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.

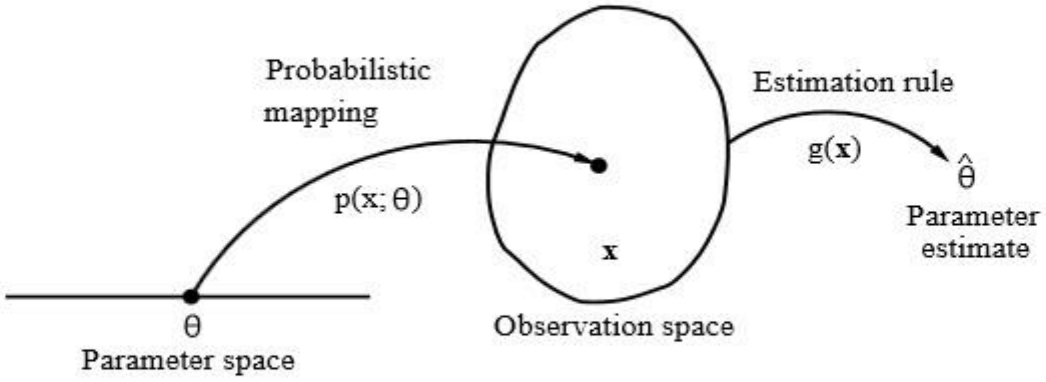
Applications of Detection

- **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

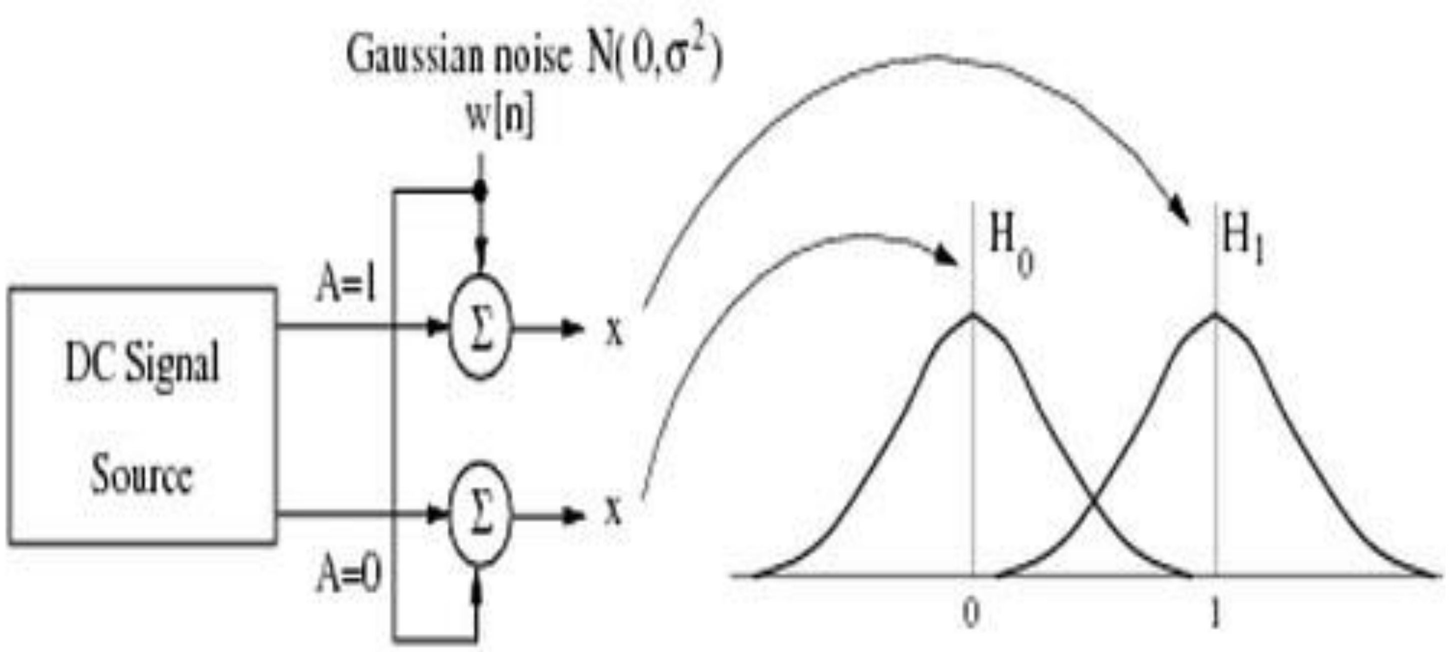
Formulation of the Estimation Problem



Classification of Estimation Approaches

Unknown param.	Probabilistic assumption	Other requirement	Estimator kind (salient property)	Practical utility
Non-random	Completely known PDF	Sufficeient statistics	Minimum variance unbiased (optimal)	Low
		Large data; no statistics	Maximum likelihood (asymptotically optimal)	Very high
	First-two moments only		Best linear unbiased (suboptimal in general)	Moderate
	Known signal model; no PDF		Least squares (suboptimal)	High
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
Level 1: Known signals in noise	1. Synchronous digital communication 2. Pattern recognition
Level 2: Signals with unknown parameters in noise	1. Digital communication system without phase reference 2. Digital communication over slowly fading channels 3. Conventional pulse radar and sonar, target detection
Level 3: Random signals in noise	1. Digital communication over scatter link 2. Passive sonar 3. Radio astronomy (detection of noise sources)

Foundations

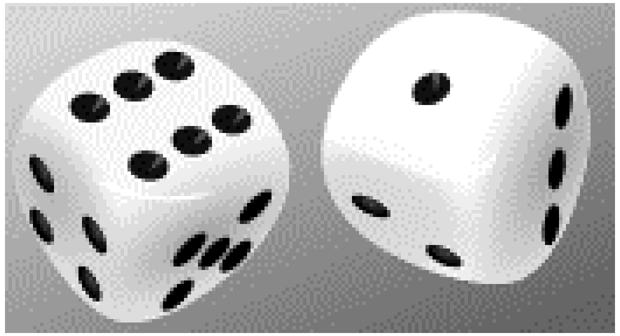
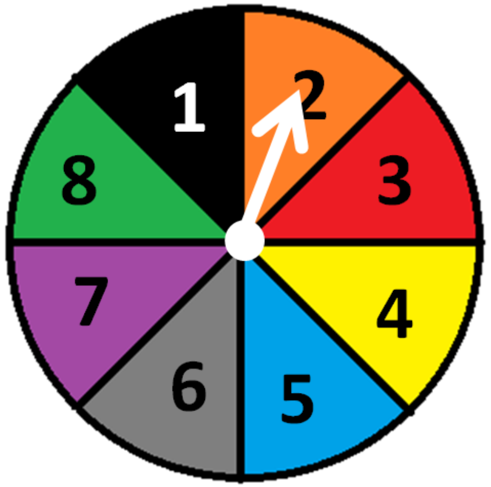
Srimanta Mandal

Outline

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

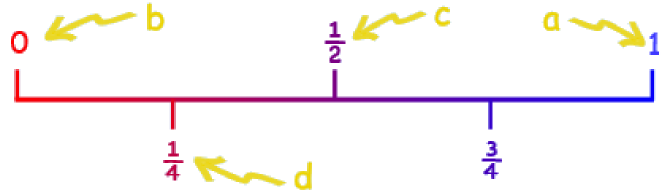
Probability

- 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.

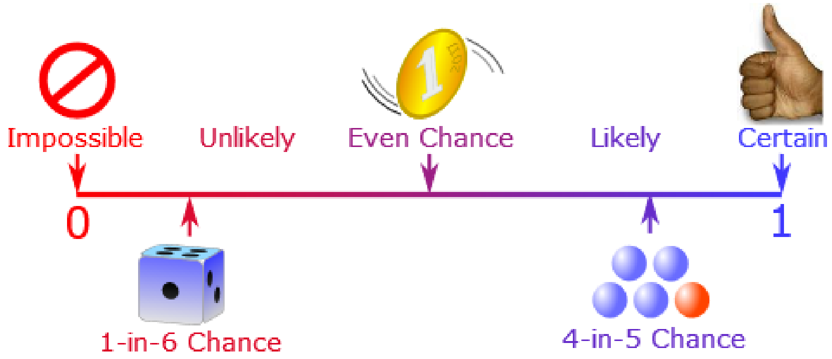




Probability Line

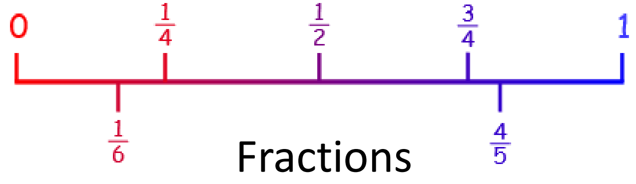
- Probability is the chance that something will happen.
  - Impossible = 0
  - Certain = 1

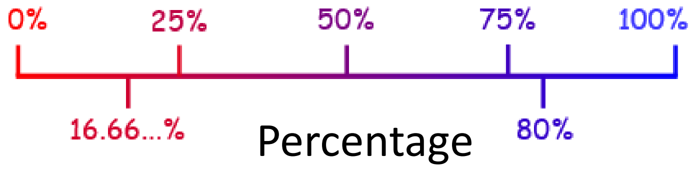


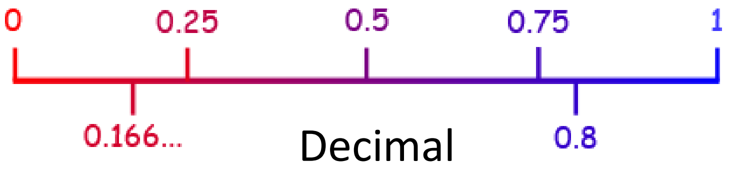
a) The sun will rise tomorrow
b) I will not have to know probability for this course
c) If I flip a coin it will land heads up
d) Choosing a red ball from a bag with 1 red ball and 3 green balls



- Representations:







Classical Definition

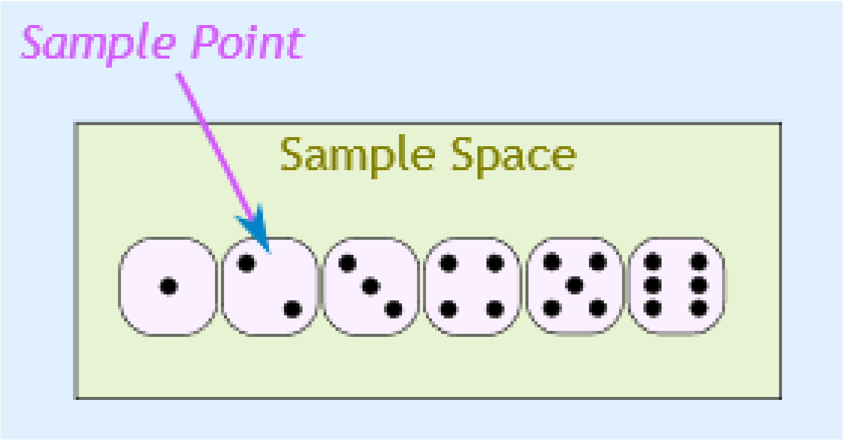
- Probability is defined as the ratio of the number of desired outcomes (what you want to happen) to the number of total possible outcomes (what could possibly happen).

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

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

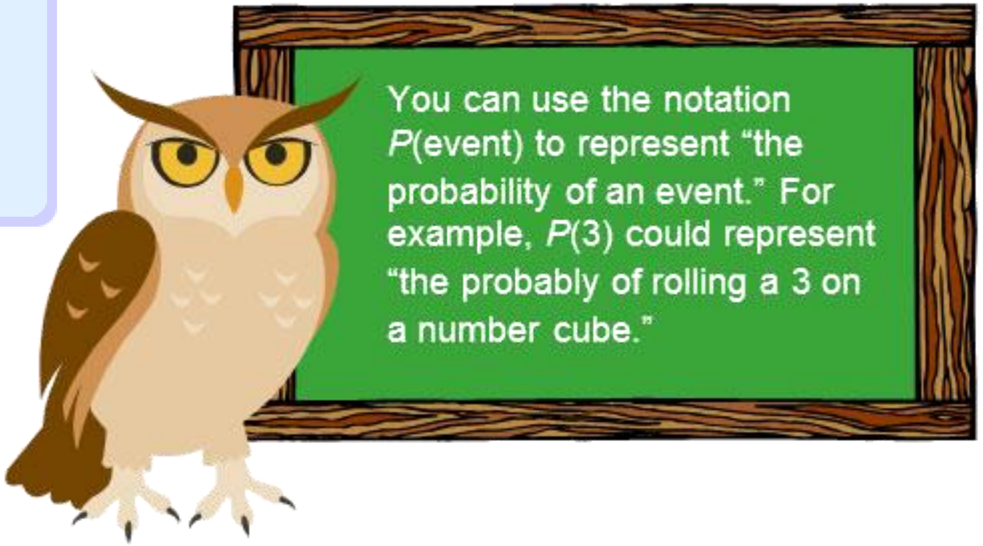
- Life is full of random events!
- Event means one (simple) or more outcomes (compound).

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



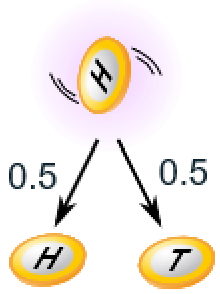
Independent Events

- 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

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

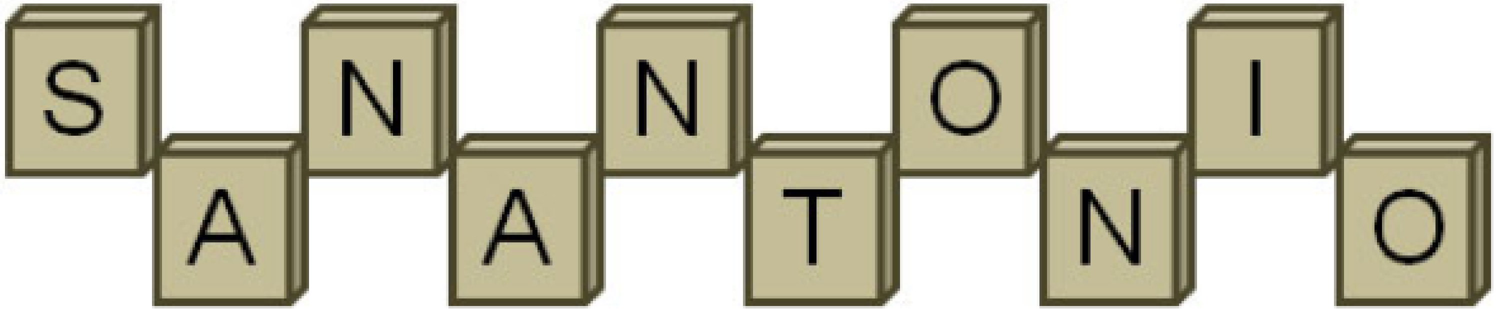
Two or more events that occur in a sequence. If the outcome of any event **does not** affect the possible outcomes of the other event(s), then the events are independent.

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

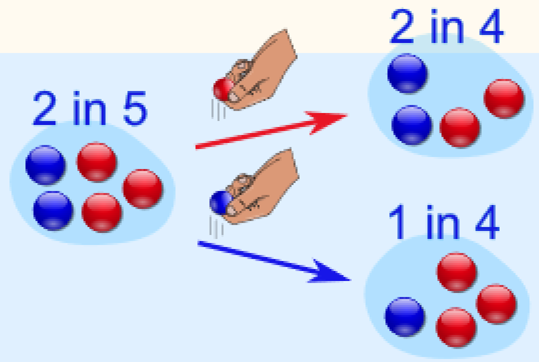
$$P(\text{vowel}) = \frac{\text{Number of Vowels}}{\text{Total Number of Tiles}} = \frac{5}{10} = \frac{5 \div 5}{10 \div 5} = \frac{1}{2}$$

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

Tree Diagram

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

Example: Marbles in a Bag



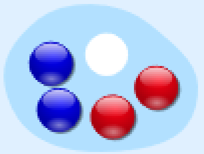
2 blue and 3 red marbles are in a bag.

What are the chances of getting a blue marble?

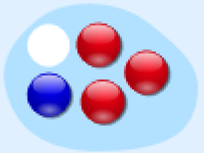
The chance is **2 in 5**

But after taking one out the chances change!

So the next time:



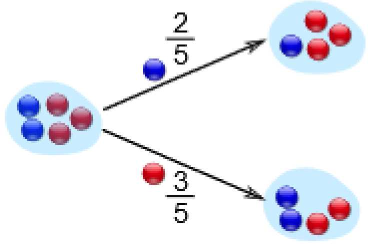
if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**



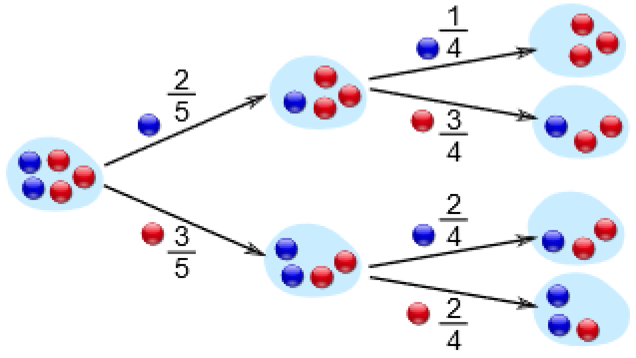
if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**

Tree Diagram (contd.)

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



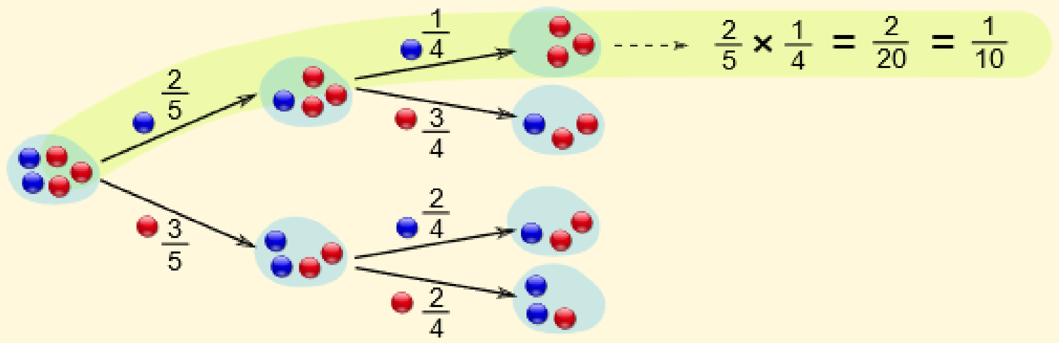
- We can go one step further and see what happens when we pick a second marble:



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?"**

Answer: it is a **2/5 chance** followed by a **1/4 chance**:



Did you see how we multiplied the chances? And got 1/10 as a result.

The chances of drawing 2 blue marbles is 1/10

Notations

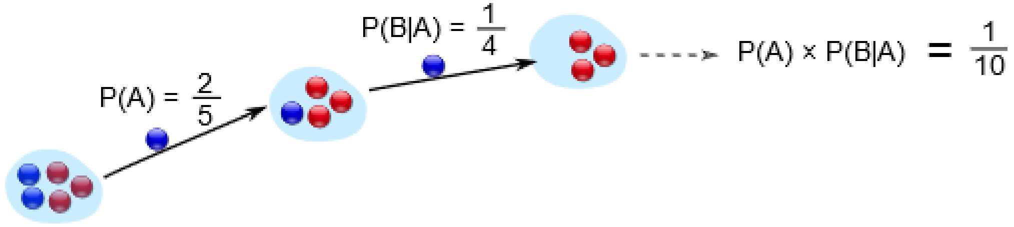
- P(A) means "Probability Of Event A“
- In our marbles example Event A is "get a Blue Marble first" with a probability of 2/5: **P(A) = 2/5**
- And Event B is "get a Blue Marble second" ... but for that we have 2 choices:
- If we got a **Blue Marble first** the chance is now **1/4**
- If we got a **Red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

"Probability Of"

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

"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:

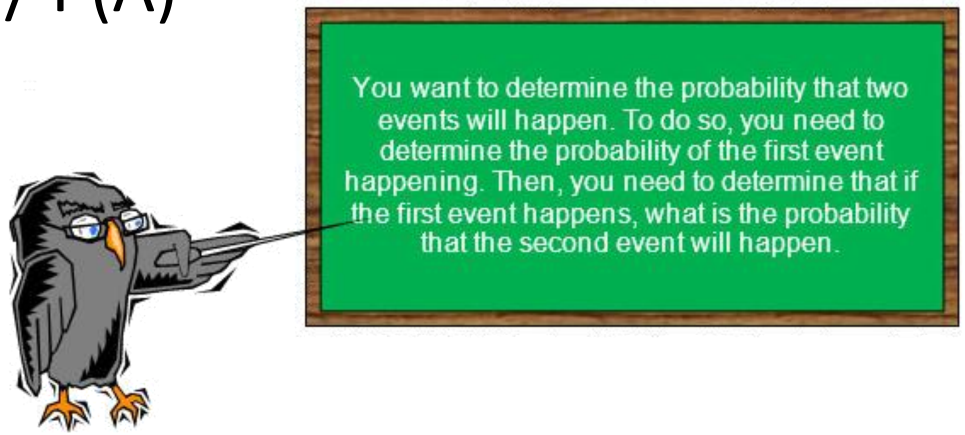
$$\mathbf{P(A\ and\ B) = P(A) \times P(B|A)} = (4/52) \times (3/51) = 12/2652 = \mathbf{1/221}$$

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

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



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

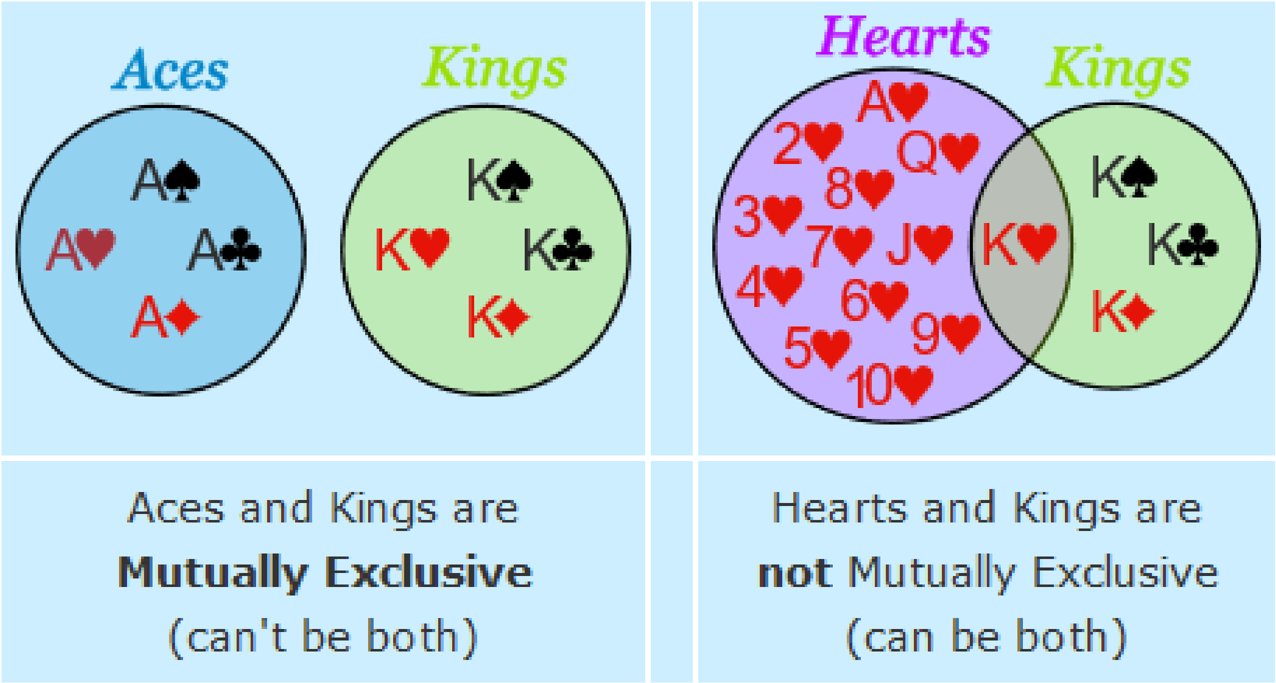
Example

- 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



Aces
Kings
A♠
A♥
A♣
A♦
K♠
K♥
K♣
K♦
Hearts
Kings
A♥
2♥
Q♥
8♥
3♥
7♥
J♥
K♥
4♥
6♥
5♥
9♥
10♥
K♠
K♣
K♦
Aces and Kings are
Mutually Exclusive
(can't be both)
Hearts and Kings are
not Mutually Exclusive
(can be both)

Mutually Exclusive

- When two events (call them "A" and "B") are Mutually Exclusive it is **impossible** for them to happen together: **P(A 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:

$$\mathbf{P(A\ or\ B) = P(A) + P(B)}$$

- *"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(King)=1/13**
- the probability of a Queen is also 1/13, so **P(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(King and Queen) = 0
- P(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(King $\cap$ Queen) = 0
- P(King $\cup$ 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% = **35%**

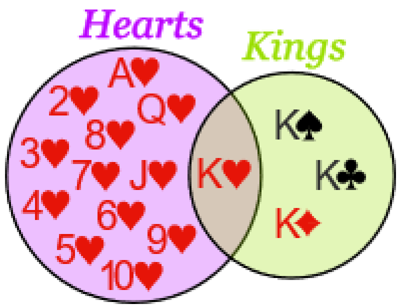
Which is written:

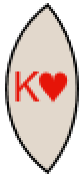
$$P(A \cap B) = 0$$

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

Not Mutually Exclusive

- **Example: Hearts and Kings**



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

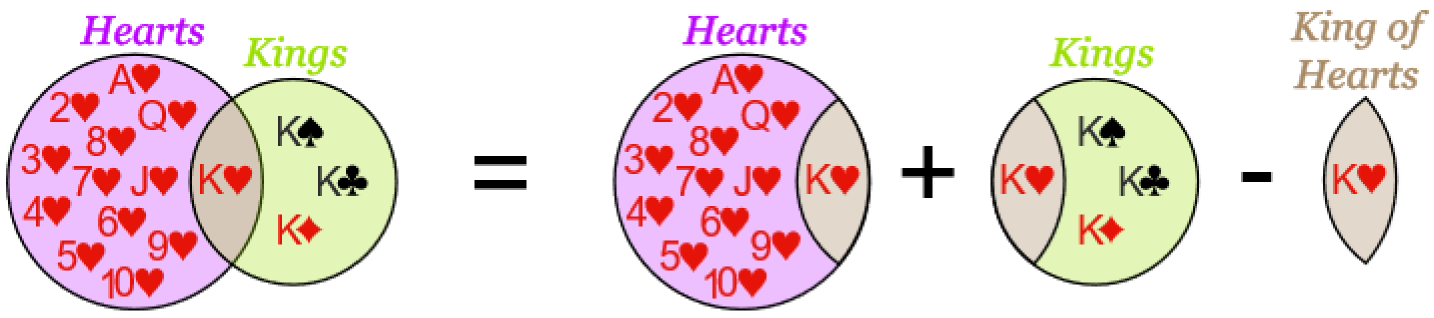
But Hearts **or** Kings is:

- 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 Cards = 13 Hearts + 4 Kings − the 1 extra King of Hearts

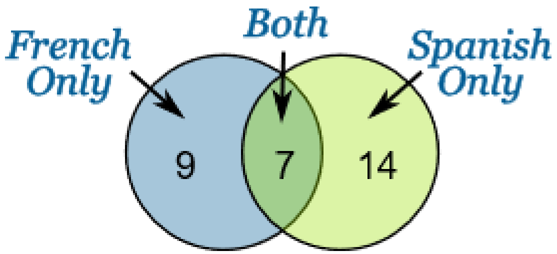
- **P(A or B) = P(A) + P(B) − P(A and B)**
- *"The probability of A **or** B equals the probability of A **plus** the probability of B **minus** the probability of A **and** B"*
- Here is the **same formula**, but using **U** and ∩:
- **P(A U B) = P(A) + P(B) − P(A ∩ B)**

Example

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

Example (contd.)

- P(French) = 16/30
- P(Spanish) = 21/30
- P(French Only) = 9/30
- P(Spanish Only) = 14/30
- P(French or Spanish) = 30/30 = 1
- P(French and Spanish) = 7/30



Lastly, let's check with our formula:
P(A or B) = P(A) + P(B) – P(A 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

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

- n(A)/n is called the relative frequency of event ***A***
- 0 ***<=*** n(A)/n <=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 : $P(A) >= 0$
- Axiom 2: $P(S) = 1$
- Axiom 3: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = 0$
- If the sample space **S** is not finite, then axiom **3** must be modified.

Elementary Properties

1. $P(\bar{A}) = 1 - P(A)$
2. $P(\varnothing) = 0$
3. $P(A) \leq P(B) \qquad \text{if } A \subset B$
4. $P(A) \leq 1$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. If $A_1, A_2, \ldots, A_n$ are n arbitrary events in S, then

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

7. If $A_1, A_2, \ldots, A_n$ is a finite sequence of mutually exclusive events in S ($A_i \cap A_j = \varnothing$ 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

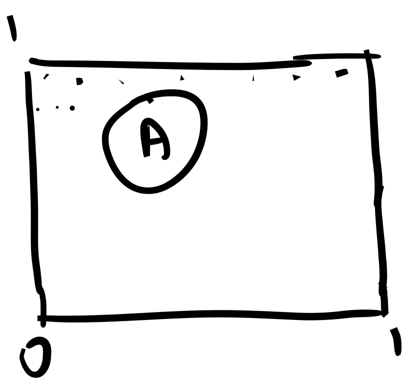
$$P(A) \leq 1$$

$$P(S) = 1$$

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

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

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



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

$$S = \bigcup_{x,y} \{x,y\}$$

$$P(S) = 1 = P\left(\bigcup_{x,y} \{x,y\}\right)$$

$$\neq \sum_{x,y} P(\{x,y\})$$

$$= 0$$

$A \cap B = \phi$

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

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

X = First roll

Y = 2nd "

→ B : $\min(X, Y) = 2$

→ $m = \max(x, y)$

(i) $P(m = 1 \mid B) = 0$

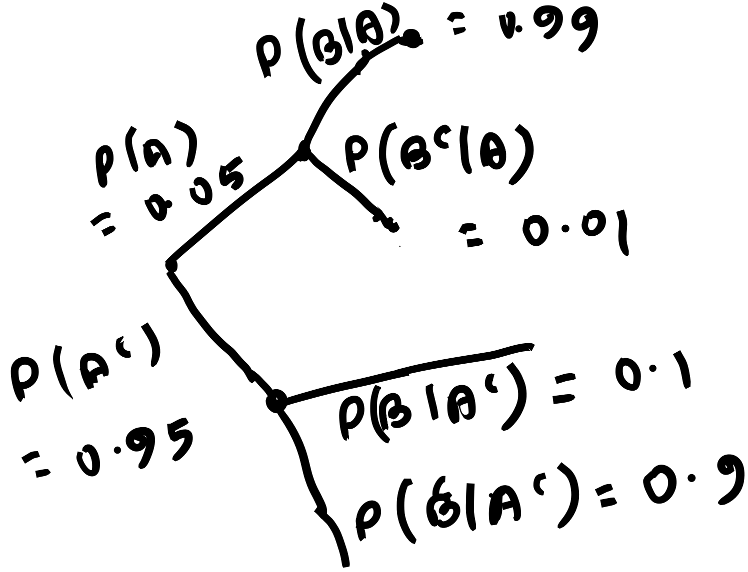
(ii) $P(m = 2 \mid B) = \dfrac{P(m = 2 \cap B)}{P(B)} = \dfrac{1/16}{5/16} = \dfrac{1}{5}$

* A : { an aircraft present }

B : { RADAR generates an alarm }

A^c = { aircraft not present }

B^c = { no Alarm }.



$$P(A\cap B) = P(A)\,P(B|A)$$
$$= 0.05 \times 0.99 = 0.0495$$

$$P(B) = P(A)\,P(B|A) + P(A^c)\,P(B|A^c)$$
$$= 0.1445$$

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

Total Probability Theorem:-

$$P(B) = P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2) + P(A_3)\,P(B|A_3) + P(A_4)\,P(B|A_4)$$

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

Bayes Rule

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

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

Cause $\qquad\qquad$ effect

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

$$A_i \xleftarrow[P(A_i|B)]{\text{inference}} B$$

Bayes Theorem

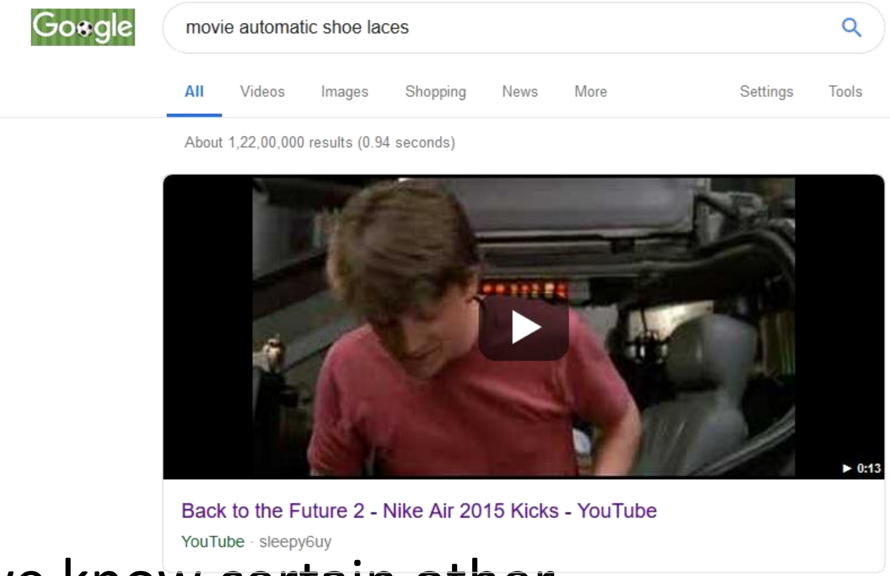
Bayes Theorem

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.



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

$$P(\text{Fire}|\text{Smoke}) = \frac{P(\text{Fire})\ P(\text{Smoke}|\text{Fire})}{P(\text{Smoke})}$$

$$= \frac{1\% \times 90\%}{10\%}$$

$$= 9\%$$

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(Rain|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(Rain) is Probability of Rain = 10%
- P(Cloud|Rain) is Probability of Cloud, given that Rain happens = 50%
- P(Cloud) is Probability of Cloud = 40%

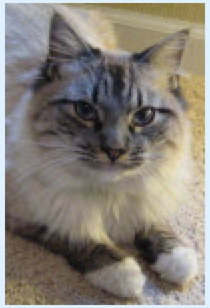
$$P(\text{Rain}|\text{Cloud}) = \frac{0.1 \times 0.5}{0.4} = .125$$

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(Allergy|Yes)**

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

  - P(Allergy) is Probability of Allergy = 1%
  - P(Yes|Allergy) is Probability of test saying "Yes" for people with allergy = 80%
  - P(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(Yes) = 1% × 80% + 99% × 10% = 10.7%
- Which means that about 10.7% of the population will get a "Yes" result.
- So now we can complete our formula:

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

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

- In fact we can write a special version of the Bayes' formula just for things like this:

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

For more cases

$$P(A1|B) = \frac{P(A1)P(B|A1)}{P(A1)P(B|A1) + P(A2)P(B|A2) + P(A3)P(B|A3) + \ldots etc}$$

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.

$$X : \Omega \longrightarrow \mathbb{R}$$

numerical value x

uppercase letter for R.V.
lowercase letter for real no.

PMF:-

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

- $p_X(x) \geq 0$
- $\sum_x p_X(x) = 1$

Geometric PMF:-

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

X : no of tosses until first head occurs.

X : Geometric R.V.

\# 1 Two independent rolls of a fair tetrahedral dice.

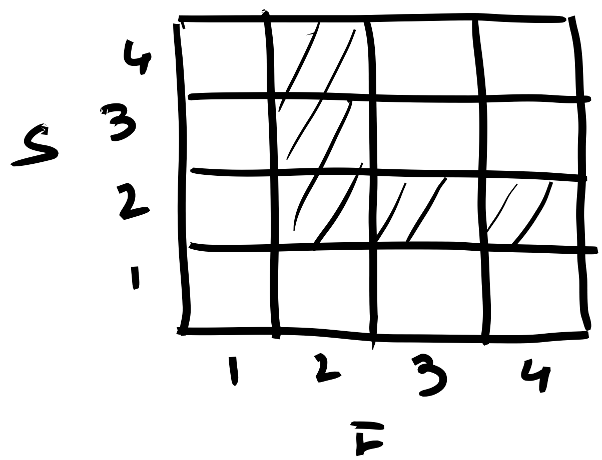
F : out-come of the first roll

S : " " " 2nd "

$X = \min(F, S)$

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

$p_X(0) = 0$



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

$P(H) = p$

let $n = 4$

$k = 2$

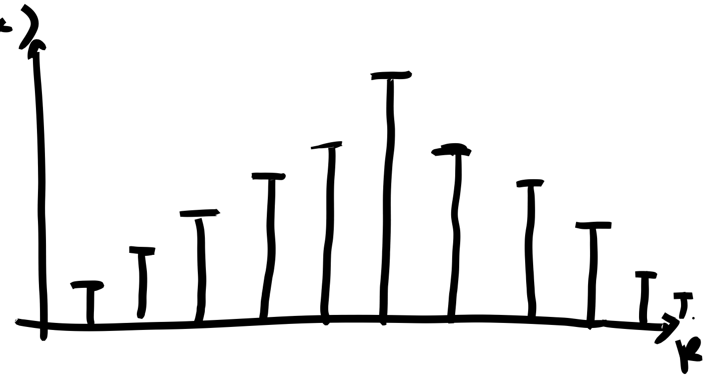
$p_X(2) = P(HTHT) + P(HHTT) + P(HTTH) + P(THHT) + P(THTH) + P(TTHH)$

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

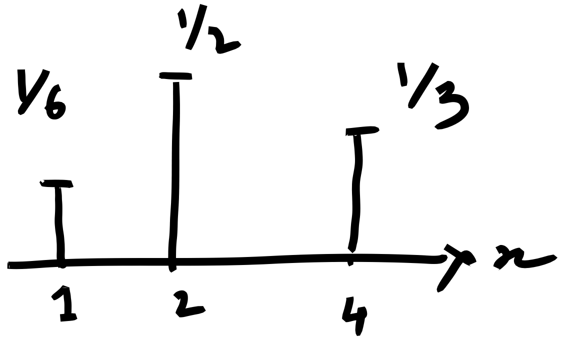
$p_X(2) = \binom{4}{2} p^2 (1-p)^{4-2}$

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

Binomial PMF



* Expectation:-



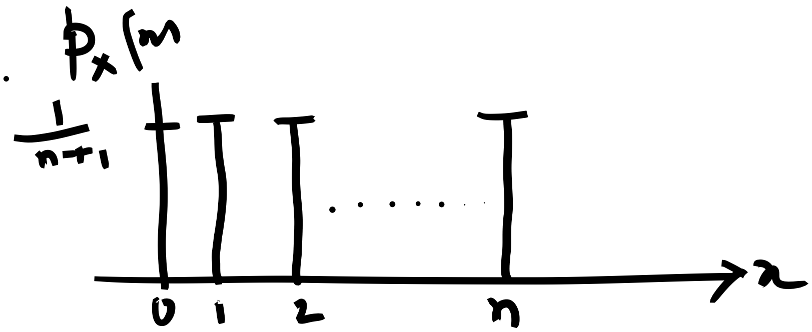
$$\frac{1}{6}\cdot 1 + \frac{1}{2}\cdot 2 + \frac{1}{3}\cdot 4$$

$$= 2.5$$

$$E[X] = \sum_x p_X(x)\cdot x$$

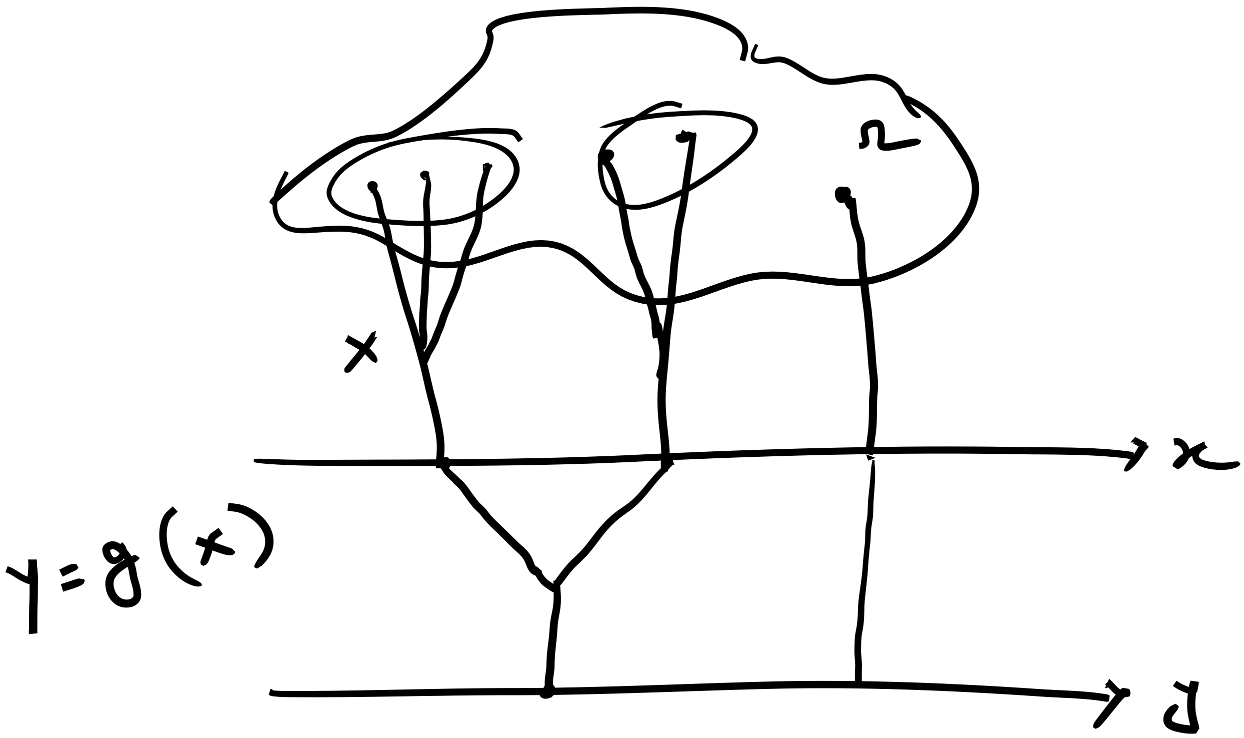
- Average in large no. of repetitions of the experiment
- Centre of gravity.

Example:-



$E[X] = 0 \times \frac{1}{n+1} + 1 \times \frac{1}{n+1} \cdots + n \times \frac{1}{n+1}$ = Answer.

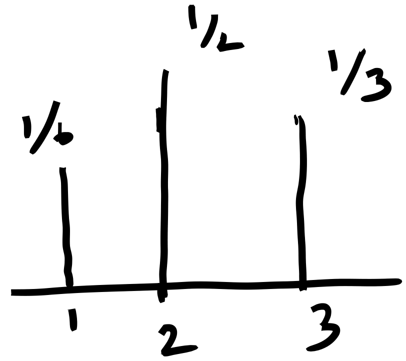
$E[X] = \frac{n}{2}$.



$$E[Y] = \sum_{y} y \, P_Y(y)$$

$$E[Y] = \sum_{x} g(x) \, P_X(x)$$

Lecture - 5



Expectation:-

$$E[X] = \sum_x x\, P(x)$$

$$= 2.5$$

#

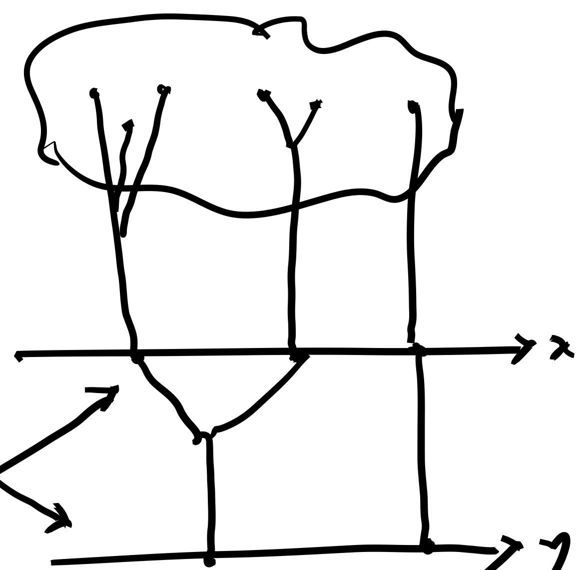
$$E[X] = 0 \times \frac{1}{n+1} + 1 \times \frac{1}{n+1} + 2 \times \frac{1}{n+1} + \cdots + n \times \frac{n}{n+1}$$

$$E[X] = \frac{n}{2}.$$

* $Y = g(X)$

$$E[Y] = \sum_y y\, p_Y(y)$$

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



Properties:- (i) $E[\alpha] = \alpha$

(ii) $E[\alpha X] = \alpha E[X]$

(iii) $E[\alpha X + \beta] = E[\alpha X] + E[\beta] = \alpha E[X] + \beta$

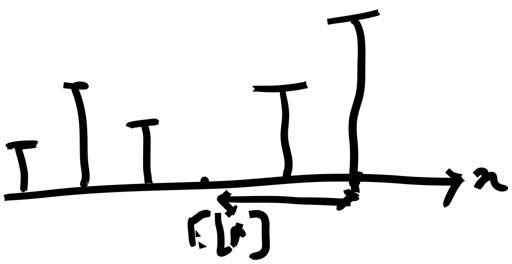
* $g(X) = X^2$

$$E[g(X)] = E[X^2] = \sum_x x^2 p_X(x)$$

second moment

* $g(X) = (X - E[X])^2$

$$\mathrm{var}(X) = E[g(X)] = \sum_x (x - E[X])^2 p_X(x)$$



$$= E[(X - E[X])^2]$$

$$= E[X^2] - (E[X])^2$$

Properties:

(i) $\mathrm{var}(X) \geq 0$

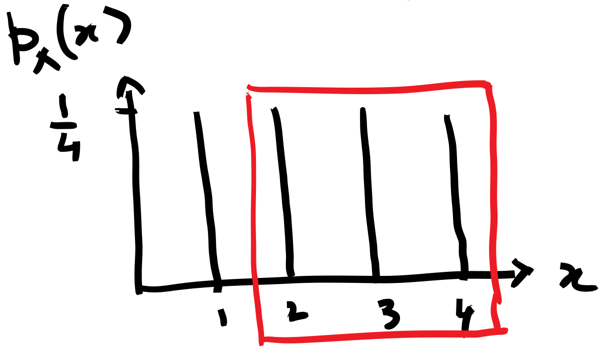
(ii) $\mathrm{var}(\alpha X + \beta) = \alpha^2 \mathrm{var}(X)$

$$\sigma = \sqrt{\mathrm{var}(X)}$$

(Standard deviation)

Conditional PMF :

$$p_{X|A}(x) \triangleq P(X = x \mid A)$$



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.

$$p_X(k) = (1-p)^{k-1} p \qquad k = 1, 2, \ldots\ldots$$

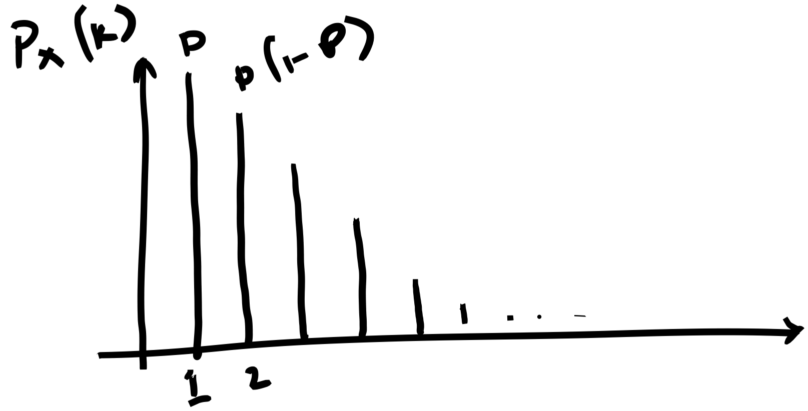
$$E[X] = \frac{1}{p}.$$

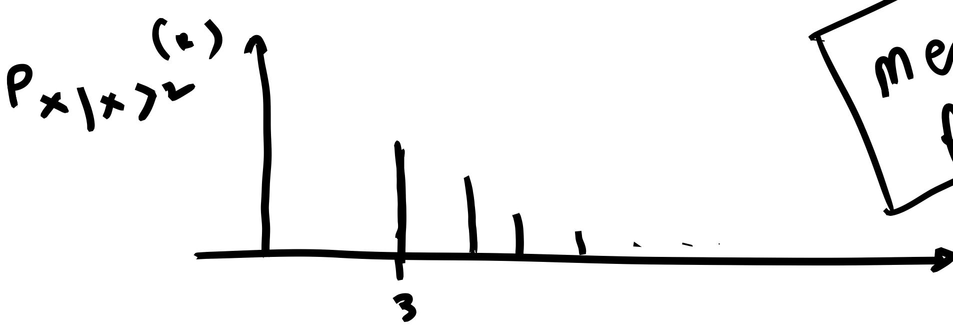
$X: \quad \underbrace{T\ T}_{2} \ \vdots\ \underbrace{T\ T\ T \cdots H}_{X-2}$

$Y: \quad \vdots\ T\ T\ T \cdots H$

$$p_Y(k) = (1-p)^{k-1}p \qquad k \geq 1$$

$$P(X-2 \mid X>2) = (1-p)^{k-1}p$$





$\boxed{\text{memoryless property}}$

Lecture - 6

X: Geometric R.V.

$$p_X(k) = (1-p)^{k-1} p \qquad k = 1, 2, \cdots$$

$$A_1, A_2, \cdots A_n$$

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

$$p_X(x) = P(A_1) p_{X|A_1}(x) + \cdots + P(A_n) p_{X|A_n}(x)$$

$$\sum_x x p_X(x) = \sum_x x \left(P(A_1) p_{X|A_1}(x) + \cdots + P(A_n) p_{X|A_n}(x) \right)$$

$$E[X] = P(A_1) E[X|A_1] + \cdots + P(A_n) E[X|A_n]$$

<u>Geometric</u>:

$$A_1 : \{X = 1\}$$

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

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

$$E[X|X>1] = E[X | X-1 > 0] = E[X-1 | X-1 > 0] + 1$$

$$= E[X] + 1$$

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

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

$-$

$pE[X] = 1$

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

<u>Joint PMF</u> $\quad p_{X,Y}(x,y) = P(X = x \text{ and } Y = y)$

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

$y=2$

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

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

- marginal PMF $\quad p_X(x) = \sum_y p_{X,Y}(x,y)$

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

- $p_{X|Y} = P(X = x \mid Y = y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$

Conditional PMF

$y = 2 \qquad 0,\ \frac{1}{5},\ \frac{3}{5},\ \frac{2}{5}.$

- $\sum_x p_{X|Y}(x|y) = 1$

- $p_{X,Y}(x,y) = p_X(x)\, p_{Y|X}(y|x)$

Multiple R.V.

$p_{X,Y,Z}(x,y,z)$

- $p_X(x) = \sum_{y,z} p_{X,Y,Z}(x,y,z)$

- $p_{X,Y,Z}(x,y,z) = p_X(x)\, p_{Y|X}(y|x)\, p_{Z|X,Y}(z|x,y)$

- $p_{X,Y,Z}(x,y,z) = p_X(x)\, p_Y(y)\, p_Z(z)$ $\boxed{\text{if } X,Y,Z \text{ independ.}}$

For Independent events:-

- $p_{X|Y}(x|y) = p_X(x)$

- $p_{X|Y,Z}(x|y,z) = p_X(x)$

*

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

(i) $p_{X,Y}(2,1) = \frac{1}{20}$

(ii) X & Y are not independ.

(iii) A: $X \le 2$ and $Y \ge 3$

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

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

conditionally independent.

* (i) If $X = Y$ $\quad var(X+Y) = ?$

$4\, var(X)$

(ii) If $X = -Y$ $\quad var(X+Y) = 0$

X, Y Dependent

(iii) If X & Y are independent and $Z = X - 3Y$

$var(Z) = var(X) + 9\, var(Y)$

Binomial Mean & Variance:

X = # of successes in n independent trials.

p = probability of success.

$E[X] = ?$ $\quad var(X) = ?$

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

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

$$\sum_i X_i = X$$

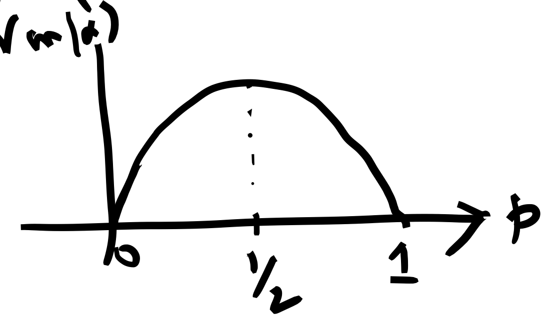
$$E[X] = \sum_i E[X_i] = \sum_{i=1}^{n} \left(1 \cdot p + 0 \cdot (1-p)\right) = np$$

$$\mathrm{var}(X_i) = p(1-p)^2 + (1-p)(0-p)^2$$

$$\mathrm{var}(X_i) = E[X_i^2] - (E[X_i])^2$$

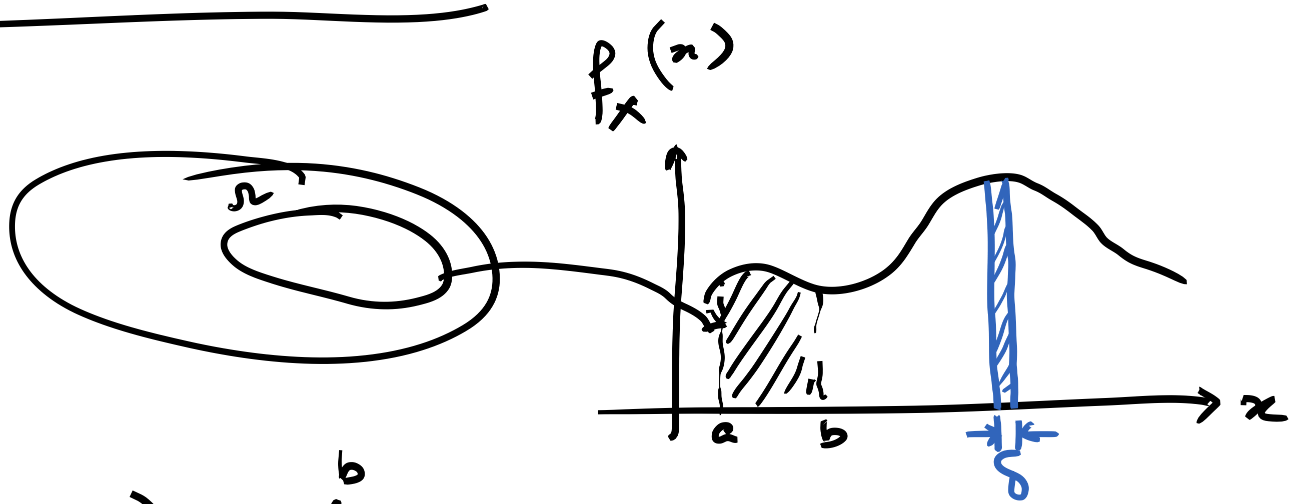
$$= p - p^2 = p(1-p)$$

$$\mathrm{var}(X) = \sum_i \mathrm{var}(X_i) = np(1-p)$$



Lecture - 7

Continuous R.V. :-



(i) $$P(a \le X \le b) = \int_a^b f_X(x)\,dx$$

(ii) $$P(X = a) = 0$$

(iii) $$f_X(x) \geq 0$$

(iv) $$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

(v) $$P(x \le X \le x+\delta) = \int_x^{x+\delta} f_X(x)\,dx = f_X(x)\,\delta$$

$$f_X(x) = \frac{P(x \le X \le x+\delta)}{\delta}$$

(vi) $$P(X \in B) = \int_B f_X(x)\,dx$$

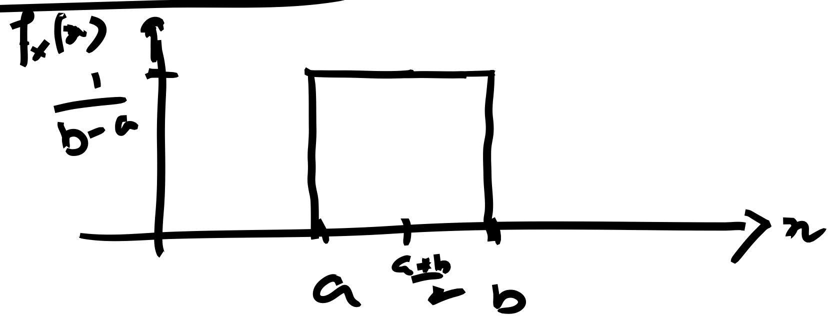
Mean and variance :-

- $E[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$

- $E[g(X)] = \int_{-\infty}^{\infty} g(x) \, f_X(x) \, dx$

- $\operatorname{var}(X) = \int_{-\infty}^{\infty} \left(x - E[X]\right)^2 f_X(x) \, dx$

Cont. Uniform R.V. :



$$f_X(x) = \frac{1}{b-a} \qquad a \le x \le b$$

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

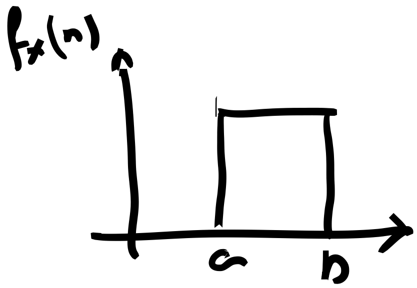
$$E[X] = \int_a^b \frac{1}{b-a} x \, dx = \frac{a+b}{2}.$$

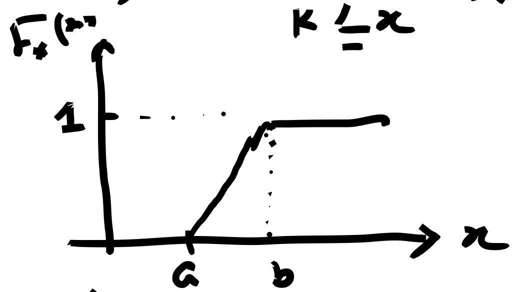
$$\sigma_X^2 = \frac{(b-a)^2}{12}$$

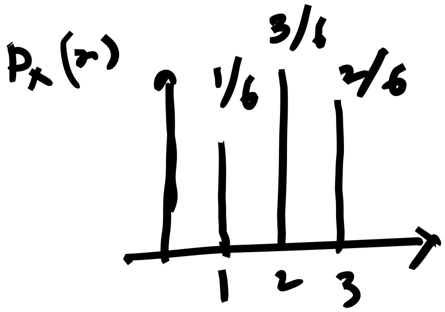
Cumulative Distribution Function (CDF)

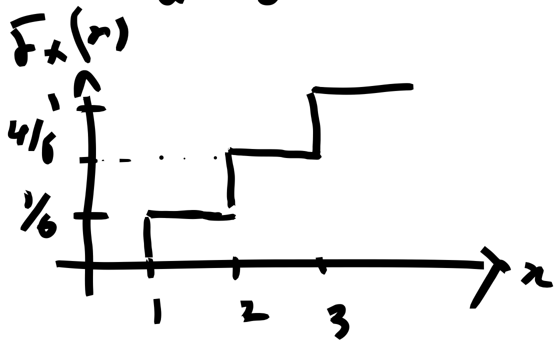
Cont. $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(\tilde{x})\, d\tilde{x}$

Dis: $F_X(x) = P(X \le x) = \sum_{k \le x} P_X(k)$



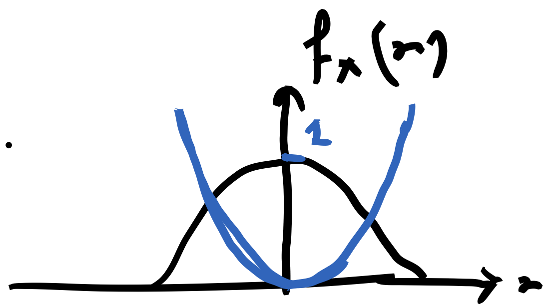






Gaussian (or normal) PDF:

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



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

$E[x] = 0$

$var(x) = 1$

General normal:

μ, σ^2

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

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

$\sigma \to$ Small

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

- Parabolla goes up quickly
- normal falls very fast
- Narrower density.

let $\quad Y = aX + b \quad [X \sim N(\mu, \sigma^2)]$

$$E[Y] = aE[X] + b = a\mu + b$$

$$Var(Y) = a^2 Var(X) = a^2\sigma^2$$

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

* $\quad CDF = \int_{-\infty}^{x} e^{-x^2/2}\,dx$

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

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

$\mu = 2 \qquad X \sim \mathcal{N}(2, 16)$

$\sigma^2 = 16$

$$P(X \leq 3)$$

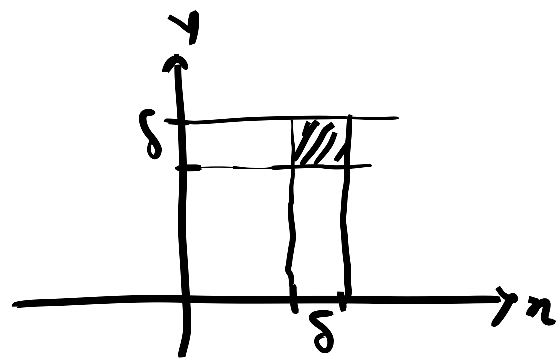
$$= P\left(\frac{X-2}{4} \leq \frac{3-2}{4}\right)$$

$$= CDF(0.25) = 0.5987$$

<u>multiple RV</u>: $\quad f_{X,Y}(x,y)$

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

(i) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$



(ii) $f_{X,Y} \geq 0$

(iii) $P(x \leq X \leq x+\delta,\ y \leq Y \leq y+\delta)$

$$\approx f_{X,Y}(x,y)\, \delta^2$$

(iv) $E[g(X,Y)] = \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy$

(v) $\displaystyle\int_{-\infty}^{\infty} f(x,y)\, dx = f_Y(y)$

(vi) $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$ for all x,y

* <u>Conditioning</u> :- $P(x \leq X \leq x+\delta)$
$\approx f_X(x)\,\delta$

(i) $P(x \leq X \leq x+\delta \mid Y \approx y) \approx f_{X|Y}(x|y)\,\delta$

(ii) $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$ 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 $X_1, X_2, \cdots$ be a sequence of i.i.d RVs each having mean μ and variance σ^2. Then the distribution of $\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ tends to standard normal as $n\to\infty$.

Random Process:-

$$\{X(t),\ t\in T\}$$

$$\omega\in\Omega$$

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

$\Rightarrow$ Fixed $t=t_k$ $\quad \boxed{X(t_k,\omega) = X_k(\omega)}$

Fixed $\omega=\omega_i$ $\quad X(t,\omega_i) = X_i(t)$

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

$$P_X(x_1,\cdots,x_n,t_1,\cdots,t_n) = P\{X(t_1)=x_1,\cdots\}$$

* If $x(t)$ is a continuous-time RP.

PDF

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

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

$(m, n)^{th}$ " " $\mu_{m,n} = \iint x^m y^n f_{x,y}(x, y)\, dx\, dy$

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

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

For R.P. :-

$$\underbrace{\mu_x(t)} = E[x(t)]$$

Ensemble average of $x(t)$

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

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

$t_i = t$

$t_j = t + \tau$

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

$$= R(\tau) = \text{Autocorrelation fn.}$$

<u>Autocovariance fn:</u>

$$K_x(t_i, t_j) = Cov[x(t_i)\, x(t_j)]$$

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

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

(i) <u>Stationary RP:-</u>

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

$t_i \in T \quad [i = 1, 2, \dots n]$

$$F_x(x_1, \dots x_n; t_1, \dots, t_n) = F_x(x_1, \dots, x_n; t_1+\tau, t_2+\tau, \dots t_n+\tau)$$

for all τ.

$K \leq n$

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

If $K = 2$: Wide-sense stationary RP.
(WSS)

(i) $E[x(t)] = \mu$

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

* Ergodic Process:-

$$\langle x(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, dt$$

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

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

Lecture - 9

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

Power Spectral Density:-

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

Inverse F.T.

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

* White noise:-

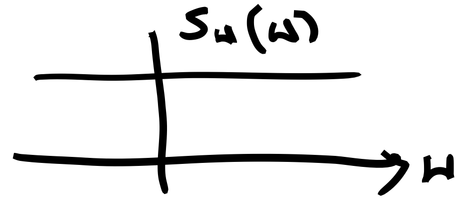
A continuous time white noise process $w(t)$ is a WSS zero-mean continuous time RP whose auto correlation f^n is given by

$$R_w(\tau) = \sigma^2 \delta(\tau) \qquad \left[\int_{-\infty}^{\infty} \delta(\tau)\, d\tau = 1\right]$$

$$S_w(w) = \int_{-\infty}^{\infty} R_w(\tau)\, e^{-jw\tau}\, d\tau$$

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

$= \sigma^2$.



Estimation Theory

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

$$\Big\downarrow \text{ Estimation Rule}$$

$$\hat{\theta}$$

Unbiased Estimator:-

$$\boxed{E[\hat{\theta}] = \theta} \quad \forall \theta, \qquad a < \theta < b$$

$x[n] = A + w[n]$ $\qquad n = 0, 1, \cdots, N-1$

$A \rightarrow$ the parameter.

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

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

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

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

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

$$= A$$

So, sample mean estimator is an unbiased estimator.

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

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

$$= A \quad \text{if } A = 0$$

$$\neq A \quad \text{if } A \neq 0$$

* $\{\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_n\}$

$$\boxed{\rho_{xy} = \frac{\operatorname{Cov}(x,y)}{\sigma_x \sigma_y}}$$

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i$$

$E[\hat{\theta}] = \theta$ [if $\hat{\theta}_i$ unbiased same variance & uncorrelated]

$\operatorname{var}(\hat{\theta}) = \frac{\sigma^2}{n}$ [$\operatorname{var}(\hat{\theta}_i) = \sigma^2$]

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

* For biased estimator

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

if $n \to \infty$; $E(\hat{\theta}_i) \neq \theta$

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

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

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

$$= var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$$

$$= var(\hat{\theta}) + b^2(\theta)$$ [For biased estimator]

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

$E[\check{A}] = aA$

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

$$\mathrm{mse}(\check{A}) = \frac{a^2\sigma^2}{N} + \left[aA - A\right]^2$$

$$\frac{d\,\mathrm{mse}(\check{A})}{da} = \frac{2a\sigma^2}{N} + 2(a-1)A \longrightarrow 0$$

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