5CS037-Concepts and Technologies of AI.
Week04, Lecture04

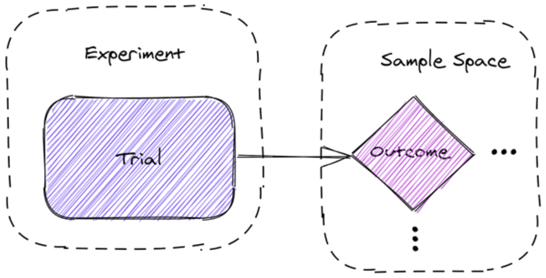
Discrete Random Variable
and
Their Probability Distributions.

Siman Giri

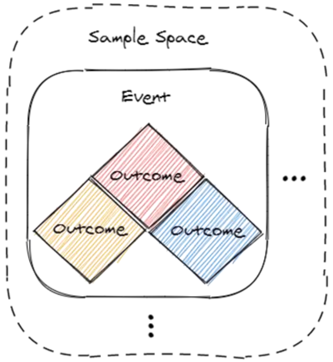
A. Revision and Preliminary Concepts.

A.1 From Last Week: Probabilistic Space.

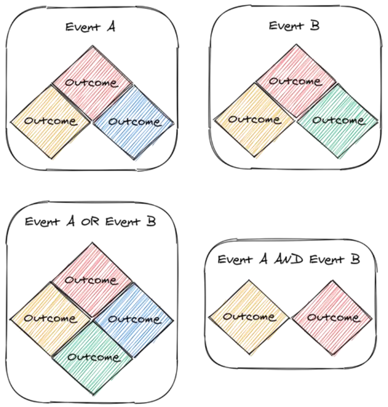
Elements of Probabilistic Space



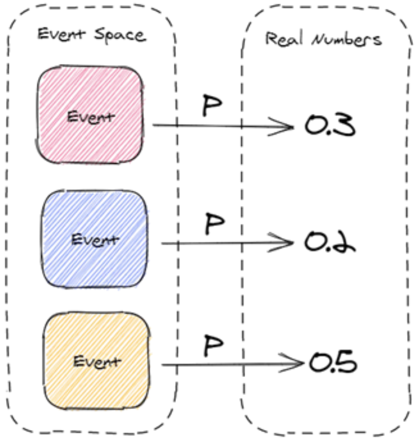
Probabilistic Experiment.



Sample Space.



Events.



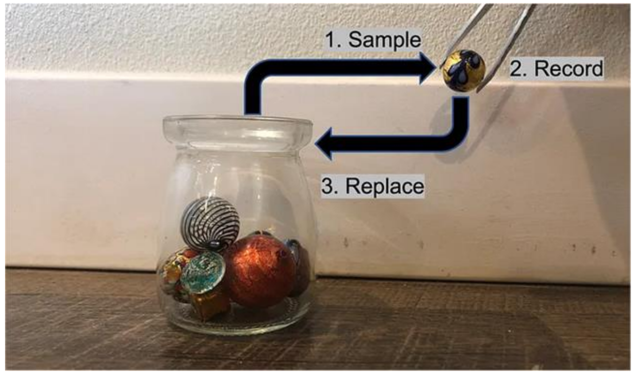
Probability Measure.

A.2 Preliminary Concepts: Statistical Experiment.

- A statistical experiment involves the observation of a sample selected from a larger body of data, existing or conceptual, called a population. The measurements in the sample, viewed as observations, are then employed to make an inference about the characteristics of the target population.
- How are these inferences made?
- The probability of the observed sample plays a major role in making an inference and evaluating the credibility of the inference.
- On the other hand, it is also clear that the method of sampling will affect the probability of a particular sample outcome.
- Let's review an example:

A.2 Preliminary Concepts: Statistical Experiment.

- For example: sampling without and with replacement,



- Thus to keep thing simple for our subsequent discussions of the probabilities associated with (random variables) we introduce simplest sampling procedure known as ***simple random sampling*** and ***random samples.***

A.3 Preliminary Concepts: Random Samples

- Simple Random Sampling and Random Samples:

Simple Random Sampling and Random Samples

Let **N** and **n** represent the numbers of elements in the **population** and **sample**, respectively. If the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples has an **equal probability** of being **selected**, the sampling is said to be **random**, and the result is said to be a **random sample.**

1. Random Variables

1.1Random Variables: Basic Definition.

- Remember the variables from programming?
  - Random variables are like programming variables, but with uncertainty.
- Random Variable is a numeric value assigned to each outcome of a chance/random experiment.
- For each element (outcomes) in a sample space, the random variable can take on exactly one value i.e. random variable must be measurable.
- Basic Definition:

Basic Definition of Random Variables.

A random variable $\mathbf{X}$ is a function from the sample space $\boldsymbol{S}$, as stated by the set of real numbers $\mathbb{R}$.

$$\boldsymbol{X}: \boldsymbol{S} \rightarrow \mathbb{R}$$

1.2Random Variables: Example.

- Experiment: Flip a three fair coins.
- Observations of interest{**Y**}: Number of **“heads”** on the **three coins**.
    - **Y** is an **Random Variable**, which **maps observation of interest** to **real numbers**.
    - **What numbers and How?**

1.2Random Variables: Example.

- Experiment: Flip a three fair coins.
- Observations of interest{Y}: Number of **"heads"** on the **three coins**.
  - Y is an **Random Variable**, which **maps observation of interest** to **real numbers**.
  - **What numbers and How?**
    - Which in general is answer to the question: **How many possible combinations of outcome for our observation of interest is possible?**
      - In above experiment **how many possible outcomes are possible**?
- Let us find out:

1.2Random Variables: Example.

- Experiment: Flip a three fair coins.
- Observations of interest{Y}: Number of **“heads”** on the **three coins**.
  - Y is an **Random Variable**, which **maps observation of interest** to **real numbers**.
  - **What numbers and How?**
    - Which in general is answer to the question: **How many possible combinations of outcome for our observation of interest is possible?**
      - In above experiment **how many possible outcomes are possible**?
- Let us find out:

Example:

Let Y be the number of heads on three coin flips:

0: Head Possible, then Y = 0: Probability of Y on 0 head is given by:
$P(Y=0) = 1/8$ {Events: (T,T,T)}

1: Head Possible, then Y = 1: Probability of Y on 1 head is given by:
$P(Y=1) = 3/8$ {Events: (H,T,T), (T,H,T),(T,T,H)}

2: Head Possible, then Y = 2: Probability of Y on 2 head is given by:
$P(Y=2) = 3/8$ {Events: (H,H,T), (H,T,H),(T,H,H)}

3: Head Possible, then Y = 3: Probability of Y on 3 head is given by:
$P(Y=3) = 1/8$ {Events: (H,H,H)}

4: Head Possible, then Y = 4: Probability of Y on 4 head is given by:
$P(Y \geq 4) = 0$ {Events: ()}

Thus our Random Variable takes value:

$Y = \{0, 1, 2, 3\}$

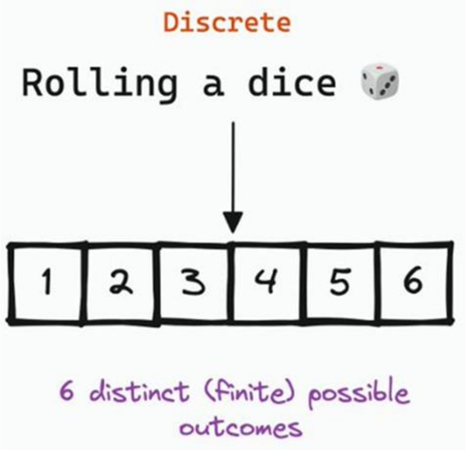
Random Variables Vs. Events

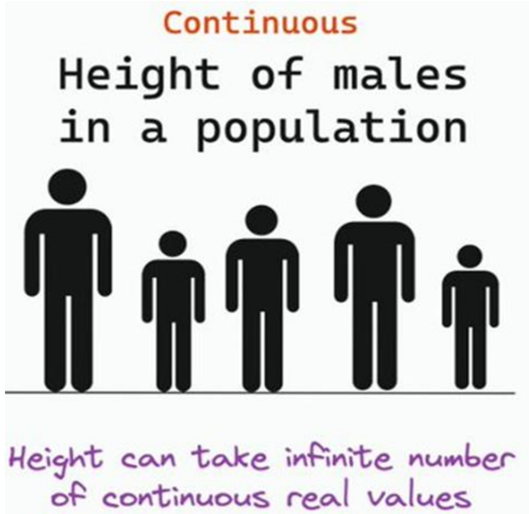
- Even though we use the same notation for random variables and for events (both use capital letters) they are distinct concepts.
- An event is a scenario, a random variable is an object.
- The scenario where a random variable takes on a particular value (or range of values) is an event.

Event	Meaning	Probability Statement
$Y = 1$	Y takes on the value 1 (there was one heads)	$P(Y = 1)$
$Y < 2$	Y takes on 0 or 1 (note this Y can't be negative)	$P(Y < 2)$
$X > Y$	X takes on a value greater than the value Y takes on.	$P(X > Y)$
$Y = y$	Y takes on a value represented by non-random variable y.	$P(Y = y)$

Random Variables: Types.

- Discrete Random Variable:
  - A random variable Y is said to be discrete if it can assume only a finite or countably infinite number of distinct values.
  - A random variable is discrete if its set of possible values is a collection of isolated points on the number line.
- Continuous Random Variable:
  - A random variable Y is said to be Continuous if it can assume any value defined in-between interval or range.
  - The variable is continuous if its set of possible values includes an entire interval on the number line.





2. The Probability Distribution for a Discrete Random Variable.

2.1 Probability Distribution.

- The probability of random variable (Discrete or Continuous) is a list of probabilities associated with each of its possible values.
- Probability Distribution for Discrete Random Variables are also known as Probability Mass Function.
- Topic for this week: Probability Mass Function.

2.2 Probability of Discrete Random Variable.

- Notationally, we will use an uppercase letter, such as Y , to denote a random variable and a lowercase letter, such as y, to denote a particular value that a random variable may assume.

$$\boxed{P(Y = y)}$$

- The expression (Y = y) can be read, the set of all points in S assigned the value y by the random variable Y .
- The probability that Y takes on the value y, P(Y = y), is defined as the sum of the probabilities of all sample points in S that are assigned the value y. We will sometimes denote P(Y = y) by p(y).

2.3 Probability Mass Function.

- The Probability mass function is probability distribution for a discrete variable Y is a representation of all the possible probabilities a random variable Y can take in the form of : a formula, a table, or a graph.
- The probability distribution for a discrete random variable assigns nonzero probabilities to only a countable number of distinct y values.
- Any value y not explicitly assigned a positive probability is understood to be such that $p(y) = 0$.
- Let's illustrate these ideas with an example.

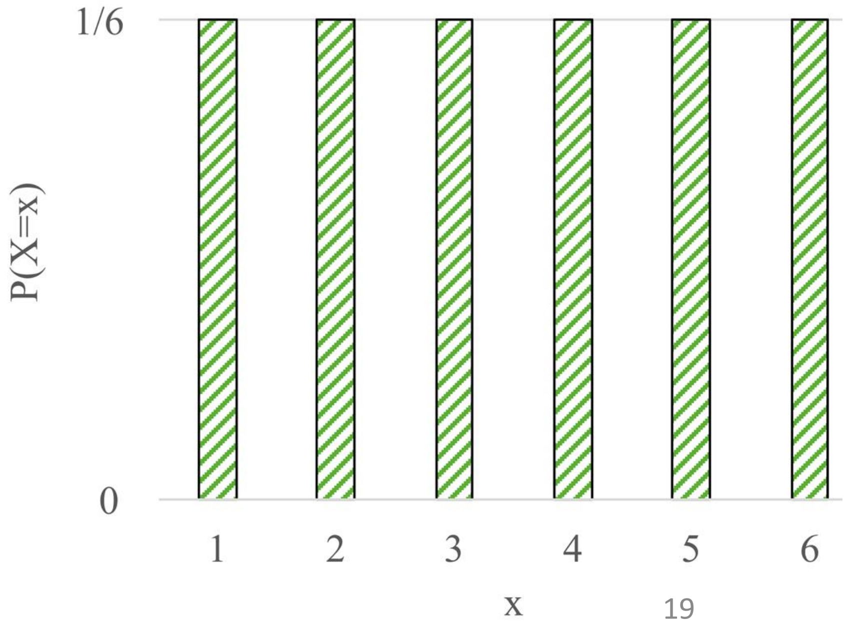
2.4 PMF: Example 1.

- PMF for Y: the value of a six sided die roll.
- By the definition:
  - P(Y=y) can take values of
    - P(Y =1); P(Y =2); P(Y =3); P(Y =4); P(Y =5); P(Y =6)
- As per definition of Probability Distribution,

we can arrange this in table, graph or formula.

Table

Value(y)	**1**	**2**	**3**	**4**	**5**	**6**
P(Y=y)	1/6	1/6	1/6	1/6	1/6	1/6



2.4 PMF: Example 2.

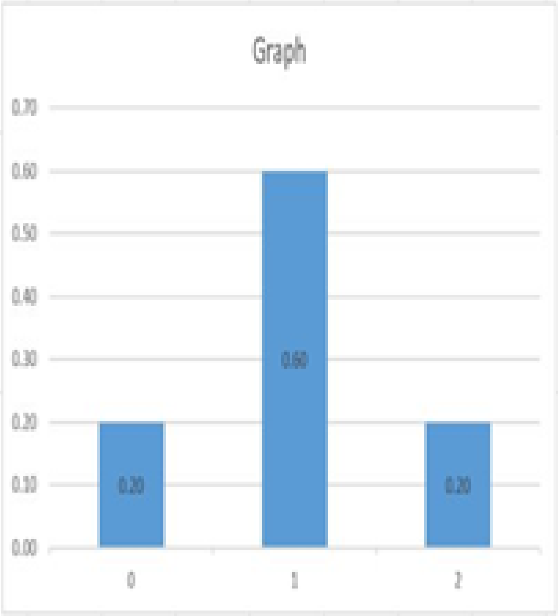
- A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y .
- *Solution:*

2.4 PMF: Example 2.

- *Solution:*
- Total possible outcomes = Total workers supervisor can select = $\binom{6}{2}$ = 15.
- **If Y = number of women in selection , then possible values for Y that can have non zero probability are {0, 1, 2}**
  - **Case1: zero women were selected:**
    - **P(Y = 0) = $\frac{\mathbf{3C0,3C2}}{\mathbf{15}} = \frac{\mathbf{1}}{\mathbf{5}}$.**
  - **Case2: One Woman and One Man are Selected:**
    - **P(Y = 1) = $\frac{\mathbf{3C1,3C1}}{\mathbf{15}} = \frac{\mathbf{3}}{\mathbf{5}}$.**
  - **Case3: Both the Worker are Women:**
    - **P(Y = 2) = $\frac{\mathbf{3C2,3C0}}{\mathbf{15}} = \frac{\mathbf{1}}{\mathbf{5}}$.**
- To Find the **Probability Distribution** we have to represent afore probability either in **Equation** or **Table** or **Graph**.

2.4 PMF: Example 2.

- *Represent in the form of formula or equation:*
- $P(Y=y) = \frac{3Cy,3C_{2-y}}{6C2}; \text{ for } y = 0,\ 1,2$



Table

y	p(y)
0	0.20
1	0.60
2	0.20

2.5 Properties of PMF.

- For any discrete probability distribution, the following must be true:
  - **$0 \le p(y) \le 1$ for all y.**
  - $\sum_y p(y) = 1$
- {where the summation is over all values of y with nonzero probability.}

3.The Expected Value of a Discrete Random Variable

3.1 Discrete Random Variable: Expectation.

- A random variable is fully represented by its probability mass function (PMF), which represents each of the values the random variable can take on, and the corresponding probabilities.
- Expectation can be thought as a Summary statistics of a Random Variable!

Definition: Expectation

The expectation of a random variable Y, written $E[Y]$ is the average of all the values the random variable can take on, each weighted by the probability that the random variable will take on that value.

$$\boldsymbol{E[Y] = \sum_{y} y . P(Y = y)}$$

Expectation goes by many other names: Mean, Weighted Average, Center of Mass, 1st Moment. All of which are calculated using the same formula.

If p(y) is an accurate characterization of the population frequency distribution, then E(Y) = μ, the population mean.

3.2 Properties of Expectations.

Property: Linearity of Expectation

$$E[aY + b] = aE[Y] + b$$

Where a and b are constants and not random variables.

Property: Expectation of Constant

$$E[a] = a$$

Sometimes in proofs, you will end up with the expectation of a constant (rather than a random variable). For example what does the E[5] mean? Since 5 is not a random variable, it does not change, and will always be 5, E[5]=5

Property: Law of Unconscious Statistician

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

One can also calculate the expected value of a function g(X) of a random variable X when one knows the probability distribution of X but one does not explicitly know the distribution of g(X). This theorem has the humorous name of "the Law of the Unconscious Statistician" (LOTUS), because it is so useful that you should be able to employ it unconsciously.

3.2 Properties of Expectations.

Property: Expectation of the Sum of Random Variables

$$E[X+Y] = E[X] + E[Y]$$

This is true regardless of the relationship between X and Y. They can be dependent, and they can have different distributions. This also applies with more than two random variables.

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Proof (Optional):

$$
\begin{aligned}
E[X+Y] &= \textstyle\sum_{x,y}(x+y)P(X=x,Y=y) \quad \{\text{Expected value of a sum.}\} \\
&= \textstyle\sum_{x,y}[xP(X=x,Y=y) + yP(X=x,Y=y)] \ \{\text{Distributive Property of sums.}\} \\
&= \textstyle\sum_{x,y} xP(X=x,Y=y) + \sum_{x,y} yP(X=x,Y=y) \{\text{Commutative property of sums.}\} \\
&= \textstyle\sum_x \sum_y \ xP(X=x,Y=y) + \sum_x \sum_y yP(X=x,Y=y) \ \{\text{Expanding Sums.}\} \\
&= \textstyle\sum_x x \sum_y P(X=x,Y=y) + \sum_y \ y \sum_y P(X=x,Y=y) \quad \{\text{Distributive property of sums.}\} \\
&= \textstyle\sum_x xP(X=x) + \sum_y yP(Y=y) \ \{\text{Marginalization.}\} \\
&= E[X] + E[Y] \ \{\text{Definition of Expectation.}\}
\end{aligned}
$$

3.1 Discrete Random Variable: Variance.

- In the last slide we showed that Expectation was a useful summary of a random variable
- it calculates the "weighted average" of the random variable.
- One of the next most important properties of random variables to understand is variance: the measure of spread.

Definition: Variance.

The variance is a measure of the "spread" of a random variable around the mean. Variance for a random variable, $\boldsymbol{Y}$, with expected value $\boldsymbol{E[Y] = \mu}$ is:

$$Var(Y) = E\left[(Y - \mu)^2\right]$$

Semantically, this is the average distance of a sample from the distribution to the mean. When computing the variance often we use a different (equivalent) form of the variance equation:

$$Var(Y) = E\left[Y^2\right] - E[Y]^2$$

4. Special Distributions

Bernoulli Distribution

4.1 What are Special Distributions?

- There are some specific distributions that are used over and over in practice, thus they have been given special names.
- Behind such distribution are a Random Experiments which
  - models a lot of real life phenomenon and
  - are frequently used in different applications including Machine Learning,
  - thus we devote a section to study them.
- For the discussed distribution (Discrete and Continuous):
  - We will try to understand the random experiment,
  - And provide a PMFs

4.3 Bernoulli Distribution: Trial

- Bernoulli Trial:
  - Bernoulli trial is an experiment with only two possible outcomes.
  - The two possible outcomes are labelled:
    - Success(S) and Failure (F).
  - The **probability of success** is
    - **$P(S) = p$**
    - and the **probability of failure** is **$P(f) = q = 1-p$**
- Examples:
  - Tossing a Coin
    - **{Success = H, Failure = T and p = P(S) = P(H)}**
  - Inspecting an item for defects
    - **{Success = defective, Failure = non-defective, and p = P(S) = P(defective)}**

4.4 Bernoulli Distribution: Experiment

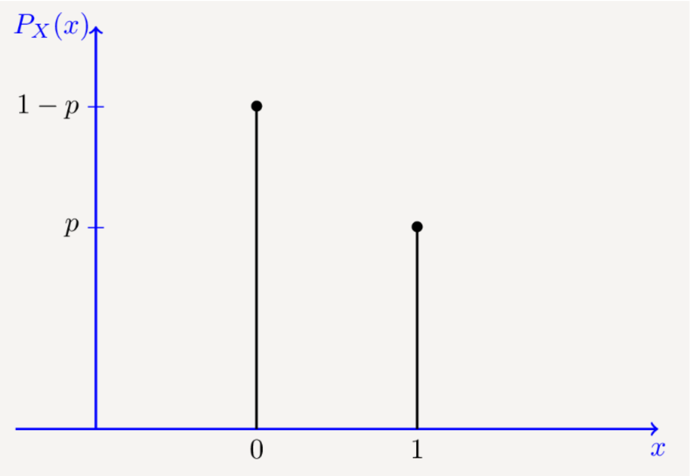
- The Bernoulli (Experiment) Process:
  - Bernoulli Process is an Random Experiment that must satisfy the following properties:
    - The Experiment consists of **n repeated Bernoulli trials**.
    - The **probability of success**,**{P(S) = p}** remains **constant** from **trial to trial.**
    - The **repeated trials** are **independent**, i.e. the **outcome** on one trial **has no effect** on the **outcome** of any other trial.

4.5 Bernoulli Distribution: PMF

Definition: Bernoulli Distribution and its PMF

A Random Variable Y is said to be a Bernoulli random variable with parameter p, shown (written) as $Y \sim Bernoulli(p)$, if it's PMF is given by:

$$P(Y = y) = \begin{cases} p & for\ x = 1 \\ 1 - p & for\ x = 0 \\ 0 & Otherwise \end{cases}$$



PMF of a $Bernoulli(p)$ random variable.

4.6 Bernoulli Distribution: Expectation

- Expectation of Bernoulli:
  - If **Y** is a Bernoulli with parameter p, **Y ~ Bernoulli(p),** then Expectation of Bernoulli Random Variable Y is :
    - $\boldsymbol{E[Y] = p}$
  - Proof(Optional):

$E[Y] = \sum_y y.P(Y = y)$	**{Definition of Expectation}**
$= 1.p + 0.(1 - p)$	**{Y can take on value 0 and 1}**
$= p$	**{Remove the zero term}**

4.6 Bernoulli Distribution: Variance

- Variance of Bernoulli:
  - If **Y** is a Bernoulli with parameter p, **Y ~ Bernoulli(p),** with Expectation$E[Y] = p$ then variance of Bernoulli Random Variable Y is :
    - $\mathbf{Var(Y) = p(1-p).}$
- Proof (Optional):

First Compute $E[Y^2]$:

$$
\begin{aligned}
E[Y^2] &= \textstyle\sum_y y^2 . P(Y=y) && \{\text{LOTUS}\}\\
&= 0^2.(1-p) + 1^2.p\\
&= p
\end{aligned}
$$

$$
\begin{aligned}
Var(Y) &= E[Y^2] - E[X]^2 && \{\text{Def. of Variance}\}\\
&= p - p^2 && \{\text{Substitute } E[X^2] = p \text{ and } E[X] = p.\\
&= p(1-p)
\end{aligned}
$$

5. Special Distributions

Binomial Distribution

5.1 Binomial Distribution: Experiment.

- Consider n independent (Bernoulli) trials of an experiment where each trial is a "success" with probability p.
- Example:
  - For 5 tosses of coin, what is the probability you flip exactly 3 heads. Let X be the random variable, Where X ={0, 1, 2, 3, 4, 5}. Find a PMF.

5.1 Binomial Distribution: Experiment.

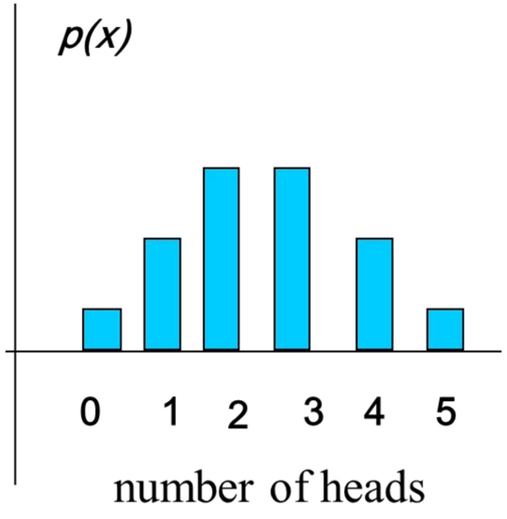
- Consider n independent (Bernoulli) trials of an experiment where each trial is a "success" with probability p.
- Example:
  - For 5 tosses of coin, what is the probability you flip exactly 3 heads. Let X be the random variable, Where X ={0, 1, 2, 3, 4, 5}. Find a PMF.

$\binom{5}{3}$ ways to arrange 3 heads in 5 trials

$_5C_3 = 5!/3!2! = 10$

Outcome	Probability
THHHT	$(1/2)^3 x (1/2)^2$
HHHTT	$(1/2)^3 x (1/2)^2$
TTHHH	$(1/2)^3 x (1/2)^2$
HTTHH	$(1/2)^3 x (1/2)^2$
HHTTH	$(1/2)^3 x (1/2)^2$
THTHH	$(1/2)^3 x (1/2)^2$
HTHTH	$(1/2)^3 x (1/2)^2$
HHTHT	$(1/2)^3 x (1/2)^2$
THHTH	$(1/2)^3 x (1/2)^2$
HTHHT	$(1/2)^3 x (1/2)^2$
10 arrangements	$x (1/2)^3 x (1/2)^2$

The probability of each unique outcome (note: they are all equal)



$$P(X=r)=\binom{n}{r}p^r(1-p)^{n-r}$$

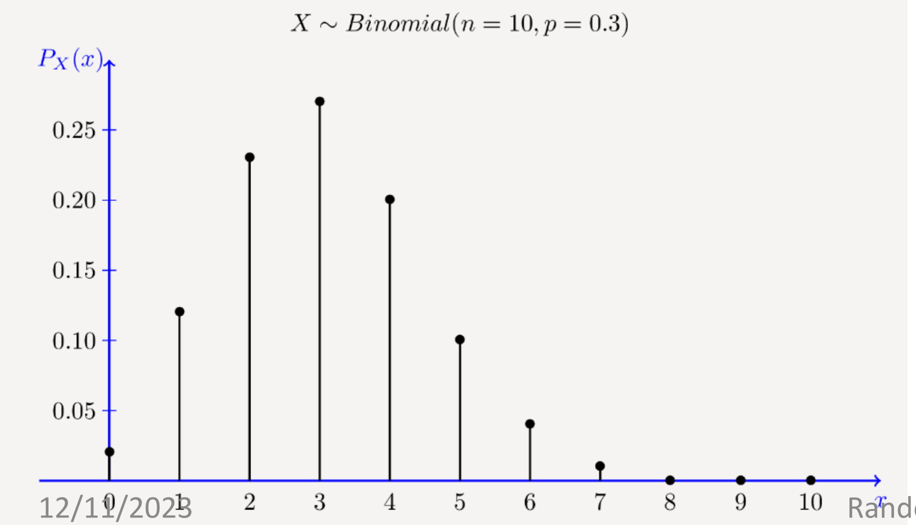
5.2 Binomial Distribution: PMF.

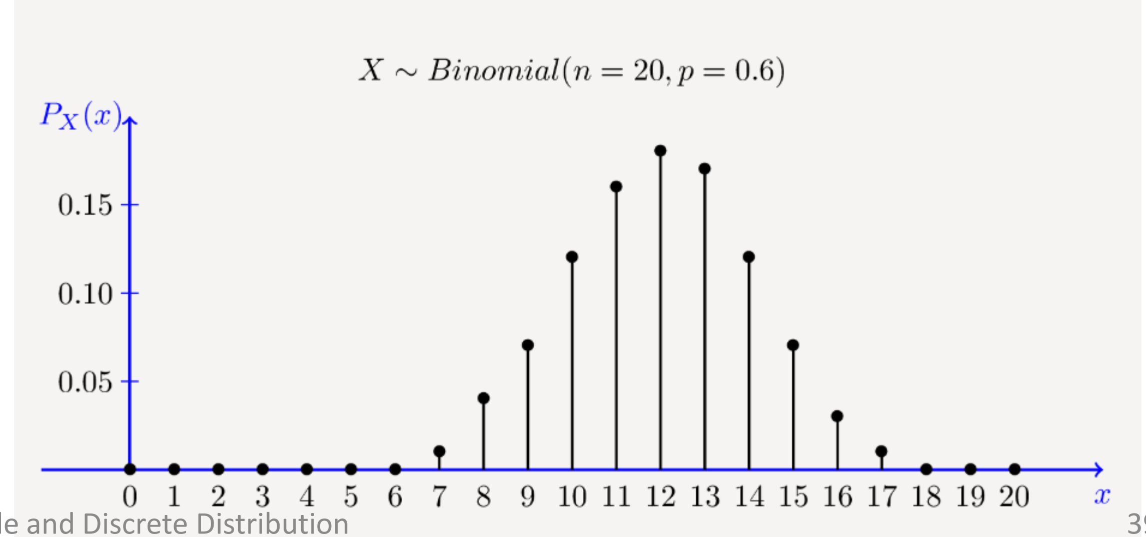
Definition: Binomial Distribution and its PMF

A random variable Y is said to be a binomial random variable with **parameters n and p**, shown as $Y \sim Binomial(n, p)$, if its PMF is given by:

$$P_Y(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & for\ k = 0, 1, 2 \dots, n \\ 0 & Otherwise \end{cases}$$

Where $0 < p < 1$.





5.2 Binomial and Bernoulli Distribution .

- Relationship is defined by:
  - If $Y_1, Y_2 \ldots, Y_n$ are independent Bernoulli(p) random variables, then the random variable X defined by $X = Y_1 + Y_2 + \cdots + Y_n$ has a Binomial (n, p) distribution.
    - $\boldsymbol{X \sim Bin(n, p) = \sum_{i=1}^{n} Y_i \sim Bern(p)}$

5.3 Binomial Distribution: E[Y] and Var(Y)

- Expectation:
  - $E[Y] = n.p$
- Variance:
  - $Var(Y) = n.p.(1 - p)$

- **Home Task:**
  - **Try proving above relationship as we did for Bernoulli Distribution.**

5.3 Binomial Distribution: Example.

- You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration. Then if you sample(randomly) 500 exposed people, how many do you expect to develop the disease?

5.3 Binomial Distribution: Example.

- You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration. Then if you sample(randomly) 500 exposed people, how many do you expect to develop the disease?
  - In this scenario, you can model the problem as a binomial distribution,
  - where each individual in the exposed group either develops the disease (success) with a probability of 0.05
  - or does not develop the disease (failure) with a probability 1−0.05=0.95.
- Expectation ~ Bin(n, p):
  - $E[Y] = n.p = 500 \times 0.05 = 25$

6. Special Distributions

Poisson Distribution.

6.1 Poisson Distribution: Intuition.

- The Poisson distribution is one of the most widely used probability distributions.
- It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space.
- In practice, it is often an approximation of a real-life random variable.
  - Suppose that we are counting the number of customers who visit a certain store from 1pm to 2 pm.
  - Based on data from previous days, we know that on average $\lambda = 15$ customers visit the store.
  - Here, we may model the random variable
    - Y: {showing the number customers} as a Poisson random variable (λ)

6.1 Poisson Distribution: PMF.

Definition Poisson Distribution and its PMF

A random variable Y is said to be a Poisson random variable with parameter λ, written as $Y \sim Poisson(\lambda)$, if its range is $R_Y = \{0,1,2,3,4 \ldots\}$, and its PMF is given by:

$$P_Y(k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!} & for\ k \in R_Y \\ 0 & otherwise \end{cases}$$