Example

- 16 people study French, 21 study Spanish and there are 30 altogether.  
Work out the probabilities!

## Example (contd.)

- $P(\text{French}) = 16/30$
- $P(\text{Spanish}) = 21/30$
- $P(\text{French Only}) = 9/30$
- $P(\text{Spanish Only}) = 14/30$
- $P(\text{French or Spanish}) = 30/30 = 1$
- $P(\text{French and Spanish}) = 7/30$



Lastly, let's check with our formula:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Put the values in:

$$30/30 = 16/30 + 21/30 - 7/30$$

Yes, it works!

# Relative Frequency Definition of Probability

- Suppose that the random experiment is repeated  $n$  times. If event  $A$  occurs  $n(A)$  times, then the probability of event  $A$ , denoted  $P(A)$ , is defined as

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$$

- $n(A)/n$  is called the relative frequency of event  $A$
- $0 \leq n(A)/n \leq 1$ , where  $n(A)/n = 0$  if  $A$  occurs in none of the  $n$  repeated trials and  $n(A)/n = 1$  if  $A$  occurs in all of the  $n$  repeated trials.

# Axiomatic Definition of Probability

- Let  $S$  be a finite sample space and  $A$  be an event in  $S$ . Then in the ***axiomatic*** definition, the probability  $P(A)$  of the event  $A$  is a real number assigned to  $A$  which satisfies the following three ***axioms*** :
- Axiom 1 :  $P(A) \geq 0$
- Axiom 2:  $P(S) = 1$
- Axiom 3:  $P(A \cup B) = P(A) + P(B)$  if  $A \cap B = \emptyset$
- If the sample space  $S$  is not finite, then axiom 3 must be modified.

# Elementary Properties

$$1. \quad P(\bar{A}) = 1 - P(A)$$

$$2. \quad P(\emptyset) = 0$$

$$3. \quad P(A) \leq P(B) \quad \text{if } A \subset B$$

$$4. \quad P(A) \leq 1$$

$$5. \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

6. If  $A_1, A_2, \dots, A_n$  are  $n$  arbitrary events in  $S$ , then

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) \\ &\quad - \cdots (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \end{aligned}$$

7. If  $A_1, A_2, \dots, A_n$  is a finite sequence of mutually exclusive events in  $S$  ( $A_i \cap A_j = \emptyset$  for  $i \neq j$ ), then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

and a similar equality holds for any subcollection of the events.

## Lecture - 3.

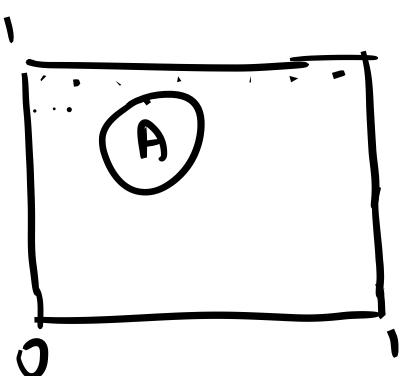
$$P(A) \leq 1$$

$$P(S) = 1$$

$$P(A \cup A^c) = P(A) + P(A^c)$$

$$1 = P(A) + P(A^c)$$

$$P(A) = 1 - P(A^c)$$



$$P(A) = \text{Area}(A)$$

$$S = \bigcup_{n,j} \{(n,j)\}$$

$$P(S) = 1 = P\left(\bigcup_{n,j} \{(n,j)\}\right)$$

$$\underset{\approx}{=} \sum_{n,j} P(\{(n,j)\})$$

$$= 0$$

$$A \cap B = \emptyset$$

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B | C) = P(A|C) + P(B|C)$$

$x = \text{First null}$

$y = 2^{\text{nd}} \dots$

$$\rightarrow B : \min(x, y) = 2$$

$$\rightarrow m = \max(x, y)$$

$$\textcircled{i} \quad P(m=1 | B) = 0$$

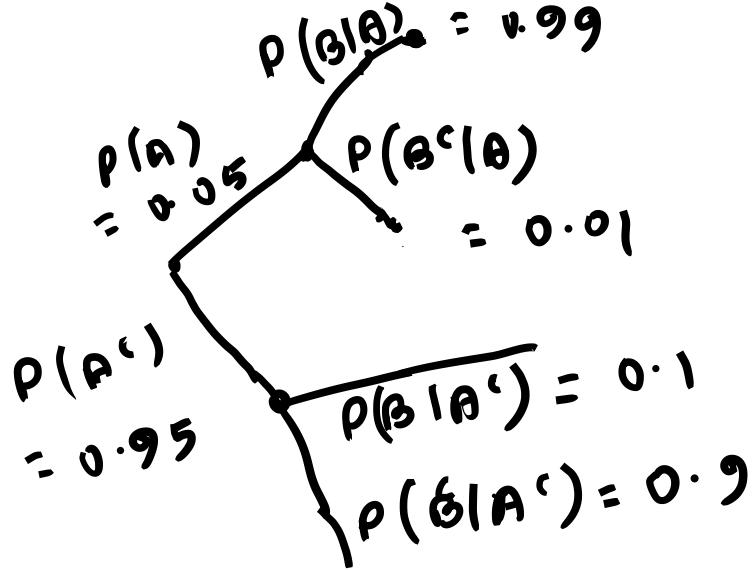
$$\textcircled{ii} \quad P(m=2 | B) = \frac{P(m=2 \wedge B)}{P(B)} = \frac{1/16}{5/16} = \frac{1}{5}$$

\*  $A = \{ \text{an aircraft present} \}$

$B = \{ \text{RADAR generates an alarm} \}$

$A^c = \{ \text{aircraft not present} \}$

$B^c = \{ \text{no Alarm} \}$



$$\begin{aligned}
 P(A \cap B) &= P(A) P(B|A) \\
 &= 0.05 \times 0.99 = 0.0495
 \end{aligned}$$

$$\begin{aligned}
 P(B) &= P(A) P(B|A) + P(A') P(B|A') \\
 &= 0.1445
 \end{aligned}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0.34$$

Total Probability Theorem:-



$$\begin{aligned}
 P(B) &= P(\theta_1) P(B|\theta_1) \\
 &\quad + P(\theta_2) P(B|\theta_2) \\
 &\quad + P(\theta_3) P(B|\theta_3) \\
 &\quad + P(\theta_4) P(B|\theta_4)
 \end{aligned}$$

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)}$$

$$= \frac{P(A_i) P(B | A_i)}{P(B)}$$

$$= \frac{P(A_i) P(B | A_i)}{\sum_j P(A_j) P(B | A_j)}$$

*Bayes Rule*

Cause

$$A_i \xrightarrow{P(B | A_i)} B$$

effect

$$B \xleftarrow{i \text{ forming}} A_i \quad P(A_i | B)$$

# Bayes Theorem

# Bayes Theorem

**Example:**

An internet search for "movie automatic shoe laces" brings up "Back to the future"

Has the search engine watched the movie? No, but it knows from lots of other searches what people are **probably** looking for.

And it calculates that probability using Bayes' Theorem.



Google movie automatic shoe laces

All Videos Images Shopping News More Settings Tools

About 1,22,00,000 results (0.94 seconds)



Back to the Future 2 - Nike Air 2015 Kicks - YouTube  
YouTube · sleepy6 guy ▶ 0:13

- Bayes' Theorem is a way of finding a probability when we know certain other probabilities.
- The formula is:  $P(A|B) = \frac{P(A) P(B|A)}{P(B)}$
- Which tells us: how often A happens given that B happens, written  $P(A|B)$ ,
- When we know: how often B happens given that A happens, written  $P(B|A)$  and how likely A is on its own, written  $P(A)$  and how likely B is on its own, written  $P(B)$ .

# Example

- Let us say  $P(\text{Fire})$  means how often there is fire, and  $P(\text{Smoke})$  means how often we see smoke, then:
  - $P(\text{Fire}|\text{Smoke})$  means how often there is fire when we can see smoke  
 $P(\text{Smoke}|\text{Fire})$  means how often we can see smoke when there is fire
- So the formula kind of tells us "forwards"  $P(\text{Fire}|\text{Smoke})$  when we know "backwards"  $P(\text{Smoke}|\text{Fire})$

Example: If dangerous fires are rare (1%) but smoke is fairly common (10%) due to barbecues, and 90% of dangerous fires make smoke then:

$$\begin{aligned} P(\text{Fire}|\text{Smoke}) &= \frac{P(\text{Fire}) P(\text{Smoke}|\text{Fire})}{P(\text{Smoke})} \\ &= \frac{1\% \times 90\%}{10\%} \\ &= 9\% \end{aligned}$$

So the "Probability of dangerous Fire when there is Smoke" is 9%

## Example: Picnic Day

You are planning a picnic today, but the morning is cloudy

- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)



### What is the chance of rain during the day?

We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.

The chance of Rain given Cloud is written  $P(\text{Rain}|\text{Cloud})$

So let's put that in the formula:

$$P(\text{Rain}|\text{Cloud}) = \frac{P(\text{Rain}) P(\text{Cloud}|\text{Rain})}{P(\text{Cloud})}$$

- $P(\text{Rain})$  is Probability of Rain = 10%
- $P(\text{Cloud}|\text{Rain})$  is Probability of Cloud, given that Rain happens = 50%
- $P(\text{Cloud})$  is Probability of Cloud = 40%

$$P(\text{Rain}|\text{Cloud}) = \frac{0.1 \times 0.5}{0.4} = .125$$

Or a 12.5% chance of rain. Not too bad, let's have a picnic!

# Bayes Theorem with Two Cases

- One of the famous uses for Bayes Theorem is False Positives and False Negatives.

## Example: Allergy or Not?

Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:

- For people that **really do** have the allergy, the test says "Yes" **80%** of the time
- For people that **do not** have the allergy, the test says "Yes" **10%** of the time ("false positive")



If 1% of the population have the allergy, and **Hunter's test says "Yes"**, what are the chances that Hunter really has the allergy?

# Bayes Theorem with Two Cases (contd.)

- We want to know the chance of having the allergy when test says "Yes", written  $P(\text{Allergy}|\text{Yes})$

$$P(\text{Allergy}|\text{Yes}) = \frac{P(\text{Allergy}) P(\text{Yes}|\text{Allergy})}{P(\text{Yes})}$$

- $P(\text{Allergy})$  is Probability of Allergy = 1%
- $P(\text{Yes}|\text{Allergy})$  is Probability of test saying "Yes" for people with allergy = 80%
- $P(\text{Yes})$  is Probability of test saying "Yes" (to anyone) = ??%
- Oh no! We **don't know** what the **general** chance of the test saying "Yes" is ..... but we can calculate it by adding up those **with**, and those **without** the allergy:
  - 1% have the allergy, and the test says "Yes" to 80% of them
  - 99% do **not** have the allergy and the test says "Yes" to 10% of them

# Bayes Theorem with Two Cases (contd.)

- Let's add that up:
  - $P(\text{Yes}) = 1\% \times 80\% + 99\% \times 10\% = 10.7\%$
- Which means that about 10.7% of the population will get a "Yes" result.
- So now we can complete our formula:
- In fact we can write a special version of the Bayes' formula just for things like this:

$$P(\text{Allergy}|\text{Yes}) = \frac{1\% \times 80\%}{10.7\%} = 7.48\%$$

$$P(\text{Allergy}|\text{Yes}) = \text{about } 7\%$$

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{not } A)P(B|\text{not } A)}$$

# For more cases

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) + \dots \text{etc}}$$

Example: The Art Competition has entries from three painters: Pam, Pia and Pablo



- Pam put in 15 paintings, 4% of her works have won First Prize.
- Pia put in 5 paintings, 6% of her works have won First Prize.
- Pablo put in 10 paintings, 3% of his works have won First Prize.

What is the chance that Pam will win First Prize?

## Lecture - 4

### Random Variable:-

- A fn. that maps sample point to real no.

$$x : \Omega \rightarrow \mathbb{R}$$

numerical value  $x$

Uppercase letter  
for R.V.

lowercase letter  
for real no.

### PMF:-

$$P_x(x) = P(x=x)$$

$$= P((\omega \in \Omega \text{ s.t. } x(\omega)=x))$$

$$- P_x(x) \geq 0$$

$$- \sum_n P_x(n) = 1$$

### Geometric PMF:-

$$P_x(k) = (1-p)^{k-1} p$$

$x$ : no. of tosses until first head occurs.

$x$ : Geometric R.V.

$\neq$  1 Two independent rolls of a fair tetrahedral die.

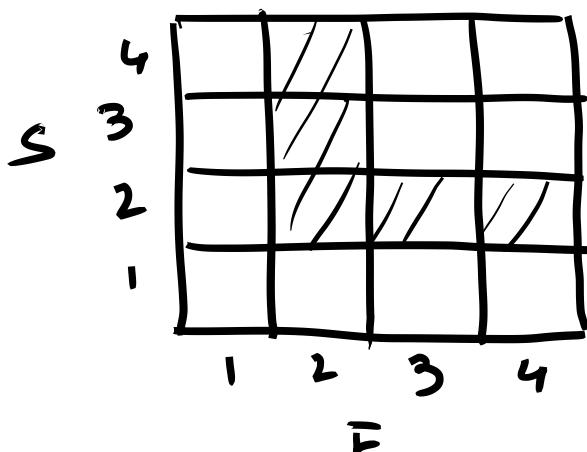
$F$ : Out-Come of the first roll

$S$ : " " " " " " 2nd "

$$X = \min(F, S)$$

$$P_X(2) = ? \quad \frac{5}{16} .$$

$$P_X(0) = 0$$



\*  $X$ : no. of heads in  $n$  independent coin tosses.

$$P(H) = p$$

$$\text{let } n = 4$$

$$k = 2$$

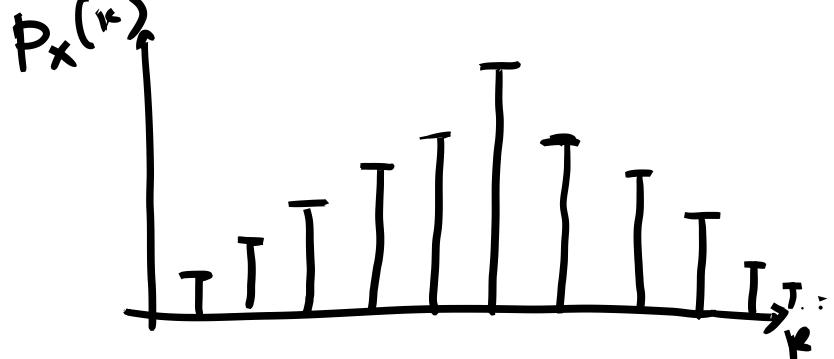
$$\begin{aligned} P_X(2) &= P(HHHT) + P(HHTT) + P(HTTH) \\ &\quad + P(THHH) + P(THTH) + P(TTHH) \end{aligned}$$

$$= 6 p^2 (1-p)^2$$

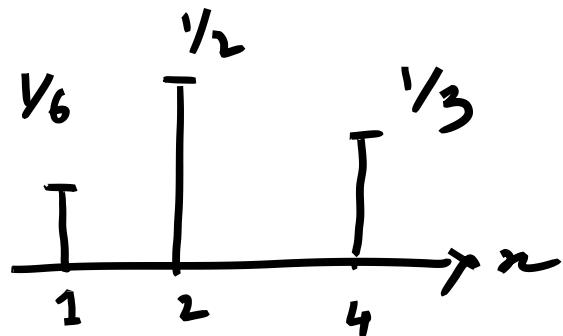
$$P_X(2) = \binom{4}{2} p^2 (1-p)^2$$

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Binomial PMF



\* Expectation:-

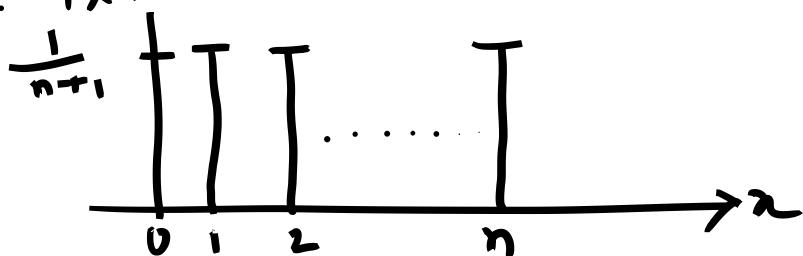


$$\frac{1}{6} \cdot 1 + \frac{1}{2} \cdot 2 + \frac{1}{3} \cdot 4 = 2.5$$

$$E[X] = \sum_n P_X(n) \cdot n$$

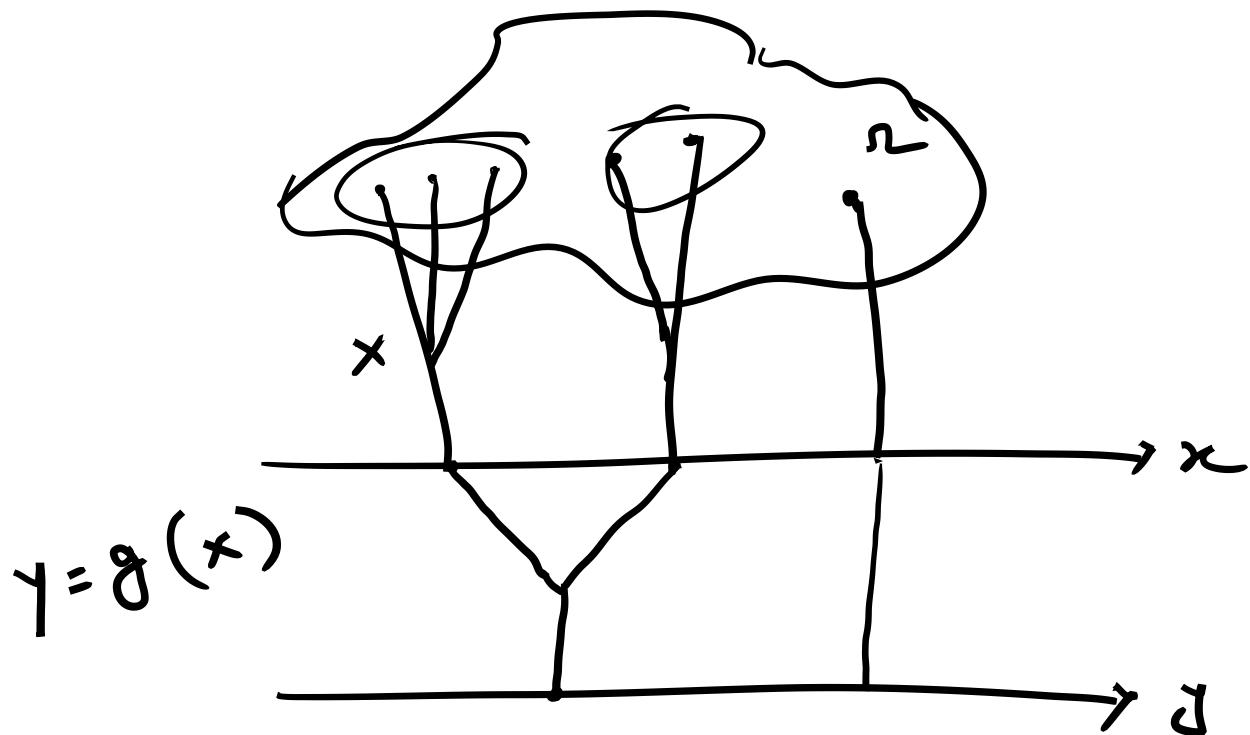
- Average in large no. of repetitions of the experiment
- Centre of gravity

Example:-  $P_X(m)$



$$f[x] = 0 \times \frac{1}{n\pi} + 1 \times \frac{1}{n\pi} + \dots + n \times \frac{1}{n\pi} = \text{Answer}$$

$$F[\bar{x}] = \frac{n}{2} \cdot$$



$$E[y] = \sum_j y P_y(y)$$

$$E[y] = \sum_n g(x) P_x(x)$$

## Lecture - 5

Expectation :-

$$E[x] = \sum_{x_i} x_i P(x_i)$$

$$= 2.5$$



$$E[x] = 0 \times \frac{1}{n+1} + 1 \times \frac{1}{n+1} + 2 \times \frac{1}{n+1} + \cdots + n \times \frac{1}{n+1}$$

$$E[x] = \frac{n}{2}$$

\*

$$Y = g(X)$$

$$E[Y] = \sum_z z p_Y(z)$$

$$E[Y] = \sum_x g(x) p_X(x)$$

Properties :- ①  $E[\alpha] = \alpha$

i)  $E[\alpha x] = \alpha E[x]$

ii)  $E[\alpha x + \beta] = E[\alpha x] + E[\beta] = \alpha E[x] + \beta$

$$* \quad g(x) = x^2$$

$$E[g(x)] = E[x^2] = \sum_n x^2 p_x(n)$$

second moment

$$* \quad g(x) = (x - E[x])^2$$

$$\begin{aligned} \text{Var}(x) &= E[g(x)] = \sum_n (n - E[x])^2 p_x(n) \\ &= E[(x - E[x])^2] \\ &= E[x^2] - (E[x])^2 \end{aligned}$$

$$\text{Properties: } \textcircled{i} \quad \text{Var}(x) \geq 0$$

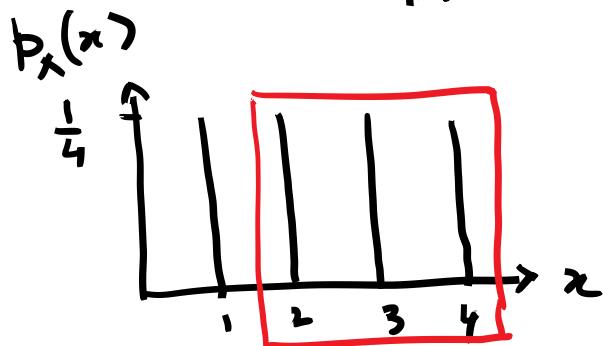
$$\textcircled{ii} \quad \text{Var}(\alpha x + \beta) = \alpha^2 \text{Var}(x)$$

$$\sigma = \sqrt{\text{Var}(x)}$$

(Standard deviation)

Conditional PMF :-

$$P_{x|A}(x) \triangleq P(x=x|A)$$



$$\text{let } A = \{x \geq 2\}$$

$$P_{x|A}(x) = \frac{1}{3} \quad [x: 2, 3, 4]$$

Conditional Expectation:-

$$E[x|A] = \sum_x x P_{x|A}(x)$$

$$E[x|A] = \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 4 = 3$$

$$E[g(x)|A] = \sum_x g(x) P_{x|A}(x)$$

Geometric PMF :-

$x$ : no. of independent coin tosses until first head.

$$P_x(k) = (1-p)^{k-1} p \quad k=1, 2, \dots$$

$$E[x] = \frac{1}{p}$$

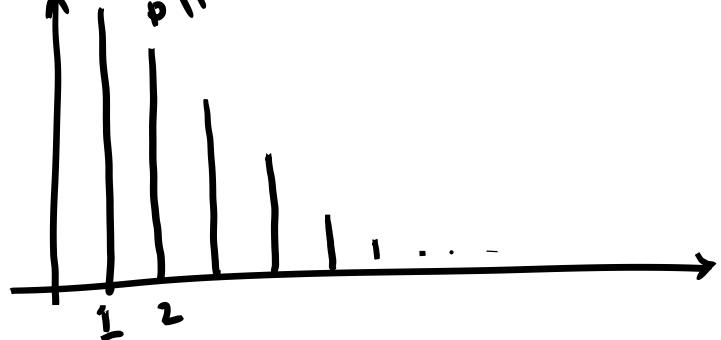
$x : \underbrace{\text{TT}}_{\text{2}}, \underbrace{\text{TTT} \dots}_{x-2}, \dots$

$y : \text{TTT} \dots H$

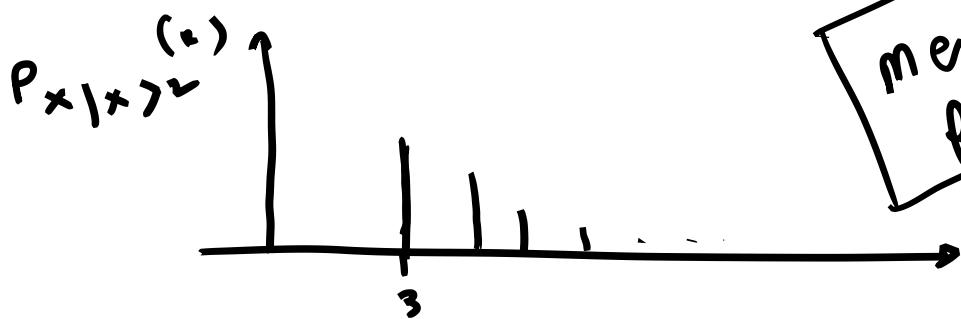
$$P_y(k) = (1-p)^{k-1} p \quad k \geq 1$$

$$P(x_{-2} | x_{>2}) = (1-p)^{k-1} p$$

$$P_x(k)$$



memoryless property



## Lecture - 6.

$x$ : Geometric R.V.

$$P_x(k) = (1-p)^{k-1} p \quad k=1, 2, \dots$$

$A_1, A_2, \dots, A_n$

$$P(B) = P(A_1) P(B|A_1) + \dots + P(A_n) P(B|A_n)$$

$$P_x(n) = P(A_1) P_{x|A_1}(n) + \dots + P(A_n) P_{x|A_n}(n)$$

$$\sum_n n P_x(n) = \sum_n n (P(A_1) P_{x|A_1}(n) + \dots + P(A_n) P_{x|A_n}(n))$$

$$E[x] = P(A_1) E[x|A_1] + \dots + P(A_n) E[x|A_n]$$

Geometric :-  $A_1 : \{x=1\}$

$$A_2 : \{x > 1\}$$

$$E[x] = P(x=1) E[x|x=1] + P(x>1) \underline{E[x|x>1]}$$

$$\begin{aligned} E[x|x>1] &= E[x|x-1>0] = E[x-1|x-1>0] \\ &= E[x]+1 \end{aligned}$$

$$\rightarrow E[x] = p \cdot 1 + (1-p)(E[x]+1)$$

$$E[X] = p + E[X] + 1 - pE[X] - p$$

-  $P(E[X]) = 1$

$$E[X^2] = \frac{1}{p}$$

Joint PMF  $P_{X,Y}(x,y) = P(X=x \text{ and } Y=y)$

$y$	4	3	2	1
1	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	
2	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
3			$\frac{1}{20}$	
4				$\frac{1}{20}$
$x$	1	2	3	4

$$P_{X,Y}(2,3) = \frac{4}{20}$$

-  $\sum_x \sum_y P_{X,Y}(x,y) = 1$

- Marginal PMF  $P_X(x) = \sum_y P_{X,Y}(x,y)$

$$P_X(3) = \frac{2}{20} + \frac{1}{20} + \frac{3}{20} = \frac{6}{20}$$

-  $P_{X|Y} = P(X=x | Y=y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

Conditional PMF

$$y=2$$

$$0, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}$$

$$- \sum_n P_{x,y} (x|y) = 1$$

$$- P_{x,y}(x,y) = P_x(x) P_{y|x}(y|x)$$

Multivariate RV.

$$P_{x,y,z}(x,y,z)$$

$$- P_x(x) = \sum_{y,z} P_{x,y,z}(x,y,z)$$

$$- P_{x,y,z}(x,y,z) = P_x(x) P_{y|x}(y|x) P_{z|y,x}(z|y,x)$$

$$- P_{x,y,z}(x,y,z) = P_x(x) P_y(y) P_z(z) \quad \begin{cases} \text{if } x, y, z \text{ independent} \\ \text{indep.} \end{cases}$$

For Independent Events:

$$- P_{x|y}(x|y) = P_x(x)$$

$$- P_{x|y,z}(x|y,z) = P_x(x)$$

\*

	x	1	2	3	4
y		$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	
z		$\frac{1}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
x	1	$\frac{1}{20}$			
y	2				
z	3				
4					

(i)  $P_{x,y}(2,1) = \frac{1}{20}$

(ii) x, y, z are not independent.

(iii)  $A: x \leq 2 \text{ and } y \geq 3$

$$P_{x,y|A}(x,y) =$$

$$P_{x|B}(x,y)$$

$\frac{1}{3}$	$\frac{1}{9}$	$\frac{2}{9}$
$\frac{2}{3}$	$\frac{2}{9}$	$\frac{4}{9}$

Conditionally independent.

\* i) If  $x = y$   $\text{Var}(x+y) = ?$   
 $\text{Var}(x)$

ii) If  $x = -y$   $\text{Var}(x+y) = 0$

iii) If  $x$  &  $y$  are independent and  
 $z = x - 3y$

$$\text{Var}(z) = \text{Var}(x) + 9\text{Var}(y)$$

Binomial Mean & Variance:

$x$  = # of successes in  $n$  independent trials.

$p$  = probability of success.

$$E[x] = ? \quad \text{Var}(x) = ?$$

$$E[x] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$$

$x, y$   
Dependent

$$x_i = \begin{cases} 1 & \text{if success in trial } i \\ 0 & \text{otherwise} \end{cases}$$

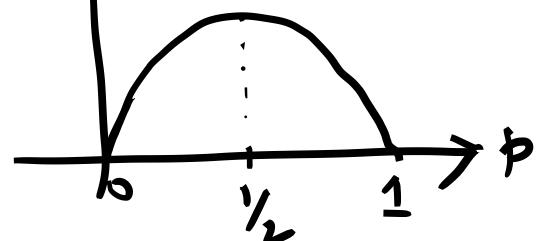
$$\sum_i x_i = X$$

$$E[X] = \sum_i E[x_i] = \sum_{i=1}^n (1 \cdot p + 0 \cdot (1-p)) = np$$

$$\text{var}(x_i) = p(1-p)^2 + (1-p)(0-p)^2$$

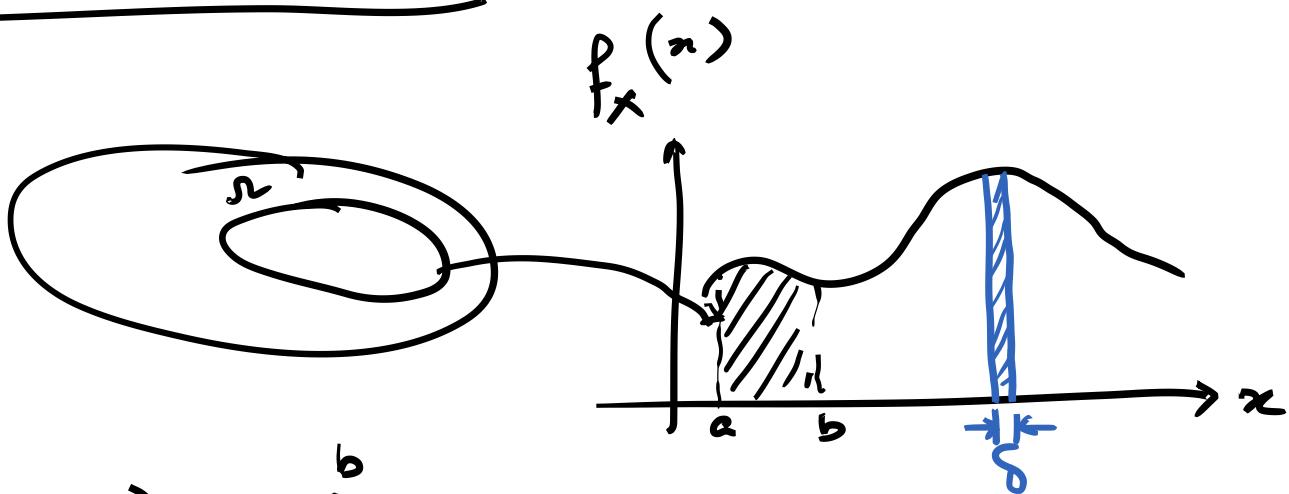
$$\begin{aligned}\text{var}(x_i) &= E[x_i^2] - (E[x_i])^2 \\ &= p - p^2 = p(1-p)\end{aligned}$$

$$\text{var}(X) = \sum_i \text{var}(x_i) = n p (1-p)$$



## Lecture - 7

### Continuous R.V.:-



i)

$$P(a \leq x \leq b) = \int_a^b f_x(x) dx$$

ii)

$$P(x=a) = 0$$

iii)

$$f_x(x) \geq 0$$

iv)

$$\int_{-\infty}^{\infty} f_x(x) dx = 1$$

v)

$$P(x \leq x \leq x+\delta) = \int_x^{x+\delta} f_x(x) dx = f_x(x)\delta$$

$$f_x(x) = \frac{P(x \leq x \leq x+\delta)}{\delta}$$

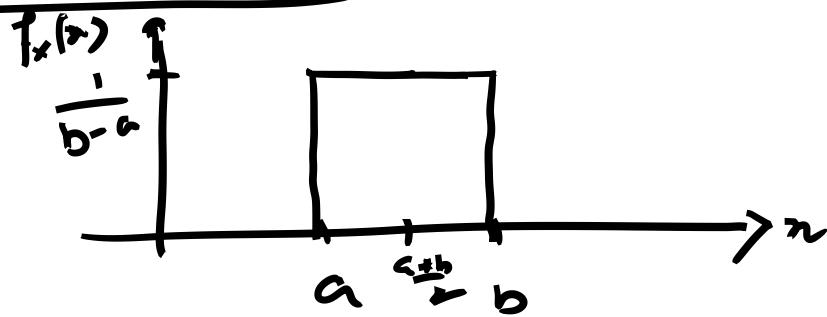
vi)

$$P(x \in \beta) = \int_{\beta} f_x(x) dx$$

## Mean and Variance :-

- $E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$
- $E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx$
- $\text{var}(x) = \int_{-\infty}^{\infty} (x - E[x])^2 f_x(x) dx$

## Cont. Uniform R.V.



$$f_x(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$= 0 \quad , \quad \text{otherwise.}$$

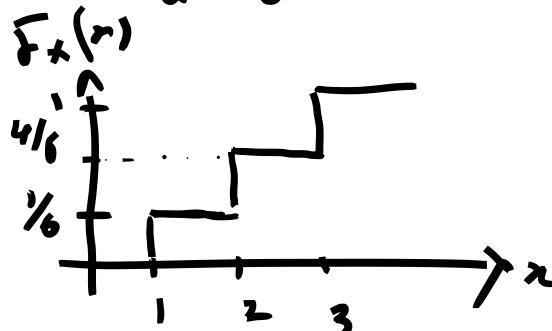
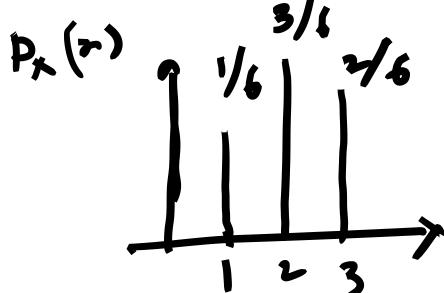
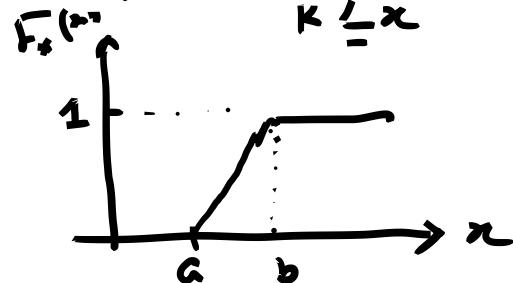
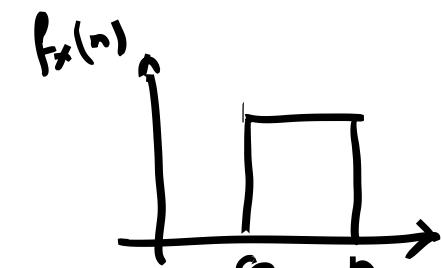
$$E[x] = \int_a^b \frac{1}{b-a} x dx = \frac{a+b}{2} .$$

$$\sigma_x^2 = \frac{(b-a)^2}{12}$$

# Cumulative Distribution Function (CDF)

Def.  $F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(z) dz$

Def:  $F_X(x) = P(X \leq x) = \sum_{k \leq x} P_X(k)$

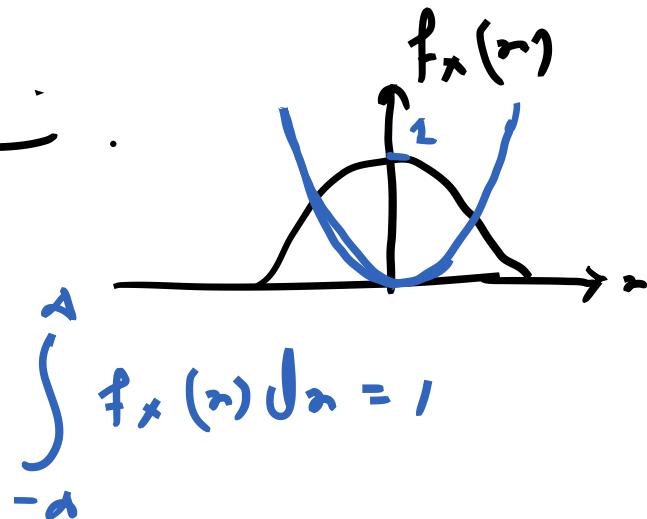


Gaussian (or normal) PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$E[X] = 0$$

$$\text{Var}(X) = 1$$



General normal:

$$\mu, \sigma^2$$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\frac{(x - \mu)^2}{\sigma^2} \rightarrow \text{Small}$$

$$\frac{(x - \mu)^2}{\sigma^2} \rightarrow \text{Large}$$

- Parabola goes up quickly
- normal falls very fast
- Normal density.

Let  $y = a x + b \quad [x \sim N(\mu, \sigma^2)]$

$$E[y] = a E[x] + b = a \mu + b$$

$$\text{Var}(y) = a^2 \text{Var}(x) = a^2 \sigma^2$$

$$y \sim N(a\mu + b, a^2 \sigma^2)$$

\* CDF =  $\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

\*  $x \sim N(\mu, \sigma^2) \quad | \quad \theta x - \mu \sim N(0, \sigma^2)$

④  $\frac{x - \mu}{\sigma} \sim N(0, 1)$

$$\mu = 2 \quad x \sim N(2, 16)$$

$$6^2 = 16$$

$$P(x \leq 3)$$

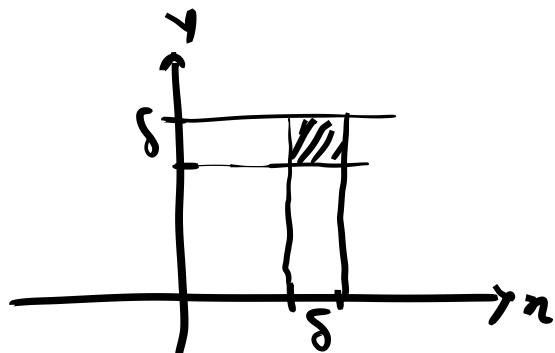
$$= P\left(\frac{x-2}{4} \leq \frac{3-2}{4}\right)$$

$$= CDF(0.25) = 0.5987$$

multiple Rv:  $f_{x,y}(x,y)$

$$P((x,y) \in S) \triangleq \iint_S f_{x,y}(x,y) dx dy$$

$$\textcircled{i} \quad \iint_{-d-d}^{d+d} f_{x,y}(x,y) dx dy = 1$$



$$\textcircled{ii} \quad f_{x,y} > 0$$

$$\textcircled{iii} \quad P(x \leq X \leq x+\delta, y \leq Y \leq y+\delta) \approx f_{x,y}(x,y) \delta^2$$

$$\textcircled{iv} \quad E[g(x,y)] = \int_{-d}^d \int_{-d}^d g(x,y) f_{x,y}(x,y) dx dy$$

$$\textcircled{v} \quad \int_{-\infty}^{\infty} f(y) dy = f_y(\delta)$$

(v)  $f_{x,y}(x,y) = f_x(x) f_y(y)$  for all  $x,y$

\* Conditioning :-  $P(x \leq X \leq x+d) \approx f_x(x) d$

(i)  $P(x \leq X \leq x+d | Y=y) \approx f_{x|y}(x|y) d$

(ii)  $f_{x|y}(x|y) = \frac{f_{x,y}(x,y)}{f_y(y)}$  if  $f_y(y) > 0$

(iii)  $f_{x,y}(x,y) = f_x f_y$  [if  $X, Y$  are independent]

$$f_{x|y}(x|y) = f_x(x)$$

## Lecture - 8

### 'The Central Limit Theorem:-'

Let  $x_1, x_2, \dots$  be a sequence of i.i.d RVs each having mean  $\mu$  and variance  $\sigma^2$ .

Then the distribution of  $\frac{x_1 + x_2 + \dots + x_n - n\mu}{\sigma\sqrt{n}}$  tends to standard normal as  $n \rightarrow \infty$ .

### Random Process:-

$$\{x(t), t \in T\} \quad \omega \in \Omega$$

$$\{x(t, \omega), t \in T, \omega \in \Omega\}$$

$\Rightarrow$  Fixed  $t = t_k$

$$x(t_k, \omega) = x_k(\omega)$$

Fixed  $\omega = \omega_i$ :  $x(t, \omega_i) = x_i(t)$

- \* If  $x(t)$  is a discrete-time process then  $x(t)$  is specified by a collection of PMFs.

$$P_x(x_1, \dots, x_n, t_1, \dots, t_n) = P\{x(t_1) = x_1, \dots\}$$

\* If  $x(t)$  is a continuous-time R.P.  
PDF

$$f_x(x_1, \dots, x_n, t_1, \dots, t_n) = \frac{\partial^n F_x(x_1, x_2, \dots, x_n, t_1, \dots, t_n)}{\partial x_1 \dots \partial x_n}$$

\*  $n^{\text{th}}$  order moment  $m_n = \int x^n f_x(x) dx$

$$(m_m)_n = \int \int x^m y^n f_{x,y}(x,y) dx dy$$

$$* f_x(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(\underline{x} - \bar{x})^T \underline{\underline{C}}^{-1} (\underline{x} - \bar{x})}$$

$$\sigma_{ij} = E[x_i, x_j]$$

For R.P. :-  $\underbrace{m_x(t)}_{\text{ensemble average of } x(t)} = E[\bar{x}(t)]$

ensemble average of  $x(t)$

$$R_x(t_i, t_j) = E[x(t_i) x(t_j)]$$

$$= \int \dots \int \underline{x_i} \underline{x_j} f_x(x_1, x_2, \dots, x_i, x_j, \dots, x_n) dx_1 dx_2 \dots dx_n$$

$$t_i = t$$

$$t_j = t + \tau$$

$$R_x(t_i, t_j) = E[x(t_i) x(t_j)]$$

=  $R(\tau)$  : Auto correlation  
fn.

Autocovariance fn:-

$$k_x(t_i, t_j) = \text{Cov}[x(t_i) x(t_j)]$$

$$= E \left\{ [x(t_i) - \mu_x(t_i)] [x(t_j) - \mu_x(t_j)] \right\}$$

$$= R_x(t_i, t_j) - \mu_x(t_i) \mu_x(t_j)$$

① Stationary RP :-

$$\{x(t_i) + \epsilon_T\}_{t_i \in T} \quad [1, 2, \dots n]$$

$$F_x(x_1, \dots, x_n; t_1, \dots, t_n) = F_x(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau)$$

for all  $\tau$ .

$K \leq n$

$x(t)$  is stationary to order  $K$ .

If  $K=2$  : wide-sense stationary rf.  
(WSS)

①  $E[x(t)] = a$

②  $R_x(t_i, t_j) = R_x(\tau) \quad [t_i = t_j + \tau]$

\* Ergodic process:-

$$\langle x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(\tau) d\tau$$

$$\bar{R}_x(\tau) = \langle x(t) x(t+\tau) \rangle$$

$$= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(\tau) x(t+\tau) d\tau$$

$$R_x(\tau) = E[x(t)x(t+\tau)]$$

Power Spectral Density :-

$$S_x(w) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau.$$

Inverse F.T.

$$\text{Auto correlation } R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(w) e^{j\omega\tau} dw$$

\* White noise :-

A continuous time white noise process  $n(t)$  is a WSS zero-mean continuous time RP whose auto correlation  $f_n = \delta$

given by

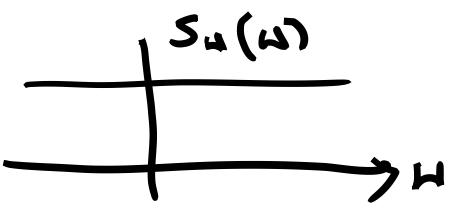
$$R_n(\tau) = \sigma^2 \delta(\tau)$$

$$\left[ \int_{-\infty}^{\infty} \delta(\tau) d\tau \right] = 1$$

$$S_n(w) = \int_{-\infty}^{\infty} R_n(\tau) e^{-j\omega\tau} d\tau$$

$$= \sigma^2 \int_{-\infty}^{\infty} \delta(\tau) e^{-j\omega\tau} d\tau$$

$$= 6^2$$



## Estimation Theory

Parameter space  $\theta$   $\xrightarrow[\text{Probabilistic mapping}]{P(n; \theta)}$  Observation space ( $x$ )



Unbiased Estimator:-

$$\boxed{E[\hat{\theta}] = \theta} + Q,$$

$$a < \theta < b$$

#  $x[n] = A + w[n]$   $n = 0, 1, \dots, N-1$

$A \rightarrow$  the Parameter.

$w[n] \rightarrow$  n.a.n  $\mathcal{N}(0, \sigma^2)$

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

$$E[\hat{A}] = \frac{1}{N} \sum_{n=0}^{N-1} E[x[n]]$$

$$E[x[n]] = A = \frac{1}{N} \sum_{n=0}^{N-1} A$$

$$= \frac{NA}{N}$$

$$= A$$

So, sample mean estimator is an unbiased estimator.

\*  $\hat{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x[n]$  Biased Estimator

$$E[\hat{A}] = \frac{NA}{2N} = \frac{A}{2}$$

$$\begin{aligned} &= A && \text{if } A = 0 \\ &\neq A && \text{if } A \neq 0 \end{aligned}$$

\*  $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n\}$   
 $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i$

$$\rho_{x,y} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$$

$$E[\hat{\theta}] = \theta \quad \left[ \begin{array}{l} \text{if } \hat{\theta}_i \text{ unbiased} \\ \text{some variance} \\ \text{& known} \end{array} \right]$$

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{n} \quad \left[ \text{var}(\hat{\theta}_i) = \sigma^2 \right]$$

$$n \rightarrow \infty \quad \hat{\theta} \rightarrow \theta$$

\* For biased estimator

$$E(\hat{\theta}_i) = \theta + \underbrace{b(\theta)}_{\text{Bias term}}$$

if  $n \rightarrow \infty$ ,  $E(\hat{\theta}_i) \neq \theta$

$$\begin{aligned} \text{mse}(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\ &= E \left\{ [(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2 \right\} \\ &= E \left\{ (\hat{\theta} - E(\hat{\theta}))^2 + 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2 \right\} \\ &= \text{var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 \\ &= \text{var}(\hat{\theta}) + b^2(\theta) \quad [\text{for Biased estimator}] \end{aligned}$$

#  $\check{A} = a - \frac{1}{n} \sum_{n=0}^{n-1} n [n] \quad [a = \text{constant not a fn of data}]$

$$E(\check{A}) = a A$$

$$\text{var}(\check{A}) = a^2 \sigma^2 / N$$

$$MSR(\hat{A}) = \frac{\alpha^2 \sigma^2}{2} + [\alpha_A - A]^2$$

$$\frac{d MSR(\hat{A})}{d\alpha} = \frac{2\alpha \sigma^2}{2} + 2(\alpha - 1)A \rightarrow 0$$

$$\alpha_{opt} = \frac{A^2}{A^2 + \frac{\sigma^2}{2}}.$$

CRLB:-

Lecture - 11/12

Reynolds  
Condition or  
Correlation

$$\textcircled{i} \quad E \left[ \frac{\partial \ln P(x; \theta)}{\partial \theta} \right] = 0 \quad \forall \theta.$$

Bound  $\text{Var}(\hat{\theta}) \geq \frac{1}{E \left[ \frac{\partial^2 \ln P(x; \theta)}{\partial \theta^2} \right]}$

$\textcircled{ii} \quad \frac{\partial \ln P(x; \theta)}{\partial \theta} = I(\theta) (\hat{\theta} - \theta)$

MVUE  $\hat{\theta} = g(x)$

$$\text{Var}(\hat{\theta}_{\text{MVUE}}) = \frac{1}{E \left[ \frac{\partial^2 \ln P(x; \theta)}{\partial \theta^2} \right]} = \frac{1}{I(\theta)}$$

$I(\theta)$  = Fisher Information.

Proof:-  $\textcircled{i} \quad E \left[ \frac{\partial \ln P(x; \theta)}{\partial \theta} \right]$

$$= \int \frac{\partial \ln P(x; \theta)}{\partial \theta} p(x; \theta) dx.$$

$$\begin{aligned}
 &= \int \frac{1}{\rho(x; \theta)} \frac{\partial \rho(x; \theta)}{\partial \theta} \rho(x; \theta) dx \\
 &= -\frac{\partial}{\partial \theta} \int \rho(x; \theta) dx \\
 &= 0
 \end{aligned}$$

i)

$$\alpha = g(\theta)$$

$$E[\hat{x}] = \alpha = g(\theta)$$

$$\int \hat{x} \rho(x; \theta) dx = g(\theta)$$

$$\int \hat{x} \frac{\partial \rho(x; \theta)}{\partial \theta} dx = \frac{\partial g(\theta)}{\partial \theta}$$

\*  $\frac{\partial \ln \rho(x; \theta)}{\partial \theta} \rho(x; \theta) = \frac{\partial \rho(x; \theta)}{\partial \theta}$

$$\int \hat{x} \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \rho(x; \theta) dx = \frac{\partial g(\theta)}{\partial \theta} - 0$$

$$\left( E \left[ \frac{\partial \ln P(x; \theta)}{\partial \theta} \right] = 0 \right)$$

$$\int \frac{\partial \ln P(x; \theta)}{\partial \theta} P(x; \theta) dx = 0$$

$$\Rightarrow \int x \frac{\partial \ln P(x; \theta)}{\partial \theta} P(x; \theta) dx = 0 \quad -\textcircled{2}$$

Eq. \textcircled{2} - \textcircled{1} :-

$$\int (\hat{x} - x) \frac{\partial \ln P(x; \theta)}{\partial \theta} P(x; \theta) dx = \frac{\partial g(\theta)}{\partial \theta} - \textcircled{3}$$

\* Cauchy-Schwarz inequality condition

$$\left[ \int w(x) g(x) h(x) dx \right]^2 \leq \int w(x) g^2(x) dx \cdot \int w(x) h^2(x) dx$$

Remark:  $g(x) = ch(x)$  [c = constant]  
Condition:  $w(x) \geq 0 \quad \forall x$

$$w(x) = P(x; \theta)$$

$$g(x) = \hat{x} - x$$

$$h(x) = \frac{\partial \ln P(x; \theta)}{\partial \theta}$$

$$\left( \frac{\partial \vartheta(\theta)}{\partial \theta} \right)^2 \leq \int \underbrace{(\hat{x} - \alpha) \rho(x; \theta)}_{\left( \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \right) \rho(x; \theta)} dx$$

$$\text{Var}(\hat{\alpha}) \geq \frac{\left( \frac{\partial \vartheta(\theta)}{\partial \theta} \right)^2}{E \left[ \left( \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \right)^2 \right]} - \textcircled{4}$$

we know

$$E \left[ \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \right] = 0$$

$$\int \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \rho(x; \theta) dx = 0$$

$$\frac{\partial}{\partial \theta} \int \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \rho(x; \theta) dx = 0$$

$$\int \left[ \frac{\partial^2 \ln \rho(x; \theta)}{\partial \theta^2} \rho(x; \theta) + \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \cdot \frac{\partial \rho(x; \theta)}{\partial \theta} \right] dx = 0$$

$$-\int \frac{\partial^2 \ln \rho(x; \theta)}{\partial \theta^2} \rho(x; \theta) dx = \int \frac{\partial \ln \rho(x; \theta)}{\partial \theta} \cdot \frac{\partial \rho(x; \theta)}{\partial \theta} dx$$

$$-\mathbb{E} \left[ -\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = \int \frac{\partial \ln p(x; \theta)}{\partial \theta} \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) dx$$

$$= \mathbb{E} \left[ \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 \right]$$

Thus  $\text{var}(\hat{\theta}) \geq \frac{\left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2}{-\mathbb{E} \left[ \left( \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right) \right]}$

If  $\alpha = g(\theta) = 0$

$$\text{var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

\* Condition for equality is

$$g(x) = c n(x)$$

$$n(x) = \frac{1}{c} g(x)$$

$$\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{1}{c(\theta)} (\hat{\theta} - \theta)$$

Cum dependent on  $\theta$  but not on  $x$

$$\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} = -\frac{1}{c(\theta)} + \frac{\partial}{\partial \theta} \left( \frac{1}{c(\theta)} \right) (\hat{\theta} - \theta)$$

$$E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = -E \left[ \frac{1}{c(\theta)} \right] + E \left[ \frac{\partial}{\partial \theta} \left( \frac{1}{c(\theta)} \right) \right] (\hat{\theta} - \theta)$$

$$-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = +\frac{1}{c(\theta)} \quad \left[ \because E[\hat{\theta}] = \theta \right]$$

$$c(\theta) = \frac{1}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]} = \frac{1}{I(\theta)}$$

\* When CRLB is attained

$$\text{var}(\hat{\theta}) = \frac{1}{I(\theta)} \quad \text{where } I(\theta) = -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]$$

# CRLB for signals in WGN:-

Deterministic signal :  $s[n; \theta]$

$$x[n] = s[n; \theta] + w[n]$$

Signal is dependent on  $\theta$

$$P(x; \theta) = \frac{1}{(2\pi\sigma^2)^N/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n; \theta])^2 \right\}$$

$$\frac{\partial \ln P(x; \theta)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n; \theta]) \frac{\partial s[n; \theta]}{\partial \theta}$$

,

,

)

,

)

,

)

|

$$\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left\{ (x[n] - s[n; \theta]) \frac{\partial^2 s[n; \theta]}{\partial \theta^2} - \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2 \right\}$$

$$-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2$$

$$vm(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2}$$

\* Transformation of parameter :-

$$x[n] = A + w[n]$$

Power:  $\tilde{A}$

$$\alpha = g(\theta) \quad \left( \frac{\partial g(\theta)}{\partial \theta} \right)^2$$

$$vm(\hat{A}) \geq -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]$$

$$\alpha = A^2 = g(A)$$

$$vm(\hat{A}^2) \geq - \frac{(2A)^2}{N\sigma^2} \quad \therefore \frac{4\theta^2\sigma^2}{N}$$

$$\begin{aligned} vm(\hat{\theta}) &\geq \frac{\sigma^2}{N} \\ I(\hat{\theta}) &= \frac{N}{\sigma^2} \\ \hat{\theta} &= \bar{x}^2 \end{aligned}$$

$$E[\hat{x}^2] \stackrel{?}{=} \bar{x}^2 \quad \leftarrow \text{unbiased?}$$

$$E[\hat{x}^2] = \text{var}(\hat{x}) + E[\hat{x}]$$

$$= \frac{\sigma^2}{n} + \bar{x}^2 \neq \bar{x}^2$$

$$\text{var}(x) = E[x^2] - E[x]^2$$

$$E[x^2] = \text{var}(x) + E[x]^2$$

Biased

\*  $\hat{g}(\theta) = a\theta + b$

$$\widehat{g(\theta)} = \hat{g}(\hat{\theta}) = a\hat{\theta} + b$$

$$E[a\hat{\theta} + b] = a\theta + b = g(\theta)$$

$\widehat{g(\theta)}$  is unbiased.

CRLB :-

$$\text{var}(\widehat{g(\theta)}) \geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{I(\theta)} = a^2 \text{var}(\hat{\theta})$$

\*  $\hat{\theta} \quad \bar{x}$

$\hat{x}^2$  is asymptotically unbiased.  
for  $N \rightarrow \infty$

$$f(n) = n^2 + 2n$$

$$f(n) \sim n^2$$

$$1. \bar{x} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

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

$$\text{var}(\bar{x}) = 4\frac{\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}$$

$$n \rightarrow \infty$$

$$\text{var}(\bar{x}) \sim \frac{4\mu^2\sigma^2}{n}$$

$$\delta(\bar{x}) \approx \delta(\mu) + \frac{d\delta(\mu)}{d\mu} (\bar{x} - \mu)$$

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots$$

$$E[\delta(\bar{x})] = \delta(\mu) = \mu$$

vector  $\rightarrow$  lower case  
vector + bold fns.

\* vector parameter :-

$$\vec{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T$$

CRLB:  $P(\vec{x}; \vec{\theta})$  satisfies regularity condition

$$E\left[\frac{\partial \ln P(\vec{x}; \vec{\theta})}{\partial \vec{\theta}}\right] = 0 \quad \forall \vec{\theta}$$

The Covariance matrix of any unbiased estimator  $\hat{\theta}$  satisfies

$$\vec{C}_{\hat{\theta}} - \vec{\Sigma}^{-1}(\vec{\theta}) \geq \vec{0}$$

$$\begin{aligned} \xi &\sim N(\mu, \sigma^2) \\ E(\xi) &= \mu + \sigma^2 \\ E(\xi^2) &= \mu^2 + 2\mu\sigma^2 + 3\sigma^4 \end{aligned}$$

$$\begin{aligned} \text{var}(\xi) &= E[\xi^2] - E[\xi]^2 \\ &= 4\mu^2\sigma^2 + 2\sigma^4 \end{aligned}$$

$\vec{I}(\vec{\theta})$  = Fisher information matrix.

$$\vec{C}_{\vec{\theta}} = \vec{I}^{-1}(\vec{\theta})$$

$$\frac{\partial \ln p(\vec{x}; \vec{\theta})}{\partial \vec{\theta}} = \vec{I}(\vec{\theta})(\vec{\theta}(\vec{x}) - \vec{\theta})$$

## vector parameter CRLB for Transformations :-

$$\vec{\alpha} = \vec{g}(\vec{\theta}) \quad \vec{g}: n\text{-dimensional f.}$$

$$C_{\vec{\alpha}} = \frac{\partial \vec{g}(\vec{\theta})}{\partial \vec{\theta}} \vec{C}'(\vec{\theta}) \frac{\partial \vec{g}(\vec{\theta})^T}{\partial \vec{\theta}} > 0$$

$$\frac{\partial \vec{g}(\vec{\theta})}{\partial \vec{\theta}} = \begin{bmatrix} \frac{\partial g_1(\vec{\theta})}{\partial \theta_1} & \frac{\partial g_1(\vec{\theta})}{\partial \theta_2} & \dots & \frac{\partial g_1(\vec{\theta})}{\partial \theta_p} \\ \frac{\partial g_2(\vec{\theta})}{\partial \theta_1} & \frac{\partial g_2(\vec{\theta})}{\partial \theta_2} & \dots & \frac{\partial g_2(\vec{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(\vec{\theta})}{\partial \theta_1} & \frac{\partial g_n(\vec{\theta})}{\partial \theta_2} & \dots & \frac{\partial g_n(\vec{\theta})}{\partial \theta_p} \end{bmatrix}$$

\* CRLB for the General Gaussian Case:-

$$\vec{x} \sim \mathcal{N}(\vec{u}(\vec{\theta}), \vec{C}(\vec{\theta}))$$

Error information matrix

$$[I(\vec{\theta})]_{i,j} = \left[ \frac{\partial \vec{u}(\vec{\theta})}{\partial \theta_i} \right] \vec{C}'(\vec{\theta}) \left[ \frac{\partial \vec{u}(\vec{\theta})}{\partial \theta_j} \right]^T + \frac{1}{2} \operatorname{tr} \left[ \vec{C}'(\vec{\theta}) \frac{\partial C(\vec{\theta})}{\partial \theta_i} \vec{C}'(\vec{\theta}) \frac{\partial C(\vec{\theta})}{\partial \theta_j} \right]$$

$$\frac{\partial \vec{u}(\vec{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [u_1(\vec{\theta})]}{\partial \theta_i} \\ \frac{\partial [u_2(\vec{\theta})]}{\partial \theta_i} \\ \vdots \\ \frac{\partial [u_N(\vec{\theta})]}{\partial \theta_i} \end{bmatrix}$$

$$\frac{\partial c(\theta)}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [c(\theta)]_{11}}{\partial \theta_1} & \frac{\partial [c(\theta)]_{12}}{\partial \theta_1} & \frac{\partial [c(\theta)]_{1n}}{\partial \theta_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial [c(\theta)]_{n1}}{\partial \theta_1} & \frac{\partial [c(\theta)]_{n2}}{\partial \theta_1} & \frac{\partial [c(\theta)]_{nn}}{\partial \theta_1} \end{bmatrix}$$

For the scalar parameter:-

$$\vec{x} \sim \mathcal{N}(\vec{\mu}(\theta), \vec{C}(\theta))$$

$$\vec{I}(\theta) = \left[ \frac{\partial \vec{\mu}(\theta)}{\partial \theta} \right]^T \vec{C}'(\theta) \left[ \frac{\partial \vec{\mu}(\theta)}{\partial \theta} \right] + \frac{1}{2} \sigma_n \left[ (\vec{C}'(\theta) \frac{\partial \vec{C}(\theta)}{\partial \theta})^T \right]$$

$$x[n] = s[n; \theta] + n[n] \quad n=0, 1, \dots, N-1$$

$$n[N] \sim \mathcal{N}(0, \tilde{\sigma}^2)$$

$$\vec{z} = \sigma \vec{I}$$

$$I(\theta) = \frac{1}{\sigma^2} \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]^T \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]$$

$$= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial [s(n; \theta)]}{\partial \theta} \right)^T$$

$$I(\theta) = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^T$$

$$\sqrt{m(\theta)} \geq \frac{\sigma^2}{\hat{I}(\theta)}$$

$\theta$  to be reestim.

$$\begin{aligned} [\vec{I}(\theta)]_{ij} &= \left[ \frac{\partial \vec{h}(\theta)}{\partial \theta_i} \right]^T \frac{\vec{I}}{\sigma^2} \left[ \frac{\partial \vec{h}(\theta)}{\partial \theta_j} \right] \\ &= \frac{1}{\sigma^2} \left[ \frac{\partial \vec{h}(\theta)}{\partial \theta_i} \right]^T \left[ \frac{\partial \vec{h}(\theta)}{\partial \theta_j} \right] \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial S[r; \vec{\theta}]}{\partial \theta_i} \frac{\partial S[n; \vec{\theta}]}{\partial \theta_j} \end{aligned}$$

#  $x[n] = w[n]$   $w[n] \sim N(0, \sigma^2)$

$$\theta: \mathbb{G}^V$$

$$C(\sigma^2) = \sigma^2 I \leftarrow \text{Identity matrix}$$

$$\begin{aligned} \vec{I}(\sigma^2) &= \frac{1}{2} + n \left[ (C^{-1}(\theta)) \frac{\partial C(\theta)}{\partial \theta} \right]^2 \\ &= \frac{1}{2} + n \left[ (C^{-1}(\sigma^2)) \frac{\partial C(\sigma^2)}{\partial \sigma^2} \right]^2 \\ &= \frac{1}{2} + n \left[ \left( \frac{1}{\sigma^2} \frac{\partial \sigma^2 I}{\partial \sigma^2} \right)^2 \right] \end{aligned}$$

$$= \frac{1}{2} + n \left[ \frac{1}{\sigma^4} \mathbb{I} \right] \\ = \frac{N}{2\sigma^4}$$

\*  $x[n] = \underbrace{A \cos(2\pi f_0 n + \phi)}_{n=0, 1, \dots, N-1} + w[n]$

$$A > 0 \\ 0 < f_0 < \frac{1}{2} \\ \vec{\theta} = [A \ f_0 \ \phi]^T$$

$$[\mathbf{I}(\theta)]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \theta]}{\partial \theta_i} \frac{\partial s[n; \theta]}{\partial \theta_j}$$

$$\alpha = 2\pi f_0 n + \phi$$

$$[\mathbf{I}(\theta)]_{11} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \cos^2 \alpha = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{1}{2} + \frac{1}{2} \cos 2\alpha \right) \\ \approx \frac{N}{2\sigma^2}$$

$$[\mathbf{I}(\theta)]_{12} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} 2\pi n \cos \alpha (-A \sin \alpha)$$

$$[\mathbf{I}(\theta)]_{13} = ? \quad \text{?} \quad \approx 0 \\ = -\frac{\pi A}{\sigma^2} \sum_{n=0}^{N-1} 2n \cos \alpha \sin \alpha \\ = -\frac{\pi A}{\sigma^2} \sum_{n=0}^{N-1} n \sin 2\alpha \approx 0$$

$$\begin{array}{ll} [\mathbf{I}(\theta)]_{21} = ? & [\mathbf{I}(\theta)]_{31} = ? \\ [\mathbf{I}(\theta)]_{12} = ? & [\mathbf{I}(\theta)]_{32} = ? \\ [\mathbf{I}(\theta)]_{13} = ? & [\mathbf{I}(\theta)]_{33} = ? \end{array}$$

$\mathbf{I}(\theta)$

$\mathbf{I}'(\theta) =$

$\text{vm}(\hat{\theta}) \gamma,$

$\text{vm}(\hat{\mu}_0) \gamma,$

$\text{vm}(\hat{\phi}) \gamma,$



Linear model:-

$$\vec{x} = \vec{H} \vec{\theta} + \vec{w}.$$

$\vec{x} = N \times 1$  observation vector.

$\vec{H} = N \times P$  " matrix

$\vec{\theta} = P \times 1$  vector of parameters.

$w = N \times 1 \sim N(0, \sigma^2 I)$

$$\hat{\theta}_{\text{MLRE}} = g(x)$$

$$\frac{\partial \ln P(x; \theta)}{\partial \theta} = I(\theta) (g(x) - \theta)$$

$$C_{\hat{\theta}} = I^{-1}(\theta)$$

$$P(x; \theta) = \frac{1}{(2\pi\sigma^2)^N} \exp\left\{-\frac{1}{2\sigma^2}(x - H\theta)^T(x - H\theta)\right\}$$

$$\ln P(x; \theta) = \left[ -\ln (2\pi\sigma^2)^N - \frac{1}{2\sigma^2}(x - H\theta)^T(x - H\theta) \right]$$

$$\frac{\partial \ln P(x; \theta)}{\partial \theta} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta} \left[ (x - H\theta)^T(x - H\theta) \right]$$

$$= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta} \left[ x^T x - 2x^T H\theta + \theta^T H^T H\theta \right]$$

$$\frac{\partial b^T \theta}{\partial \theta} = b = -\frac{1}{2\sigma^2} \left[ -H^T x + H^T H\theta \right]$$

$$\frac{\partial \theta^T A \theta}{\partial \theta} = 2A\theta$$

$$\frac{\partial \ln P(x; \theta)}{\partial \theta} = \frac{H^T H}{\sigma^2} \left[ (H^T H)^{-1} H^T x - \theta \right]$$

$$I(\theta) = \frac{H^T H}{\sigma^2}$$

$$\hat{\theta} = g(x) = (H^T H)^{-1} H^T x$$

$$C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1}$$

# Lecture - 14

$$x = H\theta + w$$

$$H = DB$$

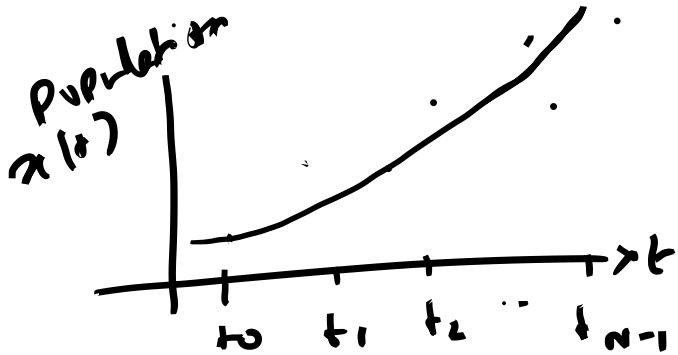
$$\hat{\theta}_{MLSE} = (H^T H)^{-1} H^T x$$

$$\underline{y = DBx + w}$$

$$I(\theta) = \frac{H^T H}{\sigma^2}$$

$$C_{\theta} = \sigma^2 (H^T H)^{-1}$$

#1.



$$x(t_n) = \theta_0 + \theta_1 t_n + \theta_2 t_n^2 + w(t_n)$$

$$\vec{x} = \vec{H} \vec{\theta} + \vec{w}$$

$$n = 0, 1, \dots, N-1$$

$$\vec{x} = \begin{bmatrix} x(t_0) \\ x(t_1) \\ \vdots \\ x(t_{N-1}) \end{bmatrix} \quad \vec{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}_{N \times 3} \quad \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{N-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{N-1} & w^{2(N-1)} & \cdots & w^{(N-1)^2} \end{bmatrix}$$

$$w = e^{j \frac{2\pi}{N}}$$

$$\hat{\theta} = (H^T H)^{-1} H^T x$$

$$\xi(t) = \sum_{i=1}^3 \theta_i t^{i-1}$$

↓  
(unre)

#2.

$$x[n] = \sum_{k=1}^m a_k \cos\left(\frac{2\pi k n}{N}\right) + \sum_{k=1}^m b_k \sin\left(\frac{2\pi k n}{N}\right) + w[n]$$

$$\Theta = [a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_m]^T$$

$$\vec{H} = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & 0 & \dots & 0 \\ \cos\left(\frac{2\pi}{N}\right) & \dots & \cos\left(\frac{2\pi m}{N}\right) & \sin\left(\frac{2\pi}{N}\right) & \dots & \sin\left(\frac{2\pi m}{N}\right) \\ \vdots & & \ddots & & \ddots & & \ddots & & \vdots \\ \vdots & & & & & & & & \vdots \\ \end{bmatrix}_{N \times 2m}$$

$$\vec{h} = [\vec{h}_1, \vec{h}_2, \dots, \vec{h}_{2m}]$$

$$\vec{h}_i^T \vec{h}_j = 0 \quad \text{for } i \neq j$$

$$\vec{h}^T \vec{h} = \begin{bmatrix} h_1^T \\ \vdots \\ h_{2m}^T \end{bmatrix} \begin{bmatrix} h_1 & h_2 & \dots & h_{2m} \end{bmatrix}$$

$$= \begin{bmatrix} h_1^T h_1 & h_1^T h_2 & \dots & h_1^T h_{2m} \\ h_2^T h_1 & h_2^T h_2 & \dots & h_2^T h_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ h_{2m}^T h_1 & h_{2m}^T h_2 & \dots & h_{2m}^T h_{2m} \end{bmatrix}$$

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi i n}{N}\right) \cos\left(\frac{2\pi j n}{N}\right) = \frac{N}{2} \delta_{ij}$$

$$\sum_{n=0}^{N-1} \sin\left(\frac{2\pi i n}{N}\right) \sin\left(\frac{2\pi j n}{N}\right) = \frac{N}{2} \delta_{ij}$$

$$H^T H = \begin{bmatrix} N/2 & 0 & 0 & \dots & 0 \\ 0 & N/2 & 0 & \dots & 0 \\ 0 & 0 & N/2 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & N/2 \end{bmatrix} = \frac{N}{2} I$$

$$\int_0^T \cos n \omega_0 t dt = 0$$

$$\int_0^T \sin n \omega_0 t dt = 0$$

$$\hat{\theta} = (H^T H)^{-1} H^T x = \frac{N}{2} H^T x = \frac{N}{2} \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_{N/2}^T \end{bmatrix} x$$

$$= \begin{bmatrix} \frac{N}{2} x_1^T x \\ \frac{N}{2} x_2^T x \\ \vdots \\ \frac{N}{2} x_{N/2}^T x \end{bmatrix}$$

$$\hat{a}_k = \frac{N}{2} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi k n}{N}\right)$$

$$\hat{b}_k = \frac{N}{2} \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi k n}{N}\right)$$

$$\mathbb{E}[\hat{a}_k] = a_k$$

$$\mathbb{E}[\hat{b}_k] = b_k$$

$$C_B = \sigma^2 (H^T H)^{-1} = \sigma^2 \left(\frac{N}{2} I\right)^{-1} = \frac{2\sigma^2}{N} I$$

## Lecture - 15

$$x = H\theta + w$$

$$\hat{\theta} = (H^T H)^{-1} H^T x ; \quad I(\theta) = \frac{H^T H}{\sigma^2}$$

\* General Linear Model :-

① Noise is not white  $\rightarrow w \sim \mathcal{N}(0, C)$

②  $x = H\theta + s + w$        $s$  is signal vector (known)

③  $C$  is not a scaled identity matrix.

$C$  is +ve definite matrix.

$C^{-1}$  is +ve " "

$$C^{-1} = D^T D \quad [D = non-singular matrix]$$

$$C = E[w w^T]$$

$$= E[(Dw)(Dw)^T] = E[Dw w^T D^T]$$

$$= D E[w w^T] D^T$$

$$= D C D^T$$

$$= D (D^T D)^{-1} D^T$$

$$= D D^{-1} (D^T)^{-1} D^T$$

$$= I.$$

$$x = H\theta + w$$

$$Dx = D(H\theta + w)$$

$$x' = H'\theta + w'$$

$$w \sim N(0, \Sigma)$$

$$\hat{\theta} = (H^T H)^{-1} H^T x'$$

$$= (H^T D^T D H)^{-1} H^T D^T D \hat{\theta}$$

mrug:  $\hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} x$

$$C_{\hat{\theta}} = (H^T H)^{-1} = (H^T D^T D H)^{-1} = (H^T C^{-1} H)^{-1}$$

\* 2nd Extension:  $x = H\theta + S + w$ .

$$x - S = H\theta + w$$

$$x' = H\theta + w$$

$$\hat{\theta} = (H^T H)^{-1} H^T x' = (H^T H)^{-1} H^T (x - S)$$

$$C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1}$$

#1.  $x[n] = A + \alpha^n + w[n]$  for  $n = 0, 1, \dots, N-1$

$A$  is known,  $A$  needs to be estimated.

$$w[n] \sim N(0, \sigma^2)$$

$$\vec{x} = A \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \vec{s} + \vec{w}$$

$$\vec{s} = [1 \quad x \quad x^2 \dots x^{n-1}]^T$$

$$\hat{\theta} : (H^T H)^{-1} H^T x' \quad [x' := x - s]$$

$$\hat{A} : \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})$$

$$var(\hat{\mu}) = \frac{\sigma^2}{N}$$

(N > P)

$$\vec{x} = H \hat{\theta} + \vec{s} + \vec{w}$$

$$\hat{\theta} : (H^T C^{-1} H)^{-1} H^T C^{-1} (x - s)$$

$$C_S : (H^T C^{-1} H)^{-1}$$

$\vec{x}$  : N x 1 : Observation vector  
 $H$  : N x P : Observation matrix  
 $\hat{\theta}$  : P x 1 : vector param.  
 $\vec{s}$  : N x 1 : Known signal  
 $\vec{w}$  : N x 1  $\rightarrow \mathcal{G}_N \sim \mathcal{N}(0, C)$

RBLSS : (Rao-Blackwell-Lehman-Scheffe)

Sufficient Statistic :-

$$x[n] = A + w[n]$$

$$[w[n] \sim \mathcal{N}(0, \sigma^2)]$$

$$\hat{A} : \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \text{minimum variance } \frac{\sigma^2}{N}$$

$$S_1 = \{x[0], x[1], \dots, x[n-1]\}$$

$$S_2 = \{x[0] + x[1], x[2], \dots, x[n-1]\}$$

$$S_3 = \left\{ \sum_{n=0}^{N-1} x[n] \right\} \rightarrow \text{minimal Sufficient Statistic.}$$

$$P(x; \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \theta)^2 \right]$$

$$T(x) = \sum_{n=0}^{N-1} x[n] = T_0$$

$$P(x \mid T(x) = T_0; \theta) = \underbrace{P(x \mid T(\bar{x}) = T_0)}_{\text{Independent of } \theta \text{ parameter}}$$

Neyman - Fisher Factorization:

$$P(\vec{x}; \theta) = g(T(\vec{x}), \theta) h(\vec{x})$$

$g$ : fn depending on  $x$  only through  $T(x)$

$h$ : " " " " "  $x$ .

$T(x)$ : sufficient statistic for  $\theta$ .

$$P(x; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \theta)^2\right\}$$

$$\sum_{n=0}^{N-1} (x[n] - \theta)^2 = \sum_{n=0}^{N-1} x^2[n] - 2\theta \sum_{n=0}^{N-1} x[n] + N\theta^2$$

$$P(x; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \left(N\theta^2 - 2\theta \sum_n x[n] + \underbrace{\sum_n x^2[n]}_{h(x)}\right)\right]$$

$$T(x) = \sum_{n=0}^{N-1} x[n]$$

$$\text{#2. } x[n] = w[n]$$

$$g(T(x), \theta) h(x)$$

$$f_{\text{noise}} = G^2$$

$$P(x; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]\right] \cdot \frac{1}{h(x)}$$

$$T(x) = \sum_{n=0}^{N-1} x^2[n]$$

#3.  $x[n] = A \cos(2\pi f_0 n + \phi) + w[n]$

```

graph LR
    C[Carrier Wave] --> A[A]
    A --> M1[Modulator 1]
    M1 --> M2[Modulator 2]
    M2 --> W[Additive Noise]
    W --> S[Summing Junction]
    S --> O[Output]
  
```

The diagram illustrates the generation of a noisy signal. It starts with a "Carrier Wave" at the top, which feeds into a block labeled "A". This is followed by two modulators, "Modulator 1" and "Modulator 2", represented by arrows pointing right. The output of Modulator 2 is combined with "Additive Noise" (represented by a wavy line) at a "Summing Junction". The final output is labeled "O".

## Lecture - 16

### NF Factorization :-

$$P(x; \theta) = g(T(x), \theta) h(x)$$

$T(x)$  : Sufficient Statistics.

#  $x[n] = A \cos(2\pi f_0 n + \phi) + w[n]$

known      known      unknown  
 ↓              ↓              ↓  
 known      known      known  $\sigma^2$

$$P(x, \phi) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x[n] - A \cos(2\pi f_0 n + \phi)]^2 \right\}$$

The exponent :-

$$\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} \cos(2\pi f_0 n + \phi) x[n] + \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi)$$

$$= \sum_{n=0}^{N-1} x^2[n] - 2A \left( \underbrace{\sum_{n=0}^{N-1} \cos(2\pi f_0 n)}_{T_1(x)} \right) \cos \phi + 2A \left( \underbrace{\sum_{n=0}^{N-1} x[n] \sin(2\pi f_0 n)}_{T_2(x)} \right) \sin \phi + \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi)$$

$$P(x, \phi) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi) - 2AT_1(x) \cos \phi + 2A T_2(x) \sin \phi \right] \right\} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]$$

$T_1(n) \Delta T_2(n)$  one jointly sufficient statistic  
for  $\phi$ .

\*  $x[n] = A + w[n]$

$$T(k) = \sum_{n=0}^{N-1} x[n]$$

① Find any unbiased estimator for  $A$

$$\hat{A} = x[0]$$

$$\hat{A} = E[\hat{x}|T]$$

② Find some  $f$  s.t.  $\hat{A} = f(T)$  is an unbiased estimator of  $A$ .

①  $\hat{A} = x[0]$

$$\hat{A} = E[x[0] | \sum_{n=0}^{N-1} x[n]]$$

\*  $\begin{bmatrix} x \\ y \end{bmatrix}$ : Gaussian Random Vector

$$\vec{\mu} = \begin{bmatrix} E[x] \\ E[y] \end{bmatrix}$$

$$\vec{\Sigma} = \begin{bmatrix} \text{var}(x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{var}(y) \end{bmatrix}$$

$$P(x|y) \sim \text{JR}(A, B)$$

$$\begin{aligned} A: E[x|y] &= \int_{-\infty}^{\infty} x P(x|y) dx = \int_{-\infty}^{\infty} x \frac{P(x,y)}{P(y)} dx \\ &= \int_{-\infty}^{\infty} x \frac{P(x,y)}{\int P(x,y) dx} dx \end{aligned}$$

$$A: E[x|y] = E[x] + \frac{\text{Cov}(x,y)}{\text{var}(y)} (y - E[y])$$

$$B: \text{var}(x|y) = \text{var}(x) - \frac{\text{Cov}^2(x,y)}{\text{var}(y)}$$

$$x = x[0]$$

$$y = \sum_n x[n]$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x[0] \\ \sum_n x[n] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[n-1] \end{bmatrix}$$

The PDF of  $[x \ y]^T$  is  $\text{JR}(\vec{\mu}, \vec{C})$

$$\vec{\mu}: E[Lx] = \vec{L} E[x] = \vec{L} A \vec{x} = \begin{bmatrix} A \\ N_A \end{bmatrix}$$

$$\vec{C} = \vec{L} \vec{L}^T \sigma^2 = \sigma^2 \begin{bmatrix} 1 & 1 \\ \vdots & \ddots & N \end{bmatrix}$$

$$\hat{A} = E[\pi|y] = E[x] + \frac{\text{Cov}(x, \pi)}{\text{Var}(y)} (y - E[y])$$

$$= A + \frac{\sigma^2}{N\sigma_y^2} \left( \sum_{n=0}^{N-1} \pi[n] - NA \right)$$

$$= A + \frac{1}{N} \left( \sum_n \pi[n] \right) - A$$

$$\hat{A} = \frac{1}{N} \sum_n \pi[n]$$

MVUE

④

$$\hat{A} = g \left( \sum_{n=0}^{N-1} \pi[n] \right)$$

$$g(n) = \frac{n}{N}$$

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} \pi[n]$$

\* RBLT Theorem:-

If  $\hat{\theta}$  is an unbiased estimator of  $\theta$  and  $T(x)$  is a sufficient statistic for  $\theta$ , then  $\hat{\theta} = E[\hat{\theta}|T(x)]$  is

① a valid estimator.

② unbiased

③ of lesser or equal variance than that of  $\hat{\theta}$  for all  $\theta$ .

Additionally if the sufficient statistic is complete

then  $\hat{\theta}$  is MLE.

- \* A statistic is complete if there is only one fn of the statistic that is unbiased.

$$\delta = E[\delta | T(x)]$$

$E[\delta | T(x)]$  is a suff fct of  $T(x)$ .

$$\begin{aligned}\delta &: E[\delta | T(x)] = \int \delta p(\delta | T(x)) d\delta \\ &= g(T(x))\end{aligned}$$

\*  $x[n] = A + n[x]$

$$T = \sum_{n=0}^{N-1} x[n]$$

$$E[g(T(x))] = A$$

$$E[h(T(x))] = A$$

Suppose there exists another fn  $n$  for which  $E[n(T(x))] = A$

$$E[g(T) - h(T)] = 0 \quad \forall A$$

$$T \sim N(A, N\sigma^2)$$

$$\int_{-\infty}^{\infty} (g(\tau) - h(\tau)) \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{1}{2N\sigma^2} (T - N\tau)^2\right] d\tau \approx 0$$

$$v(\tau) = g(\tau) - h(\tau)$$

$$\tau = \frac{\tau}{N} \quad \text{and} \quad v'(\tau) = v(N\tau) = v(\tau)$$

$$\int_{-\infty}^{\infty} v'(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{N}{2\sigma^2} (n - \tau)^2\right] d\tau \approx 0$$

$$v'(\tau) * n(\tau) = 0 \quad [n(\tau) = \frac{\text{height}}{\text{width}}]$$

$$v'(\tau) = 0 \quad \forall \tau$$

$$g(\tau) = h(\tau)$$

## Lecture 17

$$\hat{\theta} = E[\tilde{\theta} | \tau(x)] = g(\tau(x))$$

#1.  $x[n] = A + w[n]$

$$\tau(x) = \sum_n x[n]$$

$$E[g(\tau(x))] = A$$

#2.  $x[0] = A + w[0]$

$$w[0] \sim U\left[\frac{1}{2}, \frac{1}{2}\right]$$

$$E[g[x[0]]] = A$$

$$h \stackrel{?}{=} g$$

$$E[h[x[0]]] = A$$

$$\tau(x) = x[0]$$

$$v(\tau) = g(\tau) - h(\tau)$$

$$E[v(\tau)] = 0$$

$$\int_{-\infty}^{\infty} v(\tau) p(x; A) dx = 0 \quad \forall A.$$

$$x = x[0] = T$$

$$\int_{-\infty}^{\infty} v(\tau) p(\tau; A) d\tau = 0 \quad \forall A.$$

$$\int_{A - \frac{1}{2}}^{A + \frac{1}{2}} v(\tau) d\tau = 0 \quad \forall \theta$$

$$v(\tau) = \sin 2\pi \tau$$

$$g(\tau) - h(\tau) = \sin 2\pi \tau$$

$$g(\tau) \neq h(\tau)$$

\* A sufficient statistic is complete if

$$\int_{-\infty}^{\infty} v(\tau) p(\tau; \theta) d\tau = 0 \quad \text{for all } \theta.$$

is satisfied only by  $v(\tau) = 0$  for all  $\tau$ .

\* i) Find sufficient statistic for  $\theta$  by using NF factorization theorem.

ii) Check if  $T(x)$  is complete. If yes then proved. If not this approach can not be applied.

$$\hat{\theta} = E(\check{\theta} | T(x))$$

$$\hat{\theta} = g(T(x))$$

# Best Linear Unbiased Estimator (BLUE)

$$\{x[0], x[1], \dots, x[n-1]\} \quad \Theta$$

Linear:  $\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$

Unbiased:  $E[\hat{\theta}] = \Theta$

$$E\left[\sum_{n=0}^{N-1} a_n x[n]\right] = \Theta \quad -①$$

Best:  $\text{Var}(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2]$

$$= E\left[\left(\sum_n a_n x[n] - E\left[\sum_n a_n x[n]\right]\right)^2\right]$$

$$\vec{a} = [a_0 \ a_1 \ \dots \ a_{N-1}]^\top$$

$$\text{Var}(\hat{\theta}) = E\left[(\vec{a}^\top \vec{x} - \vec{c}^\top E[\vec{x}])^2\right]$$

$$= E\left[(\vec{a}^\top (\vec{x} - E[\vec{x}]))^2\right]$$

$$= E\left[\underbrace{\vec{a}^\top (\vec{x} - E[\vec{x}])}_{\vec{a}^\top \vec{r}} \underbrace{(\vec{x} - E[\vec{x}])^\top \vec{a}}\right]$$

$$\text{Var}(\hat{\theta}) = \vec{a}^\top \Sigma \vec{a} \quad -②$$

$$E[x[n]] = S[n] \theta \quad \left[ \begin{array}{l} S[n] \text{ known} \\ -\textcircled{3} \end{array} \right]$$

$$\text{minimize } \text{Var}(\hat{\theta}) = \vec{a}^T \vec{C} \vec{a} \text{ s.t. } E[\hat{\theta}] = \sum_n c_n E[x[n]] = \theta.$$

$$\text{The constraint: } \sum_{n=0}^{N-1} c_n E[x[n]] = \theta.$$

$$\sum_{n=0}^{N-1} c_n S[n] \theta = \theta$$

$$\sum_{n=0}^{N-1} c_n S[n] = 1.$$

$$\vec{a}^T \vec{s} = 1 \quad \left[ \vec{s} = \begin{bmatrix} S[0] \\ S[1] \\ \vdots \\ S[N-1] \end{bmatrix} \right]$$

The minimization problem—

$$J = \vec{a}^T \vec{C} \vec{a} + \lambda (\vec{a}^T \vec{s} - 1)$$

$J$ : lagrangian  
f.

$$\frac{\partial J}{\partial a} = 2 \vec{C} \vec{a} + \lambda \vec{s} \rightarrow 0$$

$$\vec{a} = -\frac{\lambda}{2} \vec{C}^{-1} \vec{s} - \textcircled{4}$$

$$\vec{a}^T \vec{s} = 1$$

$$\left( -\frac{\lambda}{2} \vec{C}^{-1} \vec{s} \right)^T \vec{s} = 1$$

$$-\frac{\lambda}{2} \vec{s}^T \vec{C}^{-1} \vec{s} = 1$$

$$-\frac{1}{2} = -\frac{1}{S^T C^{-1} S} - \textcircled{5}$$

$$\hat{\alpha}_{opt} = \frac{\vec{C}^{-1} \vec{s}}{\vec{s}^T \vec{C}^{-1} \vec{s}}$$

$$\hat{\theta} = \vec{z}^T \vec{x} : \frac{\vec{s}^T \vec{C}^{-1} \vec{x}}{\vec{s}^T \vec{C}^{-1} \vec{s}}$$

$$Var(\hat{\theta}) = \vec{a}_{opt}^T \vec{C} \vec{a}_{opt} = \frac{1}{\vec{s}^T \vec{C}^{-1} \vec{s}}$$

$$E[\hat{\theta}] = \frac{S^T C^{-1} \Theta [t]}{S^T C^{-1} S} = \frac{S^T C^{-1} \Theta S}{S^T C^{-1} S} = \Theta$$

unbiased

1.  $S$  : scaled mean

2.  $C$  : covariance matrix.

Entire PDF  
is multivariate

$$\#1. \quad x[n] = A + w[n]$$

$w[n] \rightarrow$  unspecified  
PDF  $\sigma^2$

$n = 0, 1, \dots, N-1$

$$E[x[n]] = A$$

$$E[x[n]^T] = S[n] \Theta$$

$$\begin{cases} \Theta = A \\ S[n] = I \end{cases}$$

$$\hat{A} = \frac{\vec{s}^T \vec{C}^{-1} \vec{x}}{\vec{s}^T \vec{C}^{-1} \vec{s}} = \frac{\vec{I} \frac{1}{\sigma^2} \vec{I}^T \vec{x}}{\vec{I}^T \frac{1}{\sigma^2} \vec{I}}$$

$$= \frac{\vec{z}^T \vec{x}}{\vec{z}^T \vec{z}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

$$\text{Var}(\hat{\theta}) = \frac{1}{\vec{z}^T \vec{z}} = \frac{\sigma^2}{\vec{z}^T \vec{z}} = \frac{\sigma^2}{N}$$

#2. uncorrelated noise  $\vec{e}_n$

$$\vec{z} = \vec{1}$$

$$\hat{\theta} = \frac{\vec{z}^T \vec{c}^T \vec{x}}{\vec{z}^T \vec{c}^T \vec{1}}$$

$$\text{Var}(\hat{\theta}) = \frac{1}{\vec{z}^T \vec{c}^T \vec{c}}$$

$$C = \begin{bmatrix} \sigma_0^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ 0 & 0 & \dots & \sigma_{N-1}^2 \end{bmatrix}$$

$$C^{-1} = \begin{bmatrix} \frac{1}{\sigma_0^2} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_1^2} & \dots & & \\ 0 & \dots & & \frac{1}{\sigma_{N-1}^2} & \end{bmatrix}$$

$$\hat{\theta} = \underline{\hspace{2cm}}$$

$$\text{Var}(\hat{\theta}) = \underline{\hspace{2cm}}$$

## Lecture - 18.

BLVF :-  $\hat{\theta} = \frac{s^T c^{-1} x}{s^T c^{-1} s}$

$$var(\hat{\theta}) = \frac{1}{s^T c^{-1} s}$$

#  $x[n] = A + w[n]$

$$\hat{A} = \frac{\sum_n \frac{x[n]}{\sigma^2[n]}}{\sum_n \frac{1}{\sigma^2[n]}}$$

$w[n]$ : zero mean  
un-correlated noise  
with variance  $\sigma_n^2$

$$var(\hat{A}) = \frac{1}{\sum_{n=1}^{N-1} \frac{1}{\sigma_n^2}}$$

### Vector Parameter :-

$$\vec{\theta} = \vec{A} \vec{x}$$

$$E[\vec{\theta}] = A E[\vec{x}] = A \theta$$

$$E[\vec{x}] = H \theta$$

$$A H = I$$

$$H = \begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix}$$

$$\vec{C}_\theta = A^T \vec{C} A \quad \text{s.t. } A H = I$$

$\vec{C} = \frac{\text{covariance}}{\text{matrix}}$  of  $\vec{x}$

$$\hat{\theta} = A \vec{x} = (H^T C^{-1} H)^{-1} H^T C^{-1} \vec{x}$$

$$C_\theta = (H^T C^{-1} H)^{-1}$$

$$\vec{x} = H \theta + w$$

$$\hat{\theta} =$$

## Gauss-Markov Theorem:-

If the data are of general linear model

form  $\vec{x} = \vec{H}\vec{\theta} + \vec{w}$   $\vec{H}$ :  $N \times P$  observation matrix  
 $\vec{\theta}$ :  $P \times 1$  vector  
 $\vec{w}$ :  $N \times 1$  noise vector

then the GLUE of  $\hat{\theta}$

$$\text{is } \hat{\theta} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}$$
zero mean and  
covariance  $\mathbf{C}$

$$\text{var}(\hat{\theta}_i) = \left[ (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \right]_{ii}$$

$$\text{cov}(\hat{\theta}) = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}$$

## Maximum-Likelihood Estimation (MLE) :-

$$* \quad \mathbf{x}[n] = \mathbf{A} + \mathbf{w}[n] \quad [\mathbf{w}[n] \sim \mathcal{N}(0, \mathbf{A})]$$

### MVUE using RLB :-

$$P(x; A) = \frac{1}{(2\pi A)^{n/2}} \exp \left[ -\frac{1}{2A} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]$$

$$\frac{\partial \ln P(x; A)}{\partial A} = -\frac{N}{2A} + \frac{1}{n} \sum_n (x[n] - A) + \frac{1}{2A^2} \sum_n (x[n] - A)^2$$

$$\therefore I(\theta)(g(x) - \theta) \stackrel{?}{=} I(A)(\hat{A} - A)$$

$$\text{var}(\hat{A}) > B \quad [B = ?]$$

## Using Sufficient Statistic.

$$\frac{1}{A} \sum_n (x[n] - A)^2 = \frac{1}{n} \sum_n x^2[n] - 2N\bar{x} + NA$$

$$p(x; A) = \frac{\overbrace{(2\pi A)^{n/2}}^{\text{constant}} \exp\left[-\frac{1}{2}\left(\frac{1}{A} \sum_n x^2[n] + NA\right)\right]}{\underbrace{g\left(\sum_n x^2[n], A\right)}_{\text{constant}} \exp\left[\frac{n\bar{x}}{A}\right]}$$

$$T(x) = \sum_n x^2[n]$$

Assuming  $T(x)$  is complete

$$\text{i) } E[g(T(x))] = A$$

$$E\left[g\left(\sum_n x^2[n]\right)\right] = A$$

$$\begin{aligned} E\left[\sum_n x^2[n]\right] &= N E[x^2[n]] \\ &= N \left[ \text{var}(x[n]) + (E[x[n]])^2 \right] \\ &= N [A + A^2] \neq A \end{aligned}$$

$$\text{ii) } E(\tilde{A} | T(x)) \quad [\tilde{A} = \text{unbiased estimator}]$$

$$\text{If } \tilde{A} = x[0]$$

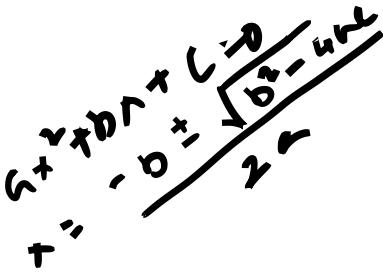
$$E(x[0] | \sum_n x^2[n])$$

For MLE:  
If  $N \rightarrow \infty : E(\hat{\theta}) \rightarrow A : \text{Var}(\hat{\theta}) \rightarrow (\text{RIB})$

$\hat{\theta} = \arg \max_{\theta} P(x; \theta)$  fun  $\times$  fixed.

#  $P(x; A) = \frac{1}{(2\pi A)^{N/2}} \exp \left\{ -\frac{1}{2A} \sum_n (x[n] - A)^2 \right\}$

$$\frac{\partial \ln P(x; A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A} \sum_n (x[n] - A) + \frac{1}{2A^2} \sum_n (x[n] - A)^2 \xrightarrow{} 0$$



$$\hat{A}^2 + \hat{A} - \frac{1}{N} \sum_n x^2[n] = 0$$

$$\hat{A} = -\frac{1}{2} \pm \sqrt{\frac{1}{N} \sum_n x^2[n] + \frac{1}{4}}$$

$$\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N} \sum_n x^2[n] + \frac{1}{4}} \quad \text{to make } A > 0$$

\*  $x[n] = \mu + w[n]$   $[w[n] \sim \mathcal{N}(0, \sigma^2)]$

$$P(x; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right\}$$

$$\frac{\partial \ln P(x; A)}{\partial A} = \frac{1}{\sigma^2} \sum_n (x[n] - A) \rightarrow 0$$

$$\hat{A} = \frac{1}{N} \sum_n x[n]$$

- \* If efficient exists, the MLE will produce it.

### Asymptotic Property of the MLE :-

If the PDF  $P(x; \theta)$  satisfies some regularity conditions, then the MLE of the unknown parameter  $\theta$  is "asymptotically distributed according to"

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta))$$

$I(\theta)$  : Fisher information evaluated at  $\theta$ .

$$\frac{\partial \ln P(x; \theta)}{\partial \theta} = I(\theta)(\hat{\theta}(x) - \theta)$$

$$\hat{\theta}_{\text{MVR}} = g(x)$$

$$I(\theta)(\hat{\theta}(x) - \theta) = 0$$

$$\hat{\theta}_{\text{MLE}} = g(x) = \hat{\theta}_{\text{MVR}}$$