HCAI5DS02 – Data Analytics and Visualization.
Lecture – 03
Statistical Modeling: Turning Uncertainty into Insight.
Quantifying Uncertainty Using:
Random Variables and Probability Distributions.

Siman Giri

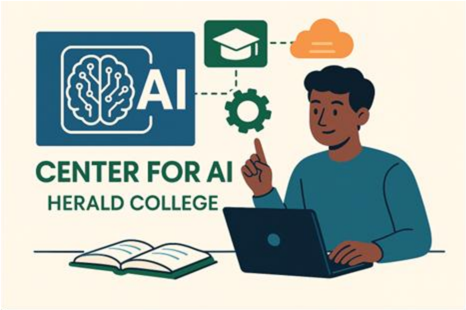


image generated via copilot.

1. Understanding The Random Variable.

{ Need, Motivation and Definition of a Random Variable.}

1.1 Recap: What did we Discuss Last Week.

- Last week, we built the foundation of probability thinking. We learned how to:
  - **Define a sample space (S)** — the set of all possible outcomes
    - *e.g., flipping a coin, rolling a die, or tracking customer behavior*
  - **Define events ($A \subseteq S$)** — outcomes of interest
    - *e.g., getting a 6 on a die, or receiving an email click*
  - **Use the axioms of probability** to measure how likely events are
    - ***e.g.,*** $\boldsymbol{P(A) \in [0, 1]}$, $\boldsymbol{P(S) = 1}$ …
  - **Apply conditional probability and Bayes' rule**
    - *e.g., Updating beliefs about fraud once a flag is triggered, or estimating conversion rates after observing user clicks*

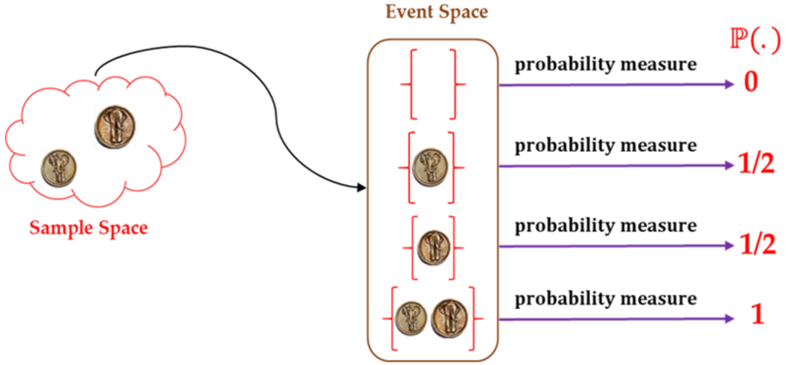
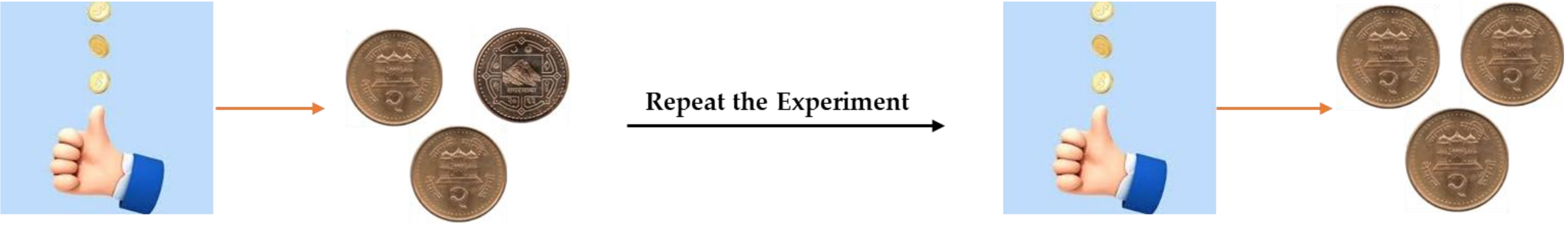


Fig: A Probabilistic Space.

←In a Figure:
Let $(\boldsymbol{S}, \boldsymbol{\mathcal{F}}, \mathbb{P})$ be a probability space, where
$\boldsymbol{S}$: Sample space (set of all possible outcomes).
$\boldsymbol{\mathcal{F}}$: Event space.
$\mathbb{P}$: $\boldsymbol{\mathcal{F}} \rightarrow [\mathbf{0}, \mathbf{1}]$: A probability measure.

? 1.2 But Here's a Problem We Haven't Solved Yet

- **Probability of Events — Yes, or No?**
  - We've been answering questions that deal with **binary events** — things that either happen or don't. For e.g.,:
    - **"What is the probability that a user clicks the email?"**
    - **"What is the probability of getting 2 heads in 3 tosses?"**
    - **"What is the probability that a transaction is fraudulent given it triggered an alert?"**
  - These questions help us understand **likelihood**, **update beliefs**, and **make decisions under uncertainty**.
  - But they all focus on ***whether*** something happens — **not** ***how much*** **or** ***how often***.
    - **Now We Want to Go Further: Quantity Matters**



1.2.1 Now we Want to Go Further: Quantity Matters.

- **In real business analytics, we often ask questions like:**
  - **"How many users will click the email out of 10,000?"**
  - **"What's the expected number of fraud alerts per day?"**
  - **"How much revenue can we expect from each customer?"**
  - **"What's the average wait time between two transactions?"**
- These are not yes/no events — these are **quantitative outcomes**.
- They require us to assign **numerical values to uncertainty**,
  - **not just label outcomes as true/false.**
- But to understand how that uncertainty unfolds in the real world — in terms of numbers we need a **mathematical function or model.**
  - That's where **random variables, probability distributions, and statistical models** come in."

1.3 So, What are Random Variables?

- In simple terms, a **random variable** is a **numerical value** $\mathbb{R}$ assigned to each outcome of a random or chance experiment.
- **Formal Definition:**
  - A **random variable** is a **function** that **maps** outcomes of a **sample space** to **real numbers:**
    - $\mathbf{X: \Omega \rightarrow \mathbb{R}}$
      - That means for each **outcome** $\boldsymbol{\omega \in \Omega}$, the **random variable** assigns a **single numeric** value:
        - $\mathbf{X(\omega) \in \mathbb{R}}$.
  - **Cautions:**
    - This **function must be measurable**,
      - meaning the **set of values**, it takes must align with **the structure of probability**
        - so, we can compute things like $\mathbf{P(X \leq x)}$.

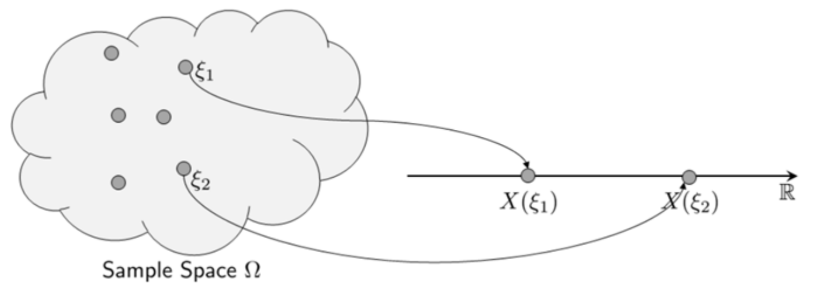


Fig: A definition of Random Variable.

To Summarize:

Concept	Description
Sample Space Ω	All possible outcomes of a random process.
Random variable X	A function assigning numbers to each outcome.
Output $X(\omega)$	A real number representing the outcome.

1.4 Random Variables in Practice.

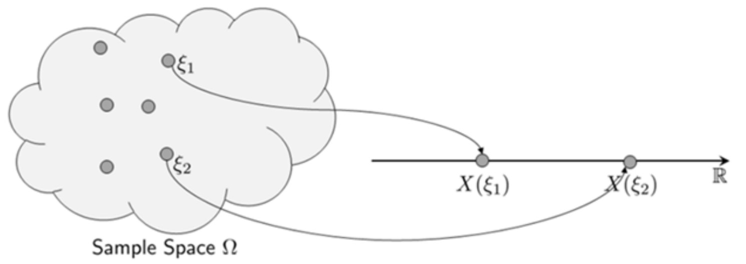


Fig: A definition of Random Variable.

- **Experiment: Flip Three Fair Coins**
  - We flip three fair coins and define a **random variable Y** :
  - If Y is a **random variable** then by definition
    - *It maps the outcomes of the experiment to real numbers.*
- **Big Question:**
  - **What Numbers and How?**
    - A **random variable** can technically **map outcomes** to any **real number**.
      - *But that doesn't mean we just assign numbers randomly!*
  - In practice, **we follow a logical convention**:
    - We assign numbers **that reflect the observation of interest**:
      - Let's say here, it's **the number of heads**.
        - **Y = Number of Heads.**
      - **How do we map to a Number?**
        - We ask a question: In our Experiment How many ways we can observe Head.

1.4.1 Random Variables in Practice.

- To Elaborate:
  - Our Experiment: Flip three Fair Coins.
  - My Sample Space: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
  - Random Variable: $Y \coloneqq$ Number of Heads.
  - Q: How many ways we can observe head.

Possible Head	Map to $\mathbb{R}$	Probability
0	$Y = 0$	Probability of Y on 0 head is given by: $P(Y = 0) = \frac{1}{8}$ {Possible Events $\Rightarrow$ (TTT)}
1	$Y = 1$	Probability of Y on 1 head is given by: $P(Y = 1) = \frac{3}{8}$ {Possible Events $\Rightarrow$ (HTT), (THT), (TTH)}
2	$Y = 2$	Probability of Y on 2 head is given by: $P(Y = 2) = \frac{3}{8}$ {Possible Events $\Rightarrow$ (HHT), (HTH), (THH)}
3	$Y = 3$	Probability of Y on 3 head is given by: $P(Y = 3) = \frac{1}{8}$ {Possible Events $\Rightarrow$ (HHH), (HTH)}
≥ 4	$Y \geq 4$	Probability of Y on 4 head is given by: $P(Y \geq 4) = \frac{\phi}{8} = 0$ {Possible Events $\Rightarrow \phi$}

1.4.2 Random Variables in Practice.

- Thus, from our example:
  - We defined: $\mathbf{Y = Number\ of\ Heads}$ → For 3 flip of a coins: $\mathbf{Y} \rightarrow \{\mathbf{0, 1, 2, 3}\}$
  - To generalize we write:
    - $(\mathbf{Y = y}): \mathbf{y} \in \{\mathbf{0, 1, 2, 3}\}$
  - We assign a probability as:
    - $\mathbf{P(Y = y)}$
  - For our example we can write:
    - $\mathbf{i.e.\ P(Y = 0), P(Y = 1), P(Y = 2), P(Y = 3)}$
- **Observation:**
  - In this example a Single Random Variable i.e. $\mathbf{Y = y}$ modeled three different **uncertainty**.
    - Through a set of countable numeric values $\mathbf{y} \in \mathbb{R}$, but
- **What If We *Can't Count* the Outcomes?**
  - For example:
    - The **observation of interest** is **not countable**?
    - The sample space is **continuous**, like **time**, **distance**, or **spend amount**?

1.5 Types of Random Variables.

Discrete Random Variable

- A **discrete random variable** is one that can take on **a countable number of distinct values.**
  - Often represents **counts** or **categories.**
  - Values can be listed individually (finite or countably infinite).
- **Examples:**
  - **Number of emails clicked in a campaign**
  - **Number of customers arriving per hour**
  - **Outcome of a dice roll $X \in \{1, 2, 3, 4, 5, 6\}$**
- **Formal:**
  - $\mathbf{X}: \boldsymbol{\Omega} \rightarrow \{x_1, x_2, x_3, \ldots\}$, **where each** $x_i \in \mathbb{R}$

Continuous Random Variable

- A **continuous random variable** can take on any **real value** within **an interval** (or union of intervals).
  - Often represents measurements like time, distance, weight, or revenue
  - Values are uncountably infinite
- **Examples:**
  - **Time until next customer arrives**
  - **Revenue generated by a customer**
  - **Temperature at noon**
- **Formal:**
  - $\mathbf{X}: \boldsymbol{\Omega} \rightarrow \mathbb{R}$, and $\mathbf{P(X = x) = 0}$ for any **single x**.
  - Instead, we compute probabilities over intervals,
    - like $\mathbf{P(a \leq X \leq b)}$.

1.5 Types of Random Variables.

Discrete Random Variable

- A **discrete random variable** is one that can take on **a countable number of distinct values.**
  - Often represents **counts** or **categories.**
  - Values can be listed individually (finite or countably infinite).
- **Examples:**
  - **Number of emails clicked in a campaign**
  - **Number of customers arriving per hour**
  - **Outcome of a dice roll $X \in \{1, 2, 3, 4, 5, 6\}$**
- **Formal:**
  - **$X: \Omega \rightarrow \{x_1, x_2, x_3, \ldots\}$, where each $x_i \in \mathbb{R}$**

Continuous Random Variable

- A **continuous random variable** can take on any **real value** within **an interval** (or union of intervals).
  - Often represents measurements like time, distance, weight, or revenue
  - Values are uncountably infinite
- **Examples:**
  - **Time until next customer arrives**
  - **Revenue generated by a customer**
  - **Temperature at noon**
- **Formal:**
  - **$X: \Omega \rightarrow \mathbb{R}$**, and **$P(X = x) = 0$** for any **single x**.
  - Instead, we compute probabilities over intervals,
    - like **$P(x \leq X \leq b)$.**

Let's first Concentrate on Discrete Example.

2. Modeling Discrete Outcomes with Random Variables.

{Quantifying Uncertainty When Outcomes Can Be Counted.}

2.1 Discrete Random Variable is …

- A **discrete random variable** is a variable that can take on a **finite or countably infinite set of values**, each with an associated probability.
  - It arises from **experiments** where **outcomes can be counted**, **not measured**.
- **Formal Definition:**
  - A discrete random variable $\mathbf{X}$ is a function: $\mathbf{X}: \boldsymbol{\Omega} \to \mathbb{R}$
    - such that the **image of** $\mathbf{X}$ is **countable**, i.e. $\mathbf{X}(\boldsymbol{\omega} \in \boldsymbol{\Omega}) = \{\, \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3} \ldots \}$
- **Key Characteristics:**
  - Takes on countable numeric values.
  - Probabilities are assigned directly to specific values: $P(X = x_i)$.
  - The sum of all probabilities is 1: $\sum_i P(X = x_i) = 1$.

Example - Experiment	Random Variable X	Possible Values
Tossing 3 Coins	Number of heads	$\{0, 1, 2, 3\}$
Email Campaign	Number of users who click	$\{0, 1, 2, \ldots, n\}$
Customer arrivals	Number of arrivals in an hour	$\{0, 1, 2, \ldots\}$

2.2 Example Use Case: Discrete Random Variable.

- **Scenario:**
  - **Email Campaign Response:**
- **Experiment:**
  - You send an email campaign to 3 customers. We want to study the response behavior i.e.
  - **"How many clicked the link?"**
- **Define Random Variable as per our Observation of Interest:**
  - Let's define a random variable **E = Number of users who clicked the email**.
  - This is a random variable
    - – it assigns a real number (**in this case, a count**) to the outcome of a **random experiment.**
- **Let's Map:**
  - To map we usually ask questions:
    - **"What is the natural numeric representation of our observation of interest?"**
      - **"How many ways we can represent our observation of interest?"**

2.2 Discrete Random Variable and Email Campaign Response.

- **How many outcomes are possible?**
  - Each customer has two possible outcomes: **Click (1) or No click (0).**
  - **Sample Space**: with **3 customers** possible outcomes could be:
    - $\mathbf{2^3 = 8}$
- **How many clicked the link?**

Outcome	Response Pattern	E = Clicked
No, No, No	(0, 0, 0)	0
Yes, No, No	(1, 0, 0)	1
No, Yes, No	(0, 1, 0)	1
No, No, Yes	(0, 0, 1)	1
Yes, Yes, No	(1, 1, 0)	2
Yes, No, Yes	(1, 0, 1)	2
No, Yes, Yes	(0, 1, 1)	2
Yes, Yes, Yes	(1, 1, 1)	3

- What are the Possible values of **E**?
  - $\mathbf{E \in \{0, 1, 2, 3\}}$
- We can also list the probabilities?
  - $\mathbf{P(E = 0) = \frac{1}{8}; P(E = 1) = \frac{3}{8}; P(E = 2) = \ ...}$

Looks Good, What is the Problem ….

2.2 Discrete Random Variable and Email Campaign Response.

- **How many outcomes are possible?**
  - Each customer has two possible outcomes: **Click (1) or No click (0).**
  - **Sample Space**: with **3 customers** possible outcomes could be:
    - $2^3 = 8$
- **How many clicked the link?**

Outcome	Response Pattern	E = Clicked
No, No, No	(0, 0, 0)	0
Yes, No, No	(1, 0, 0)	1
No, Yes, No	(0, 1, 0)	1
No, No, Yes	(0, 0, 1)	1
Yes, Yes, No	(1, 1, 0)	2
Yes, No, Yes	(1, 0, 1)	2
No, Yes, Yes	(0, 1, 1)	2
Yes, Yes, Yes	(1, 1, 1)	3

- What are the Possible values of E?
  - $E \in \{0, 1, 2, 3\}$
- We can also list the probabilities?
  - $P(E = 0) = \frac{1}{8}; P(E = 1) = \frac{3}{8}; P(E = 2) = \ ...$

- **Big Question:**
  - We do not send **email only to 3 customers …**
  - What happens if we **send email to 1000 customer …**
    - Or, **To Generalize we send email to n customer ….**

2.2.1 Email Campaign: We sent to 10,000 Users?

- **Now we want to know:**
  - **"What's the probability that exactly 50 users clicked the Email?"**
    - **Can you list all $2^{10,000}$ outcomes?**
      - **Definitely Not !!!**
  - Let's make it **more complex**: Suppose each user has **3 possible response**:
    - **Click, No Click, Purchase :**
      - Now the sample Space or the number of possible outcomes **becomes $3^{10,00}$**
      - **The Sample space becomes unimaginably large.**
- **Now imagine we are interested in more detailed questions like:**
  - **"How many clicked then purchase?"**
  - **" What is the probability that between 30 and 40 users clicked the email?"**
- **So How Do we Solve This?**
  - We need a way to **model this uncertainty efficiently.**
    - Something that gives us probabilities without listing all outcomes.
    - **Enters: The Probability Distribution Function.**

2.2.2 Probability Distribution Function: Why it Matters?

- **No need to enumerate massive sample spaces**
  - Avoid the impossible task of listing all outcomes in large experiments
- **Fast and scalable probability estimation**
  - Efficiently compute probabilities for real-world events
- **Answers complex questions like:**
  - What's the probability of conditional outcomes?
  - What's the chance that results fall within an interval (e.g., 30–40 clicks)?
  - What is the expected value or variance?
- **Enables modeling of complex random processes**
  - Handles multi-stage, multi-choice outcomes (e.g., click → purchase)
- **Computes exact or approximate probabilities**
- **Helps analyze uncertainty at scale**
  - Supports data-driven decisions in marketing, finance, operations, etc.

2.2.2 Probability Distribution Function: Why it Matters?

- **No need to enumerate massive sample spaces**
  - Avoid the impossible task of listing all outcomes in large experiments
- **Fast and scalable probability estimation**
  - Efficiently compute probabilities for real-world events
- **Answers complex questions like:**
  - What's the probability of conditional outcomes?
  - What's the chance that results fall within an interval (e.g., 30–40 clicks)?
  - What is the expected value or variance?
- **Enables modeling of complex random processes**
  - Handles multi-stage, multi-choice outcomes (e.g., click → purchase)
- **Computes exact or approximate probabilities**
- **Helps analyze uncertainty at scale**
  - Supports data-driven decisions in marketing, finance, operations, etc.

Question: "What Exactly is Probability Distribution Function?"

2.3 Probability Distribution Function.

- A **Probability Distribution Function** describes how probability is distributed over the possible values of a random variable whether **discrete or continuous**.
- **In General Terms:**
  - A **Probability Distribution Function** tells us:
    - What values a random variable can take.
    - And how likely each value (or range of values) is to occur
- **It can be represented in Different Forms:**
  - **Table: A list of values and their probabilities.**
  - **Graph: Bar chart for discrete, smooth curve for continuous**
  - **Formula/Equation: Mathematical expressions be like** $\mathbf{P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}}$

2.3.1 PDF: Discrete vs. Continuous.

For Discrete Random Variables

- We use the **Probability Mass Function (PMF)**
- It gives the probability of each **countable outcome:**
  - $\mathbf{P(X = x_i)}$
  - $\mathbf{\sum_i P(X = x_i) = 1}$

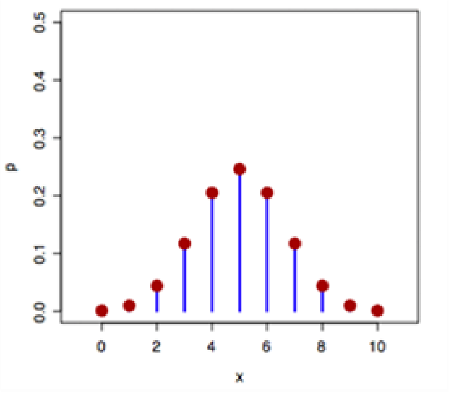


Fig: Discrete Probability Distribution

For Continuous Random Variables

- We use the **Probability Density Function (PDF).**
- Since continuous values are uncountable,
  - the probability of any exact value is
    - $\mathbf{P(X = x) = ?}$
- Instead, we calculate the probability over an interval:
  - $\mathbf{P(a \le X \le b) = \int_a^b f(x)\, dx}$

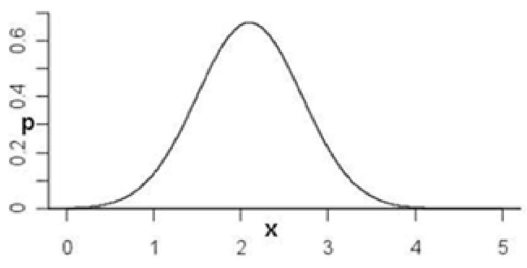


Fig: Continuous Probability Distribution

image from internet : may subjected to copyright.

2.3.1 PDF: Discrete vs. Continuous.

For Discrete Random Variables ✓

- We use the **Probability Mass Function (PMF)**
- It gives the probability of each **countable outcome:**
  - $\mathbf{P(X = x_i)}$
  - $\mathbf{\sum_i P(X = x_i) = 1}$

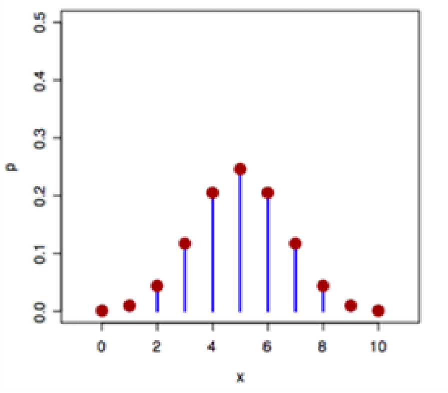


Fig: Discrete Probability Distribution

For Continuous Random Variables

- We use the **Probability Density Function (PDF).**
- Since continuous values are uncountable,
  - the probability of any exact value is
    - $\mathbf{P(X = x) = ?}$
- Instead, we calculate the probability over an interval:
  - $\mathbf{P(a \leq X \leq b) = \int_a^b f(x)\,dx}$

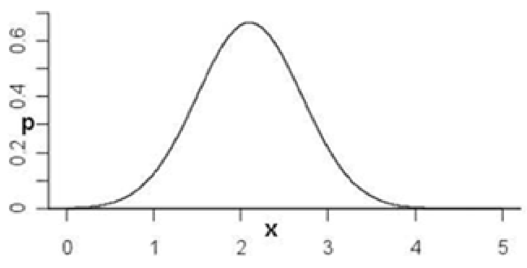


Fig: Continuous Probability Distribution
image from internet : may subjected to copyright.

3. Probability Mass Function (PMF).

{Describing the Likelihood of Countable Outcomes for Discrete Random Variables.}

3.1 PMF: A General Definition.

- A **Probability Mass Function (PMF)** is a function that describes the **probability of each possible value** that a **discrete random variable** can take.
- **In simple terms:**
  - A **PMF** tells us:
    - “What values the variable can take: e.g. {0, 1, 2, …}
    - “And how likely each of those values is to occur”
- **Formally:**
  - For a **discrete random variable X**, the **PMF** is a **function:**
    - $\mathbf{P(X = x) = p(x)}$ **such that:**
  - $\mathbf{p(x) \geq 0}$ **for all** $\mathbf{x \in Range(X)}$
  - $\mathbf{\sum_x p(x) = 1}$
- **Representations:**
  - A **PMF** can be expressed as:
    - A **table** listing values and their probabilities.
    - A **graph (bar chart)** showing probabilities visually.
    - A **formula or mathematical equations**.

3.2 PMF: Practical Example.

- **Scenario: Team Selection for A/B Testing**
  - A data science manager has a pool of 6 analysts:
    - **3 specialized in marketing analytics (M)**
    - **3 specialized in product analytics (P)**
  - To design a fair A/B test, she wants to randomly **select 2 analysts** to **analyze results**.
  - Let the **random variable Y** denote the **number of product analysts(P)** selected.
    - Find the **probability distribution for Y .**
- ***Solution:***

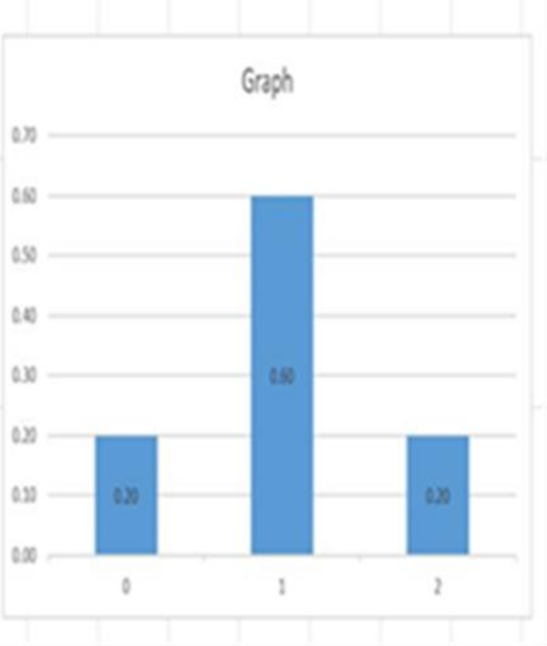
3.2 PMF: Practical Example.

- *Solution:*
  - **Total Possible Outcomes i.e.** ways manager can assemble team: $\binom{6}{2} = \mathbf{15.}$
  - Possible Values $\mathbf{Y}$ can take with **non zero probabilities:** $\mathbf{Y} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$
    - *Case 1: Select 2 marketing analysts.*
      - $\mathbf{P(Y = 0)} = \dfrac{\text{\# number of ways 2 marketing analyst could be picked}}{\text{\# Total Possible Combinations}} = \dfrac{\binom{3}{0}\binom{3}{2}}{\binom{6}{2}} = \dfrac{3}{15}$ ■
    - *Case 2: Select 1 marketing analysts and 1 product analyst.*
      - $\mathbf{P(Y = 1)} = \dfrac{\text{\# number of ways 1 marketing analyst and 1 product analyst could be picked}}{\text{\# Total Possible Combinations}} = \dfrac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}} = \dfrac{9}{15}$ ■
    - *Case 3: Select 2 product analysts.*
      - $\mathbf{P(Y = 2)} = \dfrac{\text{\# number of ways 2 product analyst could be picked}}{\text{\# Total Possible Combinations}} = \dfrac{\binom{3}{2}\binom{3}{0}}{\binom{6}{2}} = \dfrac{3}{15}$ ■
- *For PMF above possible values must be represented in the form of* **table, graph or equation.**

3.2 PMF: Practical Example.

- **PMF represented as an Equation:**
  - Let **Y** be the number of **product analysts** selected from **3 product and 3 marketing analysts**. Then:
    - $\mathbf{P(Y = y) = \frac{\binom{3}{y} \cdot \binom{3}{2-y}}{\binom{6}{2}}}$ **for** $\mathbf{y \in \{0, 1, 2\}}$
  - **PMF represented as a Table:**

Y: Product Analysts	Probability P(Y = y)
0	$\frac{3}{15} = 0.20$
1	$\frac{9}{15} = 0.60$
2	$\frac{3}{15} = 0.20$
Σ	1



PMF represented as Graph.

3.3 Transition to Empirical Distributions.

- So far, we used **combinatorics and reasoning** to **derive the PMF**.
  - But what if we **do not have theoretical assumptions** or let's say knowledge about **Experiment** that produces the data which are outcomes in sample space.
- In practice we often deal with observed data like the number of
  - daily website visitors, clicks per email batch or purchases per user.
    - That's where we use **empirical distributions** to **estimate PMF** from observed frequencies.
- **For example: Marketing Email Campaign:**
  - We ran **25 batches of an email campaign**; each batch **sent to 100 users** and following data were observed:
  - Number of users who clicked in each batch:
    - **[2, 1, 0, 3, 2, 1, 2, 0, 4, 1, 3, 2, 1, 1, 2, 3, 1, 0, 2, 3, 1, 2, 2, 1, 1]**
  - **Task: Compute the {Empirical} PMF.**

3.3.1 Transition to Empirical Distributions.

- *Solution:*

1. **Count Frequencies:**

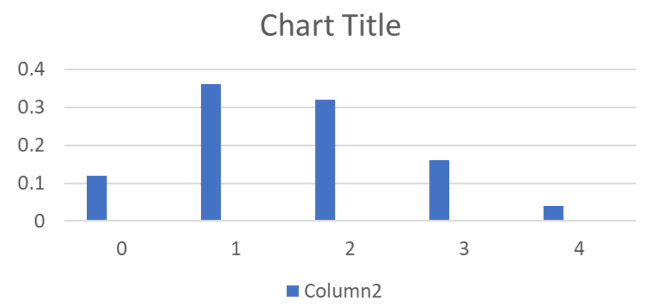
Number of Clicks:
[2, 1, 0, 3, 2, 1, 2, 0, 4, 1, 3, 2, 1, 1, 2, 3, 1, 0, 2, 3, 1, 2, 2, 1, 1]

Clicks (X)	Frequency
0	3
1	9
2	8
3	4
4	1
Total	25

2. **Convert to PMF (Relative Frequencies)**

- $P(X = x) = \frac{\text{Frequency of x}}{\text{Total Observations}}$

$X = x$	PMF: $P(X = x)$
0	$\frac{3}{25} = 0.12$
1	$\frac{9}{25} = 0.36$
2	$\frac{8}{25} = 0.32$
3	$\frac{4}{25} = 0.16$
4	$\frac{1}{25} = 0.04$



3.3.2 Empirical Distribution: Formal Definition.

- An **Empirical Discrete Distribution** is:
  - A **probability distribution** derived from **observed data**, where the probability of each unique outcome is the **relative frequency** of that outcome in **the data**.
- **Mathematical Formalization:**
  - Let:
    - $\boldsymbol{\delta_k}$**: the k-th distinct value observed in the dataset.**
      - In our **example unique observed value** $\boldsymbol{\delta_k}$ **are:** $\boldsymbol{\{0, 1, 2, 3, 4\}}$ **in**
        - **dataset** $\boldsymbol{[2, 1, 0, 3, 2, 1, 2, 0, 4, 1, 3, 2, 1, 1, 2, 3, 1, 0, 2, 3, 1, 2, 2, 1, 1]}$
    - $\boldsymbol{\omega_k}$**: the empirical weight (relative frequency) of** $\boldsymbol{\delta_k}$**.**
      - In our example relative frequency (weight) of outcome δ_k is:
        - $\boldsymbol{\omega_k = \frac{\text{count of } \delta_k}{\text{total observation}} \Rightarrow \sum \omega_k = 1}$
  - Then, For any *event* $A \subset \mathbb{R}$ *be any subset of outcomes*. Then the *empirical probability of A* is:
    - $\boldsymbol{p_{emp}(A) = \sum_{k=1}^{K} \omega_k \cdot \delta(A - \delta_k)}$
    - Where:
      - $\boldsymbol{\delta(A - \delta_k) = 1 \text{ if } \delta_k \in A \text{ ; } 0 \text{ otherwise}}$
  - This definition ensures:
    - We sum the weights of only those observed outcomes that are in the set A.
    - The result is a valid probability (between 0 and 1)

$X = x$	$\text{PMF}: P(X = x)$
0	$\frac{3}{25} = 0.12$
1	$\frac{9}{25} = 0.36$
2	$\frac{8}{25} = 0.32$
3	$\frac{4}{25} = 0.16$
4	$\frac{1}{25} = 0.04$

3.3.3 Empirical PMFs are useful But …

- Only with Empirical PMFs we may not be able to answer question like:
  - What is the probability that at least 5 users click in batch?
  - What if we scale this to 10, 000 batches?
  - What if we want to simulate new campaign?
- Challenges with Empirical Distributions:

Problem	Why It's a Problem
Only works on observed values	Can not generalize or extrapolate
Requires large datasets	Sparse or rare events are underrepresented
Hard to answer what if	No Mathematical structure
No easy parameterization	Can not simulate or optimize from it

- **Idea:** If we could model the underlying experiment we can use some of the known {Theoretical} mathematical distributions.

4. Foundations Before:
Exploring Common Discrete Distributions.

{Expectation, Variance, and Parameters of Discrete Random Variables.}

4.1 Expectation of Discrete Random Variable

- 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**} **- - Weighted Average** of a Random Variable!

Definition: Expectation

The expectation of a {discrete} 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:

$$\mathbf{E[Y]} = \sum_{\mathbf{y}} \mathbf{y}.\mathbf{p}(\mathbf{Y} = \mathbf{y})$$

Expectation goes by many other names such as:

Mean, weighted Average, Center of Mass, 1st Moment.

If $\mathbf{p(y)}$ is an **accurate characterization** of the **population frequency distribution**, then $\mathbf{E[Y]} = \boldsymbol{\mu}$ is the **population mean**.

4.2 Expectation of Discrete Random Variable: Example.

- **Scenario:**
  - A website sends a **marketing email** to users.
  - Let the random variable X represent the number of users who click in the link in a batch of 3 users. From past data, the empirical probability distribution (PMF) is:
  - From the past data, the empirical probability distribution (PMF) is:

$\mathbf{x(clicks)}$	**0**	**1**	**2**	**3**
$\mathbf{P(X = x)}$	**0.2**	**0.4**	**0.3**	**0.1**

- **What is the Expected Number of Clicks?**
  - $\mathbb{E}[\mathbf{X}] = \sum_{\mathbf{x}} \mathbf{x} \cdot \mathbf{P(X = x)} = \mathbf{0 \cdot 0.2 + 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.1 = 1.3} \blacksquare$
- **Interpretation:**
  - *On average, you can expect* ***1.3 clicks per 3 – user batch.***

4.3 Properties of Expectations.

- **Property: Linearity of Expectation**
  - $\mathbf{E[aY + b] = aE[Y] + b}$
    - Where a and b are constants and not random variables.
- **Property: Expectation of Constant**
  - $\mathbf{E[a] = a}$
  - 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**
  - $\mathbf{E[g(X)] = \sum_x g(x)P(X = x)}$
  - 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.

4.3 Properties of Expectations.

- **Property: Expectation of the Sum of Random Variables**
    - $\mathbf{E[X + Y] = E[X] + E[Y]}$
    - 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.
    - $\mathbf{E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]}$
    - **Sample – Proof:**

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

4.4 Variance of Discrete Random Variable.

- 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, $\mathbf{Y}$, with expected value $\mathbf{E[Y] = \mu}$ is:

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

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:

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

4.5 Parameters and Statistic: Definition

Parameters

- A **parameter** is a descriptive measure of the **entire population**.
- It is usually **unknown** and needs to be **estimated**.
- Parameters are **fixed** values
  - (but we usually don't know them exactly).
- Examples:
  - **Population Mean: μ**
  - **Population Variance: σ^2**
  - **True click trough rate in all customers: p**

Statistic

- A **statistic** is a descriptive measure computed from a **sample**.
- It is used to **estimate** population parameters.
- Statistics can **vary** from sample to sample.
- Examples:
  - **Sample Mean: $\bar{s}$**
  - **Sample Variance: s^2**
  - **Proportion of clicks in your test group: $\hat{p}$**

4.5.1 Cautions!!!

- **$\hat{p} \rightarrow$ Sample Proportion (Click Rate):**
  - A statistic computed from one Sample.
  - Represents the proportion of clicks in a single batch.
  - **Example (3 - users batch):**
    - **2 user clicked $\rightarrow \hat{p} = \frac{2}{3} \approx 0.67$**
    - Changes from:
      - batch (sample) to batch (sample).

- **$P(X = x) \rightarrow$ Probability of a count:**
  - A distribution tells you the likelihood of seeing a certain number of clicks.
  - **$X :=$ number of clicks in a batch**
  - Example:
  - **$P(X = 2) =$ Probability that exactly 2 out of 3 users click**.
  - Comes from PMF
    - either **theoretical or empirical**

3.3.2 Empirical Distribution: Formal Definition.

- An **Empirical Discrete Distribution** is:
  - A **probability distribution** derived from **observed data**, where the probability of each unique outcome is the **relative frequency** of that outcome in **the data**.
- **Mathematical Formalization:**
  - Let:
    - $\boldsymbol{\delta_k}$**: the k-th distinct value observed in the dataset.**
      - In our **example unique observed value** $\boldsymbol{\delta_k}$ **are:** $\{0, 1, 2, 3, 4\}$ **in**
        - **dataset** $[2, 1, 0, 3, 2, 1, 2, 0, 4, 1, 3, 2, 1, 1, 2, 3, 1, 0, 2, 3, 1, 2, 2, 1, 1]$
    - $\boldsymbol{\omega_k}$**: the empirical weight (relative frequency) of** $\boldsymbol{\delta_k}$**.**
      - In our example relative frequency (weight) of outcome δ_k is:
        - $\boldsymbol{\omega_k = \dfrac{\text{count of } \delta_k}{\text{total observation}} \Rightarrow \sum \omega_k = 1}$
  - Then, For any *event* $A \subset \mathbb{R}$ *be any subset of outcomes*. Then the *empirical probability of A* is:
    - $\boldsymbol{p_{emp}(A) = \sum_{k=1}^{K} \omega_k \cdot \delta(A - \delta_k)}$
    - Where:
      - $\boldsymbol{\delta(A - \delta_k) = 1 \text{ if } \delta_k \in A \,;\, 0 \text{ otherwise}}$
  - This definition ensures:
    - We sum the weights of only those observed outcomes that are in the set A.
    - The result is a valid probability (between 0 and 1)

$X = x$	PMF: $P(X = x)$
0	$\frac{3}{25} = 0.12$
1	$\frac{9}{25} = 0.36$
2	$\frac{8}{25} = 0.32$
3	$\frac{4}{25} = 0.16$
4	$\frac{1}{25} = 0.04$

5. Form Empirical to Theoretical Distribution (Discrete) .

{From Observation to Generalization.}

5.1 Empirical Vs. Theoretical Distributions.

Empirical Distribution.

- Based on observed frequencies from your sample data
- Useful for descriptive summaries and visualizing actual data
- Limitations:
  - Restricted to the collected data only
  - Cannot generalize to unseen or future data
  - **Do not explain how the data was generated**

Theoretical Distributions

- Model the **data-generating process** based on **assumptions**
- Provide a functional form to describe the data's behavior
- Enable simulation of new data points
- Allow calculation of probabilities for unseen events, e.g.,
  - **What is the probability $30 < X < 50$?**

5.2 What is a Data – Generating Process?

- The **Data-Generating Process** is the theoretical mechanism or "recipe" that produces the data we observe.
  - Since we rarely observe this process directly, we make **assumptions** about how data are generated to build a **statistical model**.
- These **assumptions** might include:
  - **Independence of events (e.g., each email click is independent)**
  - **Fixed probabilities (e.g., constant click-through rate)**
  - **Distribution shape (e.g., Binomial, Poisson)**
  - **Number of trials or events (e.g., number of emails sent)**
  - **By specifying these assumptions, we define a family of distributions that describe possible outcomes.**
- This framework allows us to:
  - **Predict probabilities of future events**
  - **Understand uncertainty**
  - **Perform inference about unknown parameters**

5.3 One Trial, Two Outcomes: Bernoulli Trial.

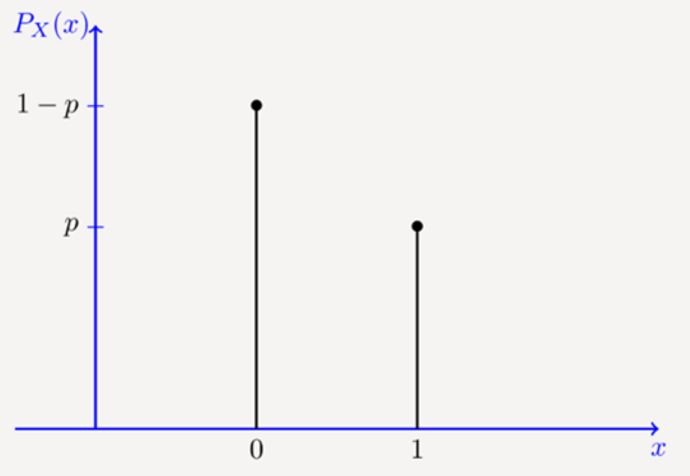
- A Bernoulli Trial is a random experiment that has only two possible outcomes:
  - **Success (usually coded as 1)**
  - **Failure (usually coded as 0)**
- Each **trial** is:
  - Independent of others
  - Has a fixed **probability of success**, denoted by ***p***

5.3.1 Bernoulli Trial to Bernoulli Distributions.

- The outcome of a **Bernoulli trial** is modeled by a **Bernoulli random variable**:
  - $Y \sim Bernoulli(p)$ called a **Bernoulli Distributions**.
  - **Here parameter:**
    - $p \rightarrow$ **probability of success and**
    - $1 - p \rightarrow$ **probability of failure.**
- **PMF of Bernoulli Distribution:**

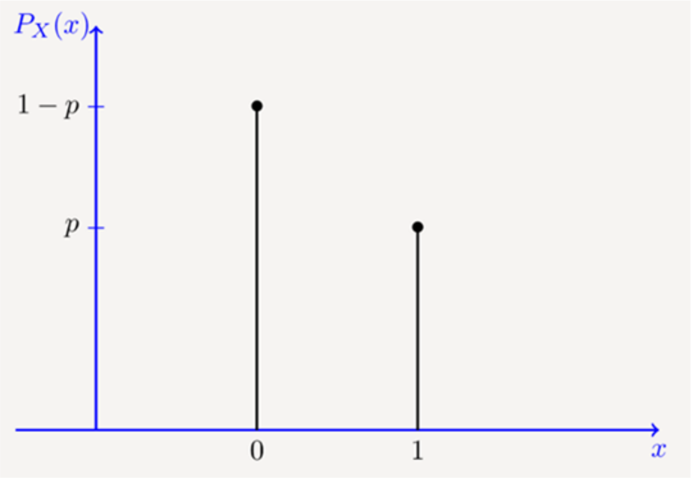
Definition Bernoulli Distribution and its PMF:
A Random Variable **Y** is said to be a Bernoulli random variable with **parameter p**, shown (written) as $\mathbf{Y \sim Bernoulli(p)}$, if it's PMF is given by:

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



5.3.2 Deriving the PMF of Bernoulli Distribution.

- **Objective:** To derive the formula for: $P(Y = y), where\ Y \sim Bernoulli(p), y \in \{0, 1\}$.
    - **Step – 1: Understand the Bernoulli Trail:**
        - Only two outcomes:
            - Success → $y = 1$, with probability p
            - Failure → $y = 0$, with probability $1 - p$
    - **Step – 2: Define the PMF for each case:**
        - **If y = 1**:
            - $P(Y = 1) = p$
        - **If y = 0:**
            - $P(Y = 0) = 1 - p$
    - **Step – 3: Combine both case in a Single Formula:**
        - We **want one formula** that works for both $y = 0$ **and** $y = 1$.
        - Use the fact that:
            - $p^y = p$ **when** $y = 1$, **and** $p^y = 1$ **when** $y = 0$
            - $(1 - p)^{1-y} = 1 - p$ **when** $y = 0$ **and** $(1 - p)^{1-y} = 1$ **when** $y = 1$
        - $P(Y = y) = p^y(1 - p)^{1-y}, y \in \{0, 1\}$■



- **This PMF satisfies:**
    - Non – negativity:
        - $0 \leq p^y(1 - p)^{1-y} \leq 1$■
    - Total probability:
        - $P(Y = 0) + P(Y = 1) = (1 - p) + p = 1$■

5.3.3 Statistics of a Bernoulli Distribution.

- **Given:**
  - Let $\boldsymbol{Y \sim Bernoulli(p)}$, where:
    - $\boldsymbol{P(Y = y) = p^y(1-p)^{1-y}, y \in \{0, 1\}}$
- **Expectations (Mean) is:**
  - $\mathbb{E}[Y] = \sum_{y\in\{0,1\}} y \cdot P(Y = y) = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = 0 \cdot (1-p) + 1 \cdot p = p\blacksquare$
  - The expected value of a Bernoulli variable is just the probability of success.
- **Variance:**
  - By definition:
    - $\boldsymbol{Var(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2}$
    - But since $\mathbf{Y \in \{0, 1\}}$, we have $\mathbf{Y^2 = Y \Rightarrow \mathbb{E}[Y^2] = \mathbb{E}[Y] = p}$, So:
      - $\boldsymbol{Var(Y) = p - p^2 = p(1-p)}\blacksquare$
  - The variance depends on both the probability of success and failure.
- **Observation:**
  - If $\boldsymbol{p = 0}$ or $\boldsymbol{p = 1}$ variance is 0 (no randomness)
  - Maximum variance is at $\boldsymbol{p = 0.5} \rightarrow$ **most uncertain outcome**.

5.4 Binomial Distribution.

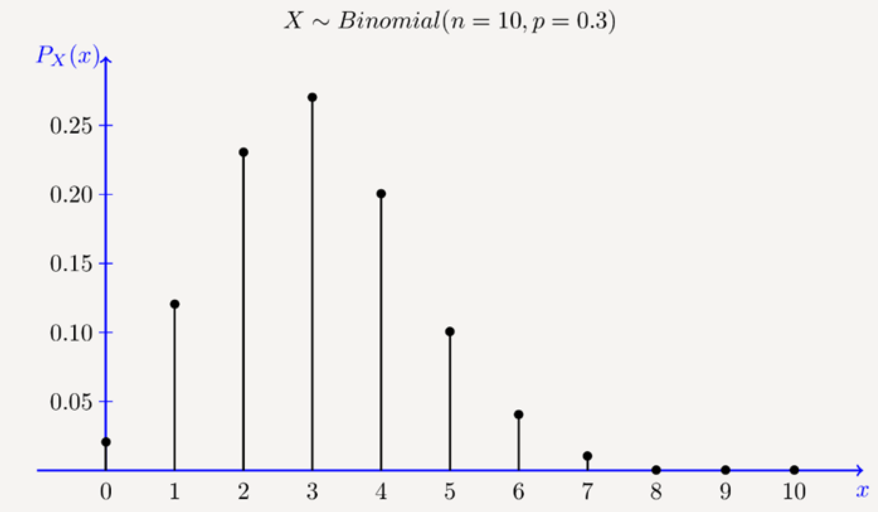
- Binomial Distribution is a discrete probability distribution that models
  - the number of successes in n independent Bernoulli trials, each with success probability p. i.e.
    - $X \sim Binomial(n, p)$
- Example:
  - Out of 100 customers sent an email, **how many click the link**?
    - Each Email is a Bernoulli Trial.
    - Parameters: $n = 100, p = 0.2$
    - $X: number\ of\ clicks$, then Random varable X is modeled by Binomial distribution with parameter $n\ and\ p$.
- Derivation of PMF:
  - Let $X \in \{0, 1, 2, \dots, n\}$ be the number of successes.
    - We want to compute:
      - $P(X = k) = Probability\ of\ getting\ exactly\ k\ successes.$
      - Choose k **successes from n trials**:
        - $\binom{n}{k}$
      - Success probability: p^k
      - Failure probability: $(1 - p)^{n-k}$
        - $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}: for\ k = 0, 1, \dots, n.$ ■

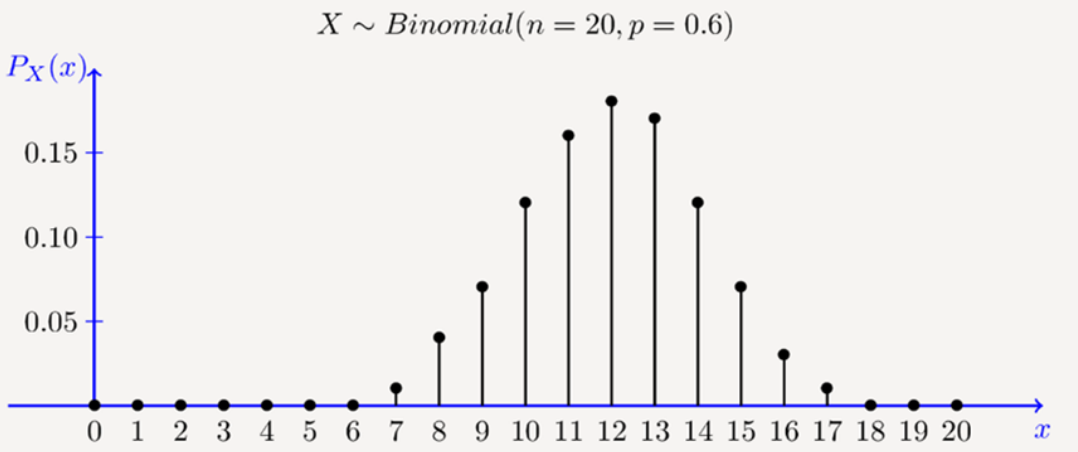
3.4.1 Binomial Distribution and PMF.

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

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

Where $0 < p < 1$.





3.4.2 Statistics of Binomial Distribution.

- **Let:**
  - $X \sim Binomial(n, p)$
    - Where X is the number of successes in **n independent Bernoulli(p) trials.**
    - So, we can write: $X = Y_1 + Y_2 + \cdots + Y_n$
    - *Here:*
      - $\Rightarrow \{Y \sim Ber(p)\}$
      - $\Rightarrow Y_i = 1$ **if the** i^{th} **trial is a successes, 0 otherwise.**
      - $\Rightarrow Y_1, Y_2, \ldots Y_n$ **are independent and identically distributed (i.i.d)**
- **Expectation of X:**
  - By linearity of expectations:
    - $\mathbb{E}[X] = \mathbb{E}[Y_1 + Y_2 + \cdots + Y_n] = \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \cdots \mathbb{E}[Y_n]$
    - Since: $\mathbb{E}[Y_i] = p$ **for all** i:
      - $\mathbb{E}[X] = np \blacksquare$

3.4.2 Statistics of Binomial Distribution.

- Variance of X:
  - $Var(X) = Var(Y_1 + Y_2 + \cdots + Y_n)$
  - If $Y_i's$ are **independent**, then:
    - $Var(X) = Var(Y_1) + Var(Y_2) + \cdots + Var(Y_n)$ $\{Property\ of\ Linearity\ of\ Variance.\}$
    - Since $Var(Y_i) = p(1 - p)$:
      - $Var(X) = n \cdot p(1 - p)$

3.4.3 Bernoulli vs. Binomial Distribution.

Feature	Bernoulli Distribution	Binomial Distribution
Type of Trial	Single Bernoulli trial (one event)	Repeated independent Bernoulli trials (n trials)
Random Variable	$\mathbf{Y} \in \{\mathbf{0}, \mathbf{1}\}$	$\mathbf{X} \in \{\mathbf{0}, \mathbf{1}, \ldots, \mathbf{n}\}$
Parameters	$p: sucess\ probability$	$n: number\ of\ trials$ $p: sucess\ probability$
PMF	$P(Y = y) = p^y(1-p)^{1-y}$	$P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$
Mean (Expected Value)	$\mathbb{E}[Y] = p$	$\mathbb{E}[X] = np$
Variance	$Var(Y) = p(1-p)$	$Var(X) = np(1-p)$
Use Case Example	Email Clicked or not.	Number of emails clicked of 1000 sent.

3.5 Geometric Distribution.

- It models the number of **trials until the first success** occurs
  - in a sequence of **independent Bernoulli trials**, each with success probability p.
  - Let the random variable:
    - $\boldsymbol{X \sim Geometric(p)}$
    - Where:
      - $\boldsymbol{X}$ → **number of trials until first success (including the success trial)**

3.5.1 Geometric Distribution: PMF.

- **Derivation of PMF:**
  - **Objective:**
    - **Find the probability that the first success occurs on the k^{th} trial in a sequence of independent Bernoulli trials.**
  - **Scenario:**
    - Each trial has only two outcomes: success (S) or failure (F).
      - Probability of success: p
      - Trials are independent.
    - **What does $P(X = k)$ mean?**
      - **The first success occurs on the k^{th} trial:**
        - Trials 1 through $k - 1$: must all be failures.
        - **Trial k: must be success.**
  - **Construct the Probability:**
    - Let's multiply the probabilities:
      - **$k - 1$ failures: $(1 - p)^{k-1}$**
      - **1 success on k^{th} trial: p**
    - So,
      - $P(X = k) = (1 - p)^{k-1}p, \mathrm{k} = 1, 2, 3, \dots$ ■
- **Interpretation: You wait $k - 1$ times (failures) before you get 1 success.**

- **Example:**
  - You are sending promotional messages to customers.
  - Each customer has a $p = 0.2$ chance of responding.
  - What is the probability the 5^{th} customer is the first to respond?
    - $\mathrm{P}(\mathrm{X} = 5) = (1 - 0.2)^4 \cdot 0.2 = (0.8)^4 \cdot 0.2 \approx 0.0819$ ■

3.5.2 Statistics of Geometric Distribution: Expectation.

- **Let:**
  - $X \sim Geometric(p)$ $(Number\ of\ trials\ until\ first\ success)$
    - $P(X = k) = (1 - p)^{k-1}p, k = 1, 2, 3, \ldots$
- **Expectation:**
  - $\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot P(X = k) = \sum_{k=1}^{\infty} k \cdot (1 - p)^{k-1}p$
  - Let's denote $q = 1 - p$, Then:
    - $\mathbb{E}[X] = p \sum_{k=1}^{\infty} kq^{k-1}$
      - Applying known identify aka power series identity: $\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}, for\ |q| < 1$
      - Since: $0 < p < 1$, **then** $|q| = |1 - p| < 1$ then:
        - $\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$
    - So, $\mathbb{E}[X] = \mathrm{p} \cdot \frac{1}{p^2} = \frac{1}{p}$ ■

- Interpretation:
  - If the probability of success in **each trial is p**, then on average, it takes $\frac{1}{p}$ **trials** to get **the first success**.
- Example:
  - If $p = 0.2$, then $\mathbb{E}[X] = 5$, expect the first success on the 5^{th} **trial.**
  - If $p = 0.5$, then $\mathbb{E}[X] = 2$.

3.5.3 Statistics of Geometric Distribution: Variance.

- **Let $X \sim Geometric(p)$:**
  - We want to derive: $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
    - We already know: $\mathbb{E}[X] = \frac{1}{p}$
  - Now we need to compute: $\mathbb{E}[X^2]$
  - **Step 1: Expand $\mathbb{E}[X^2]$:**
    - $\mathbb{E}[X^2] = \sum_{k=1}^{\infty} k^2 \cdot P(X = k) = \sum_{k=1}^{\infty} k^2 \cdot (1-p)^{k-1} p$
    - **Let's denote $q = 1 - p$**
      - **So:** $\mathbb{E}[X^2] = p \sum_{k=1}^{\infty} k^2 q^{k-1}$
  - **Step 2: Using Known Identity:**
    - $\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3}, for\ |q| < 1$
    - So: $\mathbb{E}[X^2] = p \cdot \frac{1+q}{(1-q)^3}$
      - **Recall:** $q = 1 - p$ **, so** $1 + q = 2 - p \Rightarrow 1 - q = p$
  - **Therefore:**
    - $\mathbb{E}[X^2] = p \cdot \frac{2-p}{p^3} = \frac{2-p}{p^2}$

3.5.3 Statistics of Geometric Distribution: Variance.

- **Step 3: Use Variance Formula:**
  - $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}$
- **Final Result:**
  - $Var(X) = \frac{1-p}{p^2}$ ■
- **Interpretation:**
  - Variance is the measure of uncertainty or volatility in outcomes.
    - Usually interpreted as: " How spread out the number of trials is before first success."
    - If $\mathbf{p = 0.2 \Rightarrow Var(X) = \frac{0.8}{0.04} = 20}$
    - If $\mathbf{p = 0.5 \Rightarrow Var(X) = \frac{0.5}{0.25} = 2}$
  - **So: Variance quantifies the reliability or stability of the mean.**
- **Business Analogy:**
  - Suppose the mean time until a customer buys a product is 5 days.
    - Low Variance: Most customers buy between 4 – 6 days. You can rely on the mean.
    - High Variance: Some buy in 1 day, some take 10+. The mean is still 5, but it is not reliable for planning.

- Cautions:
  - It does not tell us how well we have estimated the mean, that is done by standard error,
  - but in the context of single random variable,
    - the variance does indicate how trustworthy or representative the mean is:
- **"The mean tells you what to expect. The variance tells you how often you should expect surprises."**

3.6 Poisson Distribution

- The Poisson distribution tells you how likely it is to observe **k events** in a **fixed window**, when those events are **random but occur at a steady average rate**.
- Formal Definition:
- A **Poisson distribution** models the probability of observing a given number of events in a fixed interval of time or space, **under the following conditions**:
- Let $X \sim Poisson(\lambda)$, where:
- $P(X = k) = \frac{e^{-\lambda}.\lambda^k}{k!}, k = 0, 1, 2, \ldots$
- Here:
- $\lambda > 0 \rightarrow$ the average rate (mean) of events per interval.
- $k \rightarrow$ number of occurrences in the interval.
- $e \rightarrow$ Euler's Number, ≈ 2.718
- Key Properties:
- Expectation $\rightarrow \lambda$
- Variance $\rightarrow \lambda$

3.6.1 Using Poisson.

- **Conditions:**
  - The Poisson model applies when:
    - Events occur independently
    - Events occur one at a time
    - The **average rate λ** is constant
    - The probability of more than one event in an infinitesimally small interval is negligible.
- **Examples in Business Analytics:**

Scenario	X = Number of events …	Interval	λ = Avg. Rate
Website traffic	Page Visits	Per minute	30
Sales	Purchases	Per day	15
Support requests	Tickets filed	Per hour	5
Product defects	Defects in batch	Per 100 units	2

3.6.2 You Task:

1. **Derive** the PMF, expectation, and variance of the Poisson distribution step-by-step, starting from first principles or known limits (e.g., from Binomial distribution).
2. **Select** a real-world business scenario (e.g., number of customer support calls per hour, order arrivals per minute, product defect counts per batch).
3. **Interpret**:
   - What does the **mean (λ)** represent in your context?
   - What does the **variance** tell you about the reliability of that mean?
   - How does the **Poisson PMF** help you estimate probabilities in that scenario?

Thank You