5CS037 - Concepts and Technologies of Artificial Intelligence
Herald College
University of Wolverhampton

Tutorial - 03
General Overview about Discrete Probability Theory for Machine Learning.

Siman Giri {Module Leader - 5CS037}

November - 25 - 2025

Instructions:

- Complete all the problems.

A statistician drowned crossing a river that was, on average, three feet deep.{The Big Short}.

Contents

1 Basic Concepts.

Probability Theory is the mathematical study of uncertainty.The development of learning algorithms frequently depends on probabilistic assumptions about the data. These notes aim to provide an overview of foundational concepts in probability theory, serving as essential background material for the course.
Probability provides a consistent framework for the quantification and manipulation of uncertainty.In order to model the behavior of a process based on observed or empirical outcomes and make inferences about future events, we will discuss a formal mathematical interpretation of probability theory, grounded in the foundational work of Andrey Kolmogorov.

1.1 Probability Space:

When discussing probability, we typically refer to the likelihood of an uncertain event occurring. For instance, we might talk about the probability of rain next Tuesday. To formally address probability theory, it is essential to first define the set of possible events to which we aim to assign probabilities.Let's Start with Elements of Probabilistic Space.

1.1.1 Elements of Probabilistic Space:

A probability model is a mathematical representation of a random phenomenon experiment defined by its sample space, outcomes within the sample space and probability measure, that defines how to assign probability to each outcome governed by axioms of probability. Thus any probabilistic space has following elements:

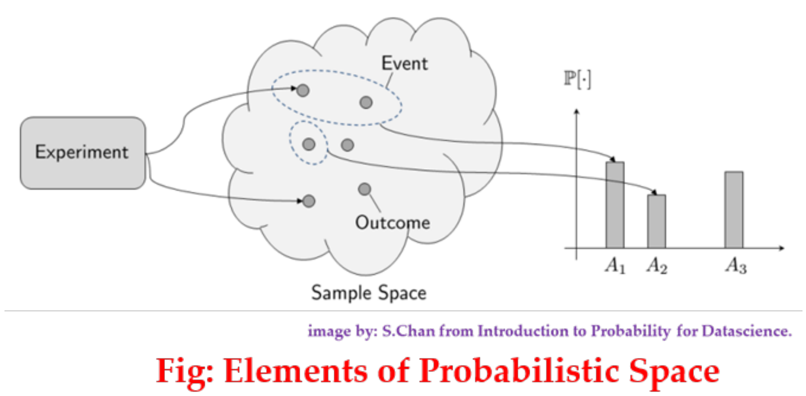


Figure 1: Elements of Probabilistic Space - {Slide - 13}.

1. A {Chance} Experiment:

A chance experiment, refers to a procedure or action that results in one of several possible outcomes, where the specific outcome is uncertain and cannot be predicted with certainty in advance.
Key features of a chance experiment are:

- **Reproducibility:** The experiment can be conducted multiple times under identical conditions.
- **Uncertainty:** The exact outcome of any single trial cannot be determined in advance.
- **Sample Space:** The set of all possible outcomes, known as the sample space, is well-defined.

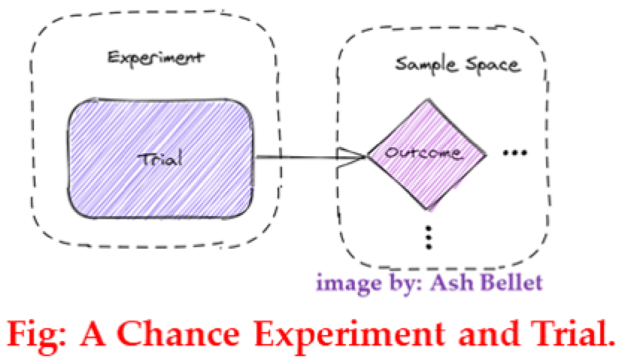


Figure 2: Chance Experiment and Trial - {Slide -14}.

2. A Sample Space:

Sample Space denoted as Ω {read - omega} or **"S"** is a set of all the outcomes of a random experiment. Here, each outcome $\omega \in \Omega$ represents the final state of the real world at the end of the experiment. Sample Space can be either discrete or continuous depending on the value outcomes ω can take.

- Discrete Sample Spaces are typically finite or countably infinite sets of distinct outcomes used to model events such as the results of rolling a die, drawing cards, or counting occurrences.
- Continuous sample spaces are typically intervals or subsets of $\mathbb{R}$ or $\mathbb{R}^n$ used to model phenomena such as time, position, temperature, or other quantities that can take on any value within a range.

Examples of sample space:

1. Coin Flip: S = Heads, Tails(Discrete)
2. Flipping two coins: S = (H,H),(H,T),(T,H),(T,T)
3. YouTube hours in a day: S $= \{x | x \in \mathbb{R}, 0 \leq x \leq 24\}$ (Continuous)

Cautions - How to Choose an appropriate Sample Space:

- The sample space should consist of distinct, mutually exclusive elements, ensuring a unique outcome for each experiment.
- It must be collectively exhaustive, covering all possible outcomes, and detailed enough to distinguish relevant outcomes while excluding unnecessary information.

3. Events:

A subset of the sample space, that is a collection of possible outcomes is called an event {Usually denoted with F}. i.e. $F \subseteq S$.

Events can be

- Simple Discrete Event: A simple event is an event consisting of exactly one outcome. For example:
  - For Single coin flip $\texttt{S} := \texttt{H,T}$

$$\texttt{Event - } F_1 = \texttt{Getting Head} = \texttt{H} \rightarrow \texttt{This is a simple Event.}$$

- Continuous Event: An event defined within Continuous Sample Space. For example:
  - Youtube Hours in a day $\texttt{S} := x|x \in \mathbb{R}, 0 \leq x \leq 24$

$$\texttt{Event - } F_{wasted\_day} = \texttt{Youtube} \geq 5\texttt{hours} = x|x \in \mathbb{R}, 5 \leq x \leq 24 \rightarrow \texttt{Continuous Event.}$$

- Event can also be set of functions: An event defined as a function of how outcomes appear in a process. In general, when we have countably infinite possible outcomes. For example:
  - Toss a fair coin infinite number of times $\texttt{S} := f : \mathbb{N} \rightarrow \texttt{H,T}$

$$\texttt{Event - } F = \texttt{First two tosses are tails} = f \in S|f(1) = T\text{and}F(2) = T$$

Events in a Event Space:

An event space is the collection of all possible events associated with a given sample space.{Denoted with $\mathfrak{F}$ and read {sigma algebra}}. Events are subsets of the sample space, and the event space includes all such subsets that are relevant to the probability experiment.

To Summarize:

- Sample SpaceΩ or "S": The set of all possible outcomes of a random experiment.
- Event $\texttt{F}$: A subset of the sample space $\texttt{F} \subseteq \texttt{S}$ that corresponds to a particular outcome or group of outcomes.
- Event Space $\mathfrak{F}$: A collection of subsets of the sample space, including:
  - the empty set is included - i.e. $\phi \in \mathfrak{F}$.
  - the entire sample space is included - i.e. $S \in \mathfrak{F}$.
  - the complement of an event is also in the event space - i.e. If $A \in \mathfrak{F}$ then $A^c \in \mathfrak{F}$.
  - the event space is closed under countable unions - i.e. If $A_1, A_2, \ldots, \in \mathfrak{F}$ then $A_i \in \mathfrak{F}$
- Example - Event Space - Consider Flipping a coin: The event space is given as:

$$\mathfrak{F} = \{\phi, \texttt{H}, \texttt{T}, \Omega\}$$

Events that are Complex:

- **Complex Events:** A complex event refers to an event composed of multiple simple events combined using logical operators such as

$$\texttt{union}(\cup), \texttt{intersection}(\cap), \texttt{or complement}(F')$$

It may involve more than one event but does not specifically require their simultaneous occurrence.

  - Example - Rolling a fair die - $S = \{1, 2, 3, 4, 5, 6\}$:
    * Event A: Rolling an even number:
$$F_A = \{2, 4, 6\}$$
    * Event B: Rolling a number greater that 4:
$$F_B = \{5, 6\}$$
    * **Complex Event:** Rolling an Even number or a number greater than 4:
$$(F_A \cup F_B) = \{2, 4, 5, 6\}$$

- **Joint Events:** A special type of complex events refers to the simultaneous occurrences of two or more events and are specially concerned with outcomes common to both events, Thus represented with symbol intersection($\cap$).
  - For Example - Rolling a die:
    * Event A: Rolling an even number:
$$F_A = \{2, 4, 6\}$$
    * Event B: Flipping heads on a single coin:
$$F_B = \{\texttt{H}\}$$
    * **Joint Event:** Rolling an Even number and a Head:
$$(F_A \cap F_B) = \{(2, H), (4, H), (6, H)\}$$

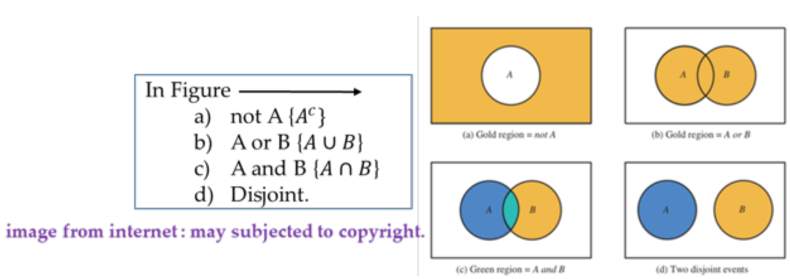


Figure 3: Using a Set Notations to Represent a Complex Events - {Slide - 18}.

4. Probability Measure:

A probability measure is a function that maps an events in an event space to a real number [0,1] satisfying `the axioms of probability`:

$$\mathbb{P} : \mathfrak{F} \to [0, 1] \tag{1}$$

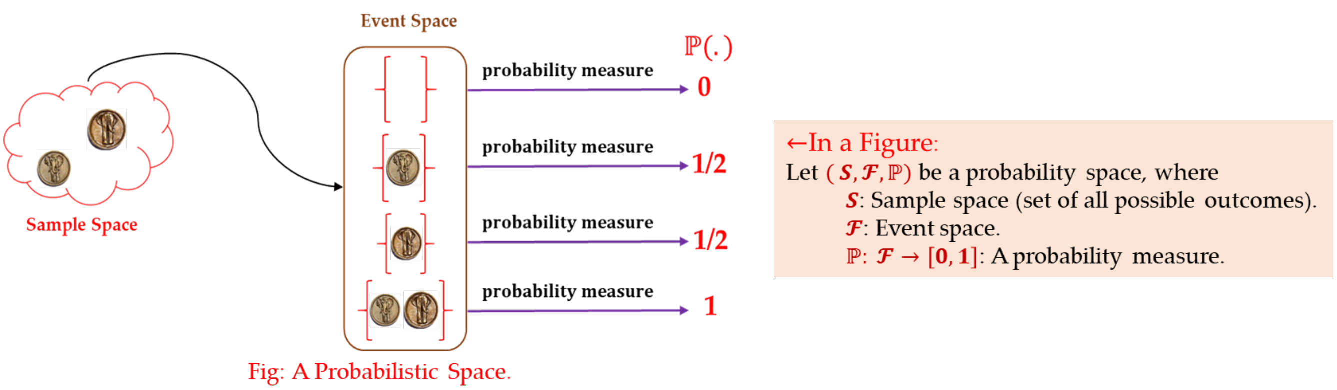


Figure 4: Ingredients of a Probabilistic Space - {slide - 19}.

1.2 Assigning a Probability:

1.2.1 Classical Approach:

In the context of the classical approach to probability, probability is defined based on the ratio of favorable outcomes to the total number of equally likely outcomes in the sample space.
The classical approach assumes that all outcomes in the sample space are equally likely. Given a finite sample space

$$\mathtt{S} = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

where n is the total number of outcomes, then the probability of any events is given by:

$$\mathbb{P}(\mathtt{F}) = \frac{1}{n}$$

Some key Points for the Classical Approach:

- Equally likely outcomes: All outcomes in the sample space must have the same probability of occurring.
- Finite sample space: The sample space consists of a finite number of possible outcomes.
- Example: If you roll a fair six-sided die, the sample space is

  $$S = \{1, 2, 3, 4, 5, 6\}$$

  Then probability of 3 is:

  $$\mathbb{P}(3) = \frac{1}{3}$$

- **Important Limitations:** The classical approach is limited to situations where all outcomes are equally likely.For cases with biased outcomes for instance

$$S = \{H, H, T\}$$

the classical approach doesn't directly apply, and a different probability model (such as the frequentist or axiomatic approach) is needed.

1.2.2 Frequentist Approach:

The frequentist approach to probability defines the probability of an event as the long-run relative frequency with which the event occurs in repeated independent trials of an experiment. In this framework, probability is based on empirical observations and the law of large numbers.

Law of Large Numbers:

The Law of Large Numbers (LLN) is a fundamental theorem in probability theory that states:

`\As the number of trials or observations increases, the average of the observed outcomes will converge to the expected value (the true probability) of the event."`

Example:

- If you repeatedly flip a fair coin:
  - In a small number of flips, the proportion of heads may vary significantly

$$\texttt{e.g.}, 3\texttt{heads out of}4\texttt{flips} = 0.75.$$

    However, as the number of flips increases, the proportion of heads will get closer to 0.50, the expected probability.

The LLN ensures that observed probabilities stabilize with a large number of trials.

Formal Definition: In the frequentist approach, if an experiment is repeated n times and an event occurs k times, then the probability of an event is estimated as:

$$\texttt{For any event}(E \in \mathfrak{F})$$

$$P(E) \approx \frac{\texttt{\# count number of favorable outcomes} \in E}{\texttt{\# Count number of outcomes in the sample space} \in \Omega} \tag{2}$$

Key Features of Frequentist Approach:

- Empirical Probability: The probability of an event is determined by conducting the experiment repeatedly and observing the proportion of times the event occurs.
- Long-Run Frequency: As the number of trials increases, the observed frequency of the event converges to a fixed value, which is interpreted as the probability of the event.
- Objective Interpretation: Probability is considered a property of the system or experiment, independent of personal belief or prior information.

- Example:Suppose you flip a fair coin 1000 times, and you observe that "Heads" occurs 510 times. The frequentist probability of getting "Heads" would be estimated as:

$$P(Heads) \approx \frac{510}{1000} = 0.51$$

  As you conduct more trials (in this case, more coin flips), the observed probability will converge to the true theoretical probability of 0.5.

Distinction from Classical Approach:

- The classical approach assigns probabilities based on equally likely outcomes and a known sample space.
- The frequentist approach estimates probabilities through repeated trials and empirical data, without assuming equally likely outcomes.

In summary, the frequentist approach treats probability as a long-run frequency of an event occurring in repeated trials, emphasizing empirical data and observations rather than theoretical assumptions about the nature of the experiment.

1.3 Background - Combinatorics.

From the lecture and this note, it is evident that a fundamental aspect of assigning probabilities in a discrete space (commonly referred to as a discrete probability model) lies in accurately enumerating the possible events. In this section, we will revisit the principles of combinatorics to refamiliarize ourselves with the concepts of counting.

1.3.1 Addition and Multiplication Principles in Counting:

In probability and combinatorics, addition and multiplication principles provide foundational rules for counting the number of possible outcomes of events or experiments. These principles are crucial for understanding more advanced topics in probability.

1. Addition Principle:

The addition principle applies when the outcomes of an experiment can be drawn from two mutually exclusive sets (sets with no overlap). If a result belongs to either Set A or Set B, and no outcome is common between the two sets, the total number of possible outcomes is the sum of the number of outcomes in each set.

- Mathematical Representation:

$$\mathtt{A} \cup \mathtt{B} = \mathtt{A} + \mathtt{B}, \quad \mathtt{if}\ \mathtt{A} \cup \mathtt{B} = \phi \tag{1}$$

Example: A restaurant offers:

- 5 main course wit chicken:
- 6 main course with mutton, and
- 12 vegetarian main courses.

Solution: Since none of these categories overlap (i.e. no dish belongs to more than one category), the total number of main courses is:

$$5 + 6 + 12 = 33 \quad \square$$

Thus, there are 23 distinct main course options in total.

2. Multiplication Principle:

The multiplication principle applies when an experiment or process can be divided into independent parts, where the number of outcomes of one part does not affect the number of outcomes of another. The total number of outcomes is the product of the outcomes of each part.

- Mathematical Representation:

$$\texttt{Total outcomes} = \mathbf{m} \times \mathbf{n} \tag{2}$$

- where m and n are the numbers of outcomes for the two parts of the experiment.

Example: Consider flipping a fair coin followed by rolling a fair six - sided die:

- The coin has 2 possible outcome $\rightarrow \{Heads, Tails\}$.
- The die has 6 possible outcome $\rightarrow \{1, 2, 3, 4, 5, 6\}$.

Solution: The total number of possible results for flipping the coin and rolling the die is:

$$2 \times 6 = 12$$

Thus, there are 12 distinct outcomes, such as $(\texttt{Heads}, 1), (\texttt{Tails}, 4), \ldots$

Practice Exercise - Addition and Multiplication Principle:

Exercise - 1:

A bookstore has:

- 5 types of fiction books,
- 6 types of non - fiction books.

For bookmarks, it offers:

- 3 plain designs,
- 4 patterned designs.

Question:

1. How many total books are available?
   - Solution: Using the addition principle for books:

$$5 + 6 = 11\texttt{books} \quad \square$$

2. If a customer buys one book and one bookmark, how many combinations are possible?
   - Solution: Using the multiplication principle for books and bookmarks:

$$11 \times (3 + 4) = 11 \times 7 = 77\texttt{combinations.} \quad \square$$

1.3.2 Permutations and Combinations:

Counting problems in probability often rely on the basic counting principles outlined earlier: the addition and multiplication principles. However, as we encounter more complex scenarios, such as determining arrangements or selections from a set, these basic principles may no longer suffice. For such problems, we use higher-level counting abstractions, namely permutations and combinations. These concepts provide a systematic framework for solving a wide range of counting problems efficiently.

1. Permutations:

Definition: A permutation is an arrangement of items in a specific order. The order in which terms are arranged matters. For example:

- Arranging the letters $\{A, B, C\}$ produces:

$$(\mathtt{ABC}), (\mathtt{ACB}), (\mathtt{BAC}), (\mathtt{BCA}), (\mathtt{CAB}), (\mathtt{CBA})$$

- If you have n items and want to arrange all of them, the number of possible permutations is $n!$.

Mathematical Formulation of Permutations:

For a subset of r items selected from a set of n items, the number of permutations is given by:

$$P(n, r) = \frac{n!}{(n-r)!} \tag{3}$$

where$P(n, r)$ represents the number of permutations of r items chosen from n.
Example:

- How many ways can 3 students (Alice, Bob, Charlie) be seated in 3 chairs?
- Solution:

$$P(3, 3) = \frac{3!}{(3-3)!} = \frac{3 \times 2 \times 1}{0!} = \frac{6}{1} = 6 \quad \square$$

Use Case: Permutations are ideal when `order matters`, such as in seating arrangements, race rankings, or arranging books on a shelf.

2. Combinations:

Definition: A combination is a selection of items where the order does not matter. For example:

- Choosing 2 letters from {A,B,C} produces {AB,AC,BC}. Here, {(AB) and (BA)} are considered the same selection.

Mathematical Formulations of Combinations:

The formula for the number of combinations is:

$$\mathtt{C}(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{4}$$

where C(n, r) represents the number of combinations of r items chosen from n.
Example:

- How many ways can you choose 2 students from a group of 3 i.e. {Alice, Bob, Charlie}?

- Solution:

$$\mathtt{C}(3,2) = \binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} = 3 \quad \square$$

The combinations are - {AB, AC, BC}

Use Case: Combinations are ideal when `order does not matter,` such as selecting a team from group, picking lottery numbers, or forming subsets.

Counting is essential in probability because it provides the foundation for calculating probabilities, particularly for discrete sample spaces. Counting helps distinguish between disjoint (mutually exclusive) and joint (overlapping) events.

1.4 Exercise - Basic Concepts:

Problem - 1 - Experiment, Sample Space, Events and Counting:

Solve all the question:

1. Write a sample space for the Given experiment:
   (a) A Die is rolled.
   (b) A die is rolled and coin is tossed
   (c) A penny and nickel are tossed.

Sample Solution b:

The outcomes include the results of rolling the die $(1, 2, 3, 4, 5, 6)$ combined with the results of tossing the coin (H, T).Using ordered pairs (x, y), where x is the die roll and y is the coin toss, the sample space is:

$$S = \{(1, H), (1, T), (2, H), (2, T), (3, H), (3, T), (4, H), (4, T), (5, H), (5, T), (6, H), (6, T)\}$$

Functional Representations:

$$S = \{(x, y) : x \in \{1, 2, 3, 4, 5, 6\}, y \in \{H, T\}\}$$

2. A Ultima has five identical ear-buds for shipment, unknown to person-in charge for shipping two of the five ear-buds are defective. A particular order calls for two of the ear-buds and is filled randomly selecting two of the five available:
   (a) List the sample space for this experiment.
   (b) Let **A** denote the event that the order is filled with two non-defective ear-buds.List the sample points in A.
3. Every person's blood type is A,B,AB,or o.In addition, each individual either has the Rhesus(RH) factor +ve or -ve. A medical technician records a person's blood and Rh factor. List the sample space for this experiment.
4. Assume to win a bet at MotoGp, you need to specify the riders that finish in the top three spots in the

exact order in which they finish. If eight riders enter the race, how many different ways can they finish in the top three spots?

5. There are 18 faculty members in the department of Mathematics and Statistics. Four people are to be in the executive committee. Determine how many different ways this committee can be created.
6. The director of a research laboratory needs to fill a number of research positions; two in biology and three in physics. There are seven applicants for the biology positions and 9 for the physicist positions. How many ways are there for the director to select these people?

Problem 2 - Assign a probability - Frequentist Approach.

Solve all the Problems:

Hint - To solve the Problems:

- Problem formalization:
  1. Describe a sample space Ω.
  2. Describe an event $\mathbf{A} \subset \mathbf{\Omega}$
  3. Count the number of elements in the event $E \in \mathfrak{F}$ (use combinatorics)
  4. Compute:

$$P(E) \approx \frac{\texttt{\# count number of favorable outcomes} \in E}{\texttt{\# Count number of outcomes in the sample space} \in \Omega} \tag{2}$$

1. n cards are drawn from a standard deck of 52. What is the probability that:
   (a) $n = 8$, all card are queens or kings?
   (b) $n = 6$, all cards contain only two spades and 1 heart?
   (c) $n = 5$, all cards contain all 4 suits(spade, heart, diamond, clubs)?
   (d) $n = 4$, all cards contain only 2 spades and 1 king?

   Please don't forget to describe the universe, space and the events!

Sample Solution:

(a) all cards are queens or kings?
- sample space: The total number of ways to draw 8 cards from a standard deck of 52 is:

$$\binom{52}{8} = \frac{n!}{K! \times (n-k)!} = \frac{52!}{8! \times (52-8)!}$$

- Favourable Events $[E_1]$: There are 4 queens and 4 kings in the deck. We need to choose 8 cards from these 8 specific ranks:

$$\binom{8}{8} = \frac{8!}{8!(8-8)!} = 1$$

$$P(E_1) = \frac{E_1}{\text{Sample space}} = \frac{1}{752,538,150} \quad \square$$

(b) n $= 6$, all cards contain only two spades and one Heart:
- There are 13 Spades; choose 2 - $\binom{13}{2}$.
- There are 13 Hearts, choose 1 - $\binom{13}{1}$.
- Total ways to choose the remaining 3 cards from the remaining 39 cards (non-spades and non-Hearts) - $\binom{39}{3}$.
- Total ways to choose 6 card from 52: $\binom{52}{6}$
- The Probability is:

$$P(E) = \frac{\binom{13}{2} \cdot \binom{13}{1} \cdot \binom{39}{3}}{\binom{52}{6}} \quad \square$$

(c) n $= 5$, all cards contain all 4 suits (Spade, Heart, Diamond, Club).To satisfy the condition:
- 4 cards must come from different suits (1 card from each suit).

$$\texttt{Choose 1 card per suit -}13 \times 13 \times 13 \times 13 = 13^4 \{\texttt{Multiplicative Rule}\}$$

- The 5th card can come from any of the 4 suits (as repetition is allowed). There are $52 - 4 = 48$ cards left to choose from.
- Total ways to choose 5 cards from 52: $\binom{52}{5}$
- The probability is:

$$P(E) \approx \frac{13^4 \cdot 48}{\binom{52}{5}} \quad \square$$

(d) n $= 4$, all cards contain only 2 Spades and 1 King
- Choose 2 spades from 13 spades: $\binom{13}{2}$
- Choose 1 king form 4 kings: $\binom{4}{1}$
- The remaining card can be any card from the remaining 36 cards (non spades and non kings): $\binom{36}{1}$
- Total way to choose 4 cards from 52 is:$\binom{52}{4}$
- The probability is:

$$P(E) \approx \frac{\binom{13}{2} \cdot \binom{4}{1} \cdot \binom{36}{1}}{\binom{52}{4}} \quad \square$$

2. A box contains 10 white and 10 black marbles. Following experiment were performed:
   - randomly drawing out with replacement,two marbles in succession and noting the color each time.

   (To draw "with replacement" means that the first marble is put back before the second marble is drawn.)

   (a) Construct a sample space for the experiment.

   (b) List the outcomes that comprise each of the following events.
   - At least one marble of each color is drawn.
   - No white marble is drawn.

3. For the experiment of rolling a single six-sided die once, define events
   - **T: the number rolled is three.**
   - **G: the number rolled is four or greater**

   Solve following:

   (a) List the outcomes that comprise T and G.

   (b) List the outcomes that comprise $T \cup G$, $T \cap G$, T^c, and $(T \cup G)^c$.

   (c) Assuming all outcomes are equally likely, find $P(T \cup G)$,$P(T \cap G)$,and $P(T^c)$.

   (d) Determine whether or not T and G are mutually exclusive. Explain why or why not.

2 Axioms of Probability.

The axiomatic approach defines probability abstractly using a set of axioms, allowing for a more general and flexible interpretation of probability that can be applied to both theoretical(classical) and empirical (Frequentist) problems.
This approach provides a rigorous mathematical foundation for probability theory, making it applicable to a wide variety of situations, including those with complex or infinite sample spaces.

2.1 Axioms of Probability:

The axioms of probability are the foundational rules that any probability measure must satisfy to ensure consistency and validity in probability theory. These axioms, introduced by Andrey Kolmogorov (that is why also known as Kolmogorov Axioms of Probability), define how probabilities are assigned to events in a mathematically rigorous way.

- For a probability space defined by $\Omega, \mathfrak{F}, \mathbf{P}$ Where:
  - Ω or "S": Sample space (Set of all possible outcomes)
  - $\mathfrak{F}$: Event Space.
  - $\mathbf{P}$: $\mathbf{P} : \mathfrak{F} \to [0, 1]$ A Probability Measure.

  Axioms are Defined as Follows:

Axioms of Probability:

1. Non - Negativity: Probability can not be negative.
$$P(A) \geq 0 \ \forall A \in \mathfrak{F}$$
2. Normalization: Sum of probability of all possible outcomes in a sample space is 1.
$$P(\omega) = 1.$$
3. Additivity (Countable additivity) - For any sequence of mutually exclusive events defined as:
$$A_1, A_2, \ldots\ldots \texttt{Such that } A_i \cap A_j = \phi \ for \ i \neq j.$$
The probability is given as:
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$
For example: For two mutually exclusive events A and B:
$$P(A \cup B) = P(A) + P(B)$$

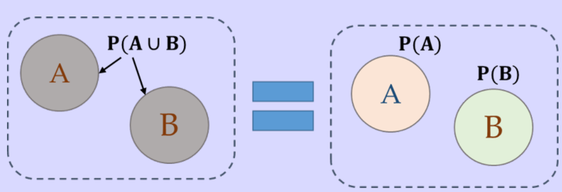


Fig: Understanding Additivity.

1. Corollaries derived from the axioms:

In mathematics and logic, a corollary is a statement that follows directly and easily from a previously proven theorem or proposition. It is considered a consequence of the main result and often requires little or no additional proof. Corollaries are typically used to highlight or formalize implications of a theorem that are significant or useful in a specific context.Following are the corollaries derived from axiom of probability:

Corollaries:

1. Corollary 1 - Probability of the Empty Set: The probability of the empty set ϕ i.e. event with no outcomes is zero i.e.
$$\mathbf{P}(\phi) = 0.$$
2. Corollary 2 - Probability of Complements: The probability of complements of an event A denoted A^c is:
$$\mathbf{P}(A^c) = 1 - A^c.$$
3. Corollary 3 - Probability of subsets i.e. If $A \subseteq B$ then:
$$\mathbf{P}(\mathbf{A}) \leq \mathbf{P}(\mathbf{B}).$$
4. Corollary 4 - Bounds of Probability - For any event A:
$$\mathbf{0} \leq \mathbf{P}(\mathbf{A}) \leq \mathbf{1}$$
For any event A and B:
$$\mathbf{P}(\mathbf{A} \cup \mathbf{B}) \leq \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}).$$
5. Corollary 5 - Inclusion - Exclusion or Probability of Union of Events: For any events A and B:
$$\mathbf{A} \cup \mathbf{B} = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{A} \cap \mathbf{B})$$

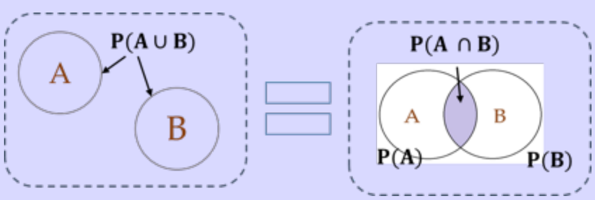


Figure: Inclusion - Exclusion (Slide 21)

2. Probability of Mutually Exclusive Events:

Definition - Mutually Exclusive Events: Mutually exclusive events are events that cannot occur simultaneously. In other words, the occurrence of one event precludes the occurrence of the other(s). For example:

- In a single coin toss, the events "Heads" and "Tails" are mutually exclusive.
- In rolling a die, the events "rolling a 3" and "rolling a 4" are mutually exclusive.

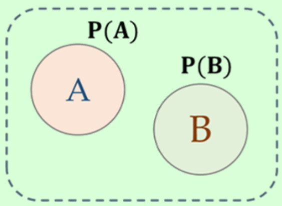
Mathematical Representation of Mutually Exclusive Events:

Two events A and B are mutually exclusive if:

$$\mathbf{A \cap B = Œ}$$

Where $A \cap B$ is the intersection of A and B, representing outcomes common to both. Since $A \cap B = \phi$,Thus the probability of both events occuring togetther is 0 i.e.

$$\mathbf{P(A \cap B) = 0}$$



← Fig: Mutually Exclusive Events.

Figure from Slide 26

1. Assigning Probability to Mutually Exclusive Events:

When events are mutually exclusive, the probability of their union is:

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

{Based on additivity axiom of probability also known as Union or Additive law of Probability.} If we are not sure about the exclusivity of two events A and B, based on inclusion - exclusion we write:

$$\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

2.2 To Sum Up - Assigning Probability to Discrete Events:

If the sample space consists of a finite number of possible outcomes, then the probability law is specified by the probabilities of the events that consist of a single element. In particular, the probability of any event $\{\mathbf{s1, s2, ..., sn}\}$ is the sum of the probabilities of its elements:

$$\mathbf{P(s1, s2, \ldots sn) = P(\{S1\}) + \cdots + P(\{Sn\})} \quad \square$$

Thus we can say that:
To evaluate the probability of discrete events, the calculation depends on counting, which in turn depends on the type of events: whether they are disjoint or joint.

- Disjoint events are always mutually exclusive, meaning they cannot occur simultaneously.

$$\mathbf{P(A \cup B) = P(A) + P(B)} \quad \square$$

- Joint events can either be independent (unrelated) or dependent (related or conditioned), depending on whether the occurrence of one event affects the probability of the other.

$$\mathbf{P(A \cap B)} = \frac{\texttt{\# count the items of A and B}}{\texttt{\# Total Outcomes in the sample space}}$$

Joint events introduces the variable of relationship between the events which has to be identified and studied to correctly assign the probabilty follow section will discuss on the same.

2.3 Exercises - Axioms of Probability.

Problem - 1:

Fill in the blanks:

1. Toss one fair coin (the coin has two sides, H and T). The outcomes are _____. Count the outcomes. There are _____ outcomes.
2. Toss one fair, six-sided die (the die has 1, 2, 3, 4, 5, or 6 dots on a side). The outcomes are $\{\_\_\_, \_\_\_, \_\_\_, \_\_\_, \_\_\_, \_\_\_\}$. Count the outcomes. There are _____ outcomes.
3. Multiply the two numbers of outcomes. The answer is __________.
4. Event A = heads (H) on the coin followed by an even number (2, 4, 6) on the die.
$$A = \{\_\_\_\_\_\_\_\_\_\_\}.$$
Find $\mathbf{P(A)}$.
5. Event B = heads on the coin followed by a three on the die.
$$B = \{\_\_\_\_\_\_\_\_\_\_\}$$
Find $\mathbf{P(B)}$.

Problem - 2:

We have a committee of n = 10 people and we want to choose a chairperson, a vice-chairperson and a treasurer. Suppose that 6 of the members of the committee are male and 4 of the members are female. What is the probability that the three executives selected are all female?

Problem - 3:

Pick an integer in [1, 1000] at random. Compute the probability that it is divisible neither by 12 nor by 15.

Sample Solution:

1. Step - 1 - Understand the Problem:
   - Sample Space: $S = \{1, 2, 3, \dots, 1000\} \rightarrow |S| = 1000.$
   - Objective: Find the probability that a number is not divisible by 12 or 15.
   - Approach: Use the complement rule:
   $$\mathbf{P(\texttt{Not divisible by 12 or 15}) = 1 - P(A_{12} \cup A_{15})}$$
   Here:
     - A_{12} is the set of numbers divisible by 12
     - A_{15} is the set of numbers divisible by 15
     - $A_{12} \cup A_{15}$ is the set of numbers divisible by either 12 or 15.

2. Step - 2 - Calculate $\mathbf{P(A_{12} \cup A_{15})}$ using Inclusion - Exclusion:
   Using the formula for probabilities of unions:

   $$\mathbf{P(A_{12} \cup A_{15}) = P(A_{12}) + P(A_{15}) - P(A_{12} \cap A_{15})}$$

   Let's Compute all the element:
   (a) Calculate $P(A_{12})$:The numbers divisible by 12 in [1, 1000] are:

   $$|A_{12}| = \lfloor \frac{1000}{12} \rfloor = 83$$

   Thus:

   $$\mathbf{P(A_{12}) = \frac{|A_{12}|}{|S|} = \frac{83}{1000}} \quad \square$$

   (b) Calculate $P(A_{15})$: The numbers divisible by 15 in [1, 1000] are:

   $$|A_{15}| = \lfloor \frac{1000}{15} \rfloor = 66$$

   Thus:

   $$\mathbf{P(A_{15}) = \frac{|A_{15}|}{|S|} = \frac{66}{1000}} \quad \square$$

   (c) Calculate $P(A_{12} \cap A_{15}$: The numbers divisible by both 12 and 15 are divisible by their least common multiple i.e.

   $$\texttt{lcm}(12, 15) = 60$$

   The numbers divisible by 60 in [1, 1000] are:

   $$|A_{60}| = \lfloor \frac{1000}{60} \rfloor = 16$$

   Thus:

   $$\mathbf{P(A_{11} \cap A_{15}) = \frac{|A_{60}|}{|S|} = \frac{16}{1000}} \quad \square$$

   (d) Combine the results: Using the inclusion - exclusion formula:

   $$\mathbf{P(A_{12} \cup A_{15})} = P(A_{12}) + P(A_{15}) - P(A_{12} \cap A_{15})$$
   $$\mathbf{P(A_{12} \cup A_{15})} = \frac{83}{1000} + \frac{66}{1000} - \frac{16}{1000}$$
   $$\mathbf{P(A_{12} \cup A_{15})} = \frac{133}{1000}$$

3. Step - 4 - Compute the Complement:

   $$\mathbf{P}(\texttt{Not divisible by 12 or 15}) = 1 - P(A_{12} \cup A_{15})$$
   $$\mathbf{P}(\texttt{Not divisible by 12 or 15}) = 1 - \frac{133}{1000}$$
   $$\mathbf{P}(\texttt{Not divisible by 12 or 15}) = \frac{867}{1000} \quad \square$$

Problem - 4:

Sit 3 men and 4 women at random in a row. What is the probability that either all the men or all the women end up sitting together?

Problem - 5:

Roll a single die 10 times. Compute the following probabilities:

1. that you get at least one 6;

    `Hint - Find Complementary Events and apply Inclusion Exclusion Principle.`

2. that you get at least one 6 and at least one 5.

    `Hint - Find Complementary Events and apply Inclusion Exclusion Principle.`

Problem - 6 - The Birthday Problem {Optional}:

Assume that there are k people in the room. What is the probability that there are two who share a birthday? Consider following:

- ignore leap years, assume all birthdays are equally likely;
- generalize the problem i.e. from n possible birthdays, sample k times with replacement.

3 Events thar are Joint - Conditional Probability and Examination of Independence.

3.1 Conditional Probability:

Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information.Here are some examples of situations we have in mind:

- In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?
- In a word guessing game, the first letter of the word is a "t". What is the likelihood that the second letter is an "h"?
- How likely is it that a person has a disease given that a medical test was negative?
- A spot shows up on a radar screen. How likely is it that it corresponds to an aircraft?

To extend further, Consider an experiment with a well-defined sample space and an associated probability law. Let A and B be two events within this space. Suppose it is known that event A occurs only if event B occurs, and the probability of event B is given. How does this information about B affect or update our assessment of the probability of A?
Thus, the theory of conditional probability construct a new probability law, which takes into account this knowledge and which, for any event A, gives us the conditional probability of A given B, denoted by $P(A|B)$ and also read A conditioned on B.

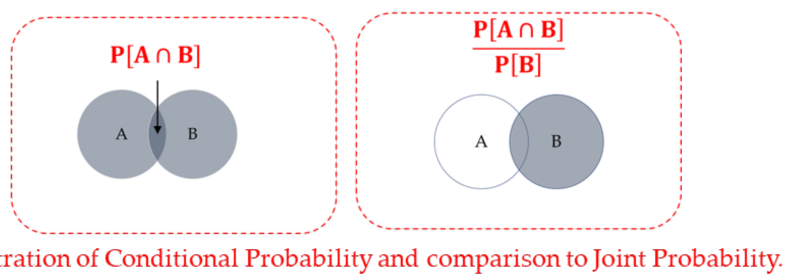


Figure 5: Conditional Probability - {Slide - 34}.

Mathematical Representation of Conditional Probability:

For any event A and B, if event A is dependent on B (or also called conditioned on B) and we know:

$$\mathbf{P(B) \neq 0}$$

Then Conditional Probability of A given B is :

$$\mathbf{P(A|B) = \frac{P(A \cap B)}{P(B)}}\{\texttt{read: A conditioned on B}\}$$

1. Conditional Vs. Joint Probability:

The difference between conditional Probability $\mathbf{P}(\mathbf{A}|\mathbf{B})$ and joint Event $\mathbf{P}(\mathbf{A}\cap\mathbf{B})$ is the denominator they carry i.e.

- Conditional Probability isolates the impact of B on A, focusing only on cases where B occurs.

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

$\rightarrow$ `focuses on likelihood of A given B has occurred.`

- Joint Probability measures the overlap of A and B in the entire sample space.

$$\mathbf{P}(\mathbf{A}\cap\mathbf{B}) = \frac{\mathbf{P}(\mathbf{A}\cap\mathbf{B})}{\mathbf{P}(\mathbf{\Omega})} = \mathbf{P}(\mathbf{A}\cap\mathbf{B})$$

$\rightarrow$ `represents the probability that both A and B occur simultaneously.`

2. Axioms of Conditional Probability:

The Axioms are:

Axioms of Probability.	Axioms – Conditional Probability.	
For any Event E – P(E)	For any Event E and F conditioned on – P(E|F)	
$0 \le P(E) \le 1$	$0 \le P(E\|F) \le 1$	Non – negativity.
$P(S) = 1$	$P(S\|F) = 1$	Normalizations.
$P(E \text{ or } F) = P(E) + P(F)$	$P(E \text{ or } G\|F) = P(E\|F) + P(F\|G)$	Additive Probability.
$P(E^c) = 1 - P(E)$	$P(E^c\|F) = 1 - P(E\|F)$	Complement Probbaility.

Table: The Paradigm of Conditional Probability.

Figure 6: Axioms of Conditional Probability - {Slide - 35}.

3.2 Independence:

We have introduced the conditional probability $\mathbf{P}(\mathbf{A}|\mathbf{B})$ to capture the partial information that event B provides about event A. An interesting and important special case arises when the occurrence of B provides no information and does not alter the probability that A has occurred, i.e.

$$\mathbf{P}(\mathbf{A}|\mathbf{B}) = \mathbf{P}(\mathbf{A})$$

When the above equality holds, we say that A is independent of B.

Mathematical Representation of Independence:

Two events are said to be independent if knowing the outcome of one event does not change your belief about whether or not the other event will occur. Mathematically:

$$\mathbf{P(A \cap B) = P(A) \times P(B)} \rightarrow \texttt{aka Multiplicative Rule of Probability.}$$

Iterative Extension to Multiple or n mutually independent events:

$$A_1, A_2, \dots, A_n \rightarrow \texttt{mututally independent of one another}$$

The independece equation holds for all susets of the events such as:

$$\mathbf{P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots P(A_n) = \prod_{i=1}^{n} P(A_i)} \quad \square$$

1. Conditional Interpretation of Independence:

From the definition of Conditional Probability:

$$\mathbf{P(A|B) = \frac{P(A \cap B)}{P(B)}} \rightarrow \mathbf{P(B) > 0}.$$

If A and B are independent, then the occurrence of B does not change the probability of A i.e.

$$\mathbf{P(A|B) = P(A)}$$

Substituting this into the definition of conditional probability:

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

Multiply both side by $\mathbf{P(B)}$:

$$\mathbf{P(A \cap B) = P(A) \cdot P(B)} \quad \square$$

3.3 Final Thoughts - Joint Probability:

Joint probability is the probability of two or more events happening at the same time. If we have two events A and B, the joint probability:

$$\mathbf{P(A \cap B)} \quad \texttt{or sometimes also written as} \quad \mathbf{P(A, B)}$$

For example, if we consider rolling two dice,

- let event A be A: "rolling a 3 on the first die," and
- event B be B: "rolling a 5 on the second die."

The joint probability i.e. $\mathbf{P(A, B)}$ would be the probability of:

$$P(A, B) = P(\texttt{rolling a 3 on the first die}) \quad \textbf{and} \quad P(\texttt{5 on the second die})$$

Probability then could be calculated:

- For independent events, where one event does not influence the outcome of the other, the joint probability is simply the product of their individual probabilities:

$$\mathbf{P(A \cap B) = P(A) \cdot P(B)}$$

- For dependent events, where the occurrence of one event affects the probability of the other, the joint probability is found by multiplying the probability of one event by the conditional probability of the second event given the first:

$$\mathbf{P(A \cap B) = P(A) \cdot P(B|A)}$$

3.4 Exercise - Conditional Probability and Independence:

Problem - 1:

E and F are mutually exclusive events. Such that:

$$\mathbf{P(E)} = 0.4$$
$$\mathbf{P(F)} = 0.5$$

Find $\mathbf{P(E|F)}$.

Sample Solution

Give:

- If E and F are mutually exclusive events then:

$$\mathbf{P(E \cap F) = 0}$$

  That is they can not occur together.
- $\mathbf{P(E) = 0.4}$
- $\mathbf{P(F) = 0.5}$

To Find:

$$\mathbf{P(E|F)} = \frac{P(E \cap F)}{P(F)} \rightarrow \texttt{From the definition of conditional probability}$$
$$\mathbf{P(E|F)} = \frac{0}{0.5} \rightarrow P(E \cap F) = 0$$
$$\mathbf{P(E|F)} = 0 \quad \square$$

This result indicates that if E and F are mutually exclusive, the probability of E occurring given that F has occurred is zero, as they cannot both happen at the same time.

Problem - 2:

J and K are independent events. $\mathbf{P(J|K) = 0.3}$. Find $\mathbf{P(J)}$.

Problem - 3:

Let event:

$$\mathbf{C} = \texttt{taking an Math Class}$$

$$\mathbf{D} = \texttt{taking an Programming Class}$$

Suppose:

$$\mathbf{P(C) = 0.75; P(D) = 0.3; P(C|D) = 0.75} \quad \texttt{and} \quad \mathbf{P(C \cap D) = 0.25}.$$

Justify your answers to the following questions numerically:

1. Are $\mathbf{C}$ and $\mathbf{D}$ independent?
2. Are $\mathbf{C}$ and $\mathbf{D}$ mutually exclusive?
3. What is $\mathbf{P(D|C)}$?

Problem - 4:

A family has two children. Given that at least one child is a girl, what is the probability that both children are girls? $[\mathbf{\frac{1}{3}}]$

Problem - 5:

A survey shows that 60% of people drink tea, 50% drink coffee, and 30% drink both. Are drinking tea and coffee independent events?

Hint:

To determine whether the events "drinking tea" (T) and "drinking coffee" (C) are independent, we need to check if the following condition holds:

$$\mathbf{P(T \cap C) = P(T) \cdot P(C)}$$

If this condition holds, the events are independent. Otherwise, they are dependent.

4 {Discrete} Random Variable.

Motivating Example - Why do we need Random variables?

Consider an experiment involving the flipping of 10 coins, where the objective is to determine the number of coins that result in heads.The sample space $\mathbf{\Omega}$consists of all possible sequences of heads and tails of length 10. For instance, one possible outcome might be:

$$\omega_0 = \{\texttt{H, H, T, H, T, H, H, T, T, T}\} \in \Omega$$

However, in practice, the probability of any specific sequence of heads and tails is often less significant than the evaluation of certain real-valued functions of these outcomes. Examples of such functions include the total number of heads obtained in the 10 tosses or the length of the longest consecutive run of tails. Under appropriate technical conditions, these functions are referred to as **random variables**.

Introduction to Random Variables{Discrete}

In a layman terms, random variables is a numeric $\mathbb{R}$ value assigned to each outcome of a chance/random experiment.

- In a way it can be thought of as a variable from programming, but this is a function that maps a uncertainty of any event in sample space with uncertainty.
- For each element (outcomes) in a sample space, the random variable can take on exactly one value i.e. random variable must be measurable.

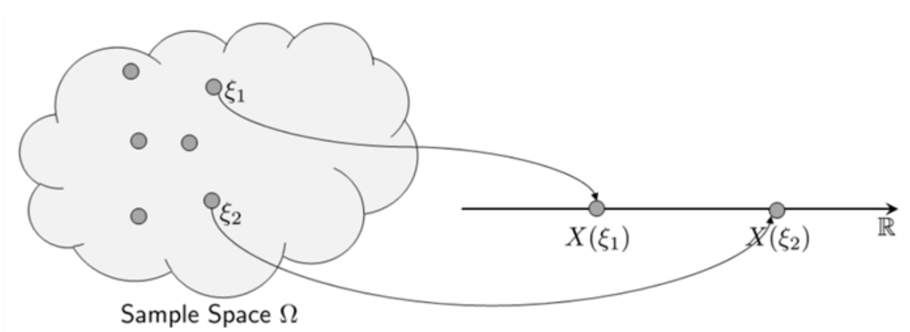


Fig: A definition of Random Variable.

Figure 7: [Pictorial Representation of Random Variable - {Slide - 46}.

Mathematical Formulation of Random Variable:

A random variable $\mathbf{X}$ is a function:

$$\mathbf{X} : \mathbf{\Omega} \to \mathbb{R}$$

that maps an outcome $\omega \in \Omega$ to a number $X(\omega) = \mathbb{R}$ on the real line. Typically, we will denote random variables using upper case letters $\mathbf{X}(\texttt{Outcome})$. We will denote the value that a random variable may take on using lower case letters x.

Example:

In our experiment above, suppose that $\mathbf{X}(\mathrm{x})$ is the number of heads which occur in the sequence of tosses $\mathtt{x}$. Given that only 10 coins are tossed, $\mathbf{X}(\mathrm{x})$ can take only a finite number of values, so it is known as a discrete random variable. Here, the probability of the set associated with a random variable X taking on some specific value k is:

$$\mathbf{P}(\mathbf{X}=\mathbf{k}) := \mathbf{P}(\{\mathtt{x:}\quad \mathtt{X(x)\ =\ k}\})$$

Here are some more examples of Discrete Random variables:

1. Choose a random point in the unit square

$$\{(\mathbf{x},\mathbf{y}) : \mathbf{0} \leq \mathbf{x},\mathbf{y} \leq \mathbf{1}\}$$

   and let X be its distance from the origin.
2. Choose a random person in a class and let X be the height of the person, in inches.
3. Let X be the value of the NASDAQ stock index at the closing of the next business day

Random Variables in Practice:

Experiment: Flip a three fair coins:

- Observations of interest$\mathbf{Y}$: Number of "heads" on the three coins.
- $\mathbf{Y}$ is an Random Variable, which maps observation of interest to real numbers.
  - A very big Question – What numbers and How?
  - By definition I can map to any number in real number line, but that does not mean we start mapping randomly by assigning any numbers,
  - In practice we follow the convention: i.e. Which in general is answer to the question: `How many possible combinations of outcome for our observation of interest is possible?`
- For above experiment:

`How many Possible outcomes are possible?`

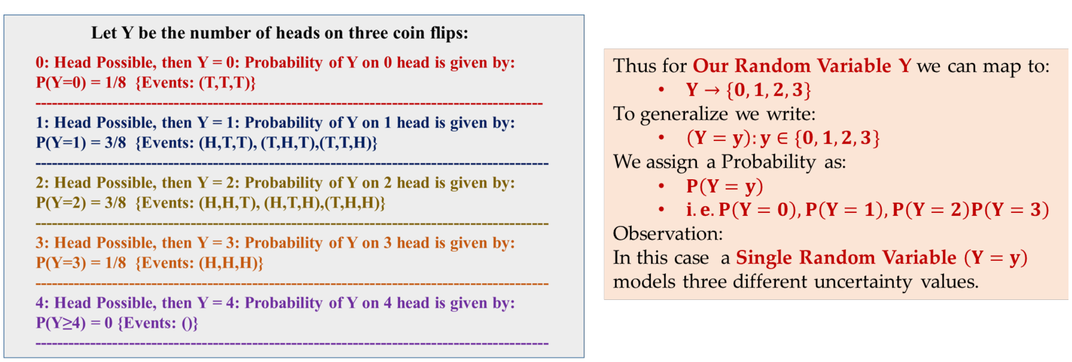


Figure 8: Assigning Random Variables - Convention - {Slide - 49 - 50}.

4.1 Distribution of Discrete Random Variable - Probability Mass Function:

The probability of a 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.
- Probability Distribution for Continuous Random Variables are also known as Probability Density Function.

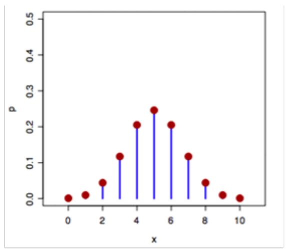


Fig: Discrete Probability Distribution

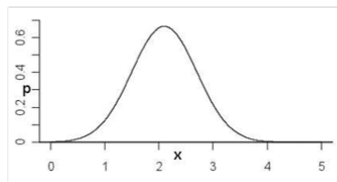


Fig: Continuous Probability Distribution
image from internet : may subjected to copyright.

Figure 9: Probability Mass Function and Probability Density Function.

Probability Mass Function (PMF)

The Probability Mass Function (PMF) is a mathematical representation of the probability distribution of a discrete random variable Y. It provides the probabilities that Y assumes specific values and can be expressed in the form of:

- A mathematical formula,
- A tabular representation, or
- A graphical visualization.

For a discrete random variable, the PMF assigns nonzero probabilities to a countable set of distinct values y. Any value of y not explicitly assigned a probability is assumed to satisfy $p(y) = 0$.

Properties of the PMF

1. **Non-negativity**: The probability for each y must satisfy:

$$\mathbf{0 \leq p(y) \leq 1} \quad \text{for all } \mathbf{y}.$$

2. **Normalization**: The sum of probabilities across all possible values of y must equal 1:

$$\sum_{\mathbf{y}} \mathbf{p(y) = 1},$$

   where the summation is taken over all values of $\mathbf{y}$ with $\mathbf{p(y) > 0}$.

These properties ensure that the PMF defines a valid probability distribution for discrete random variables.

Example Problem on PMF:

PMF of a Dice Roll:

- Question - **PMF of a Dice Roll:** A fair six-sided die is rolled once. Define the PMF $\mathbf{p(x)}$ of the outcome $\mathbf{X}$
  1. Write $\mathbf{p(x)}$ explicitly for all $\mathbf{x}$.
  2. Compute $\mathbf{P(X > 4)}$.
- **Solution:**
  (a) Writing the PMF $\mathbf{p(x)}$:
  The outcome X of rolling a fair six-sided die can be any integer from 1 to 6. Since the die is fair, each outcome has an equal probability of occurring. The PMF $p(x)$ is defined as:

$$\mathbf{p(x) = P(X = x)}$$

For a fair die, the probability of each face appearing is:

$$\mathbf{p(x) = \frac{1}{6}} \quad \text{for } \mathbf{x = 1, 2, 3, 4, 5, 6}$$

So, the PMF $p(x)$ is:

$$\mathbf{p(x)} = \begin{cases} \frac{1}{6}, & \text{if } x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

  (b) The probability $\mathbf{P(X > 4)}$ is:

$$\mathbf{P(X > 4) = \frac{1}{3}}$$

Cumulative Distribution Function{Optional}:

The Cumulative Distribution Function (CDF) for a discrete probability mass function (PMF) is a mathematical function that gives the probability that a random variable $\mathbf{Y}$ takes a value less than or equal to a specific value $\mathbf{y}$. It is defined as:

$$\mathbf{F_y = P(Y \leq y) = \sum_{t \leq y} p_Y(t)},$$

Where:

- $\mathbf{F_Y(y)}$: CDF of the random variable $\mathbf{Y}$.
- $\mathbf{p_Y(t)}$: PMF of $\mathbf{Y}$, which gives the probability $\mathbf{P(Y = t)}$,
- $\mathbf{t \leq y}$: The summation includes all values of $\mathbf{t}$ less than or equal to $\mathbf{y}$.

Key Properties of CDF:

1. Non-decreasing: The cumulative distribution function (CDF) $\mathbf{F_Y(y)}$ is non-decreasing:

$$\mathbf{F_Y(y_1) \leq F_Y(y_2)} \quad \text{for } \mathbf{y_1 < y_2}.$$

2. Limits:
$$\lim_{\mathbf{y}\to-\infty} \mathbf{F_Y(y) = 0}, \quad \text{and} \quad \lim_{\mathbf{y}\to\infty} \mathbf{F_Y(y) = 1}.$$
3. Step Function: For discrete random variables, the CDF increases in discrete steps corresponding to the probabilities at each $\mathbf{y}$.

Example Problem:

Suppose $\mathbf{Y}$ is a discrete random variable with the following PMF:
$$p_Y(y) = \begin{cases} 0.1, & \text{if } y = 1, \\ 0.2, & \text{if } y = 2, \\ 0.3, & \text{if } y = 3, \\ 0.4, & \text{if } y = 4. \end{cases}$$
Compute the CDF.

Example Problem - Solution:

Computing the CDF:

- The CDF, $\mathbf{F_Y(y)}$, is defined as:
$$\mathbf{F_Y(y) = P(Y \leq y) = \sum_{t \leq y} p_Y(t)}.$$
- Let's compute $\mathbf{F_Y(y)}$ for various values of $\mathbf{y}$:
  1. For $\mathbf{y < 1}$:
  $$\mathbf{F_Y(y) = 0} \quad \texttt{(since there is no probability for values less than 1).}$$
  2. For $\mathbf{1 \leq y < 2}$:
  $$\mathbf{F_Y(y) = P(Y = 1) = 0.1}.$$
  3. For $\mathbf{2 \leq y < 3}$:
  $$\mathbf{F_Y(y) = P(Y = 1) = 0.1}.$$
  4. For $\mathbf{2 \leq y < 3}$:
  $$\mathbf{F_Y(y) = P(Y = 1) + P(Y = 2) = 0.1 + 0.2 = 0.3}.$$
  5. For $\mathbf{3 \leq y < 4}$:
  $$\mathbf{F_Y(y) = P(Y = 1) + P(Y = 2) + P(Y = 3) = 0.1 + 0.2 + 0.3 = 0.6}.$$

  For $\mathbf{y \geq 4}$:
  $$\mathbf{F_Y(y) = P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) = 0.1 + 0.2 + 0.3 + 0.4 = 1.0}.$$

- CDF Function Representation: The CDF $\mathbf{F_Y(y)}$ can be written as:

$$\mathbf{F_Y(y)} = \begin{cases} 0, & \text{if } y < 1, \\ 0.1, & \text{if } 1 \leq y < 2, \\ 0.3, & \text{if } 2 \leq y < 3, \\ 0.6, & \text{if } 3 \leq y < 4, \\ 1.0, & \text{if } y \geq 4. \end{cases}$$

- Interpretation:
  - The value $\mathbf{F_Y(1) = 0.1}$ means that the probability that $\mathbf{Y}$ takes a value less than or equal to 1 is 0.1.
  - The value $\mathbf{F_Y(3) = 0.6}$ indicates that the probability that $\mathbf{Y}$ takes a value less than or equal to 3 is 0.6.
  - The value $\mathbf{F_Y(4) = 1.0}$ means that the probability that $\mathbf{Y}$ takes a value less than or equal to 4 is 1.0, which is expected as 1 is the total probability for the distribution.

  This step-by-step approach shows how the CDF accumulates probabilities from the PMF, giving the probability that $\mathbf{Y}$ is less than or equal to each value of $\mathbf{y}$.

4.2 Statistics of Discrete Random Variable:

1. Expected Value of a Discrete Random Variable:

Assume that $\mathbf{X}$ is a discrete random variable with possible values $\mathbf{x_i}$, $\mathbf{i = 1, 2, \ldots}$. The expected value, also called expectation, average, or mean, of $\mathbf{X}$ is defined as:

$$\mathbf{E[X] = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i).}$$

2. Properties of Expectation of Discrete Random Variable:

Property	**Mathematical Form**
`Linearity`	$\mathbf{E[aX + bY] = aE[X] + bE[Y]}$
`Expectation of a Constant`	$\mathbf{E[c] = c}$
`Multiplication by a Constant`	$\mathbf{E[aX] = aE[X]}$
`Additivity`	$\mathbf{E[X + Y] = E[X] + E[Y]}$
`Non-negativity for Non-negative Functions`	If $\mathbf{X \geq 0}$, `then` $\mathbf{E[X] \geq 0}$
`Law of the Unconscious Statistician`	$\mathbf{E[g(X)] = \sum_x g(x)p(x)}$ (discrete)

3. Variance of a Discrete Random Variable:

The variance of a random variable $\mathbf{X}$ is defined as:

$$\mathtt{Var}(\mathbf{X}) = \mathbf{E}[(\mathbf{X} - \mathbf{E}[\mathbf{X}])^{\mathbf{2}}] = \mathbf{E}[\mathbf{X}^{\mathbf{2}}] - (\mathbf{E}[\mathbf{X}])^{\mathbf{2}}$$

4. Properties of a Variance:

Property	**Mathematical Form**
`Variance of a Constant`	$\mathtt{Var}(\mathbf{c}) = \mathbf{0}$
`Scaling by a Constant`	$\mathtt{Var}(\mathbf{aX}) = \mathbf{a}^{\mathbf{2}}\mathtt{Var}(\mathbf{X})$
`Additivity for Independent Variables`	`If` $\mathbf{X}$ `and` $\mathbf{Y}$ `are independent,` $\mathtt{Var}(X + Y) = \mathtt{Var}(\mathbf{X}) + \mathtt{Var}(\mathbf{Y})$
`General Additivity (Not Independent)`	$\mathtt{Var}(\mathbf{X} + \mathbf{Y}) = \mathtt{Var}(\mathbf{X}) + \mathtt{Var}(\mathbf{Y}) + \mathbf{2}\mathtt{Cov}(\mathbf{X}, \mathbf{Y})$
`Non-negativity`	$\mathtt{Var}(\mathbf{X}) \geq \mathbf{0}$
`Law of Total Variance`	$\mathtt{Var}(\mathbf{X}) = \mathbf{E}[\mathtt{Var}(\mathbf{X}\|\mathbf{Y})] + \mathtt{Var}(\mathbf{E}[\mathbf{X}\|\mathbf{Y}])$

4.3 Exercise - Discrete Random Variable:

Problem - 1 - Basic PMF Evaluation:

Let the PMF of a discrete random variable X be given by:

$$p(x) = \begin{cases} \frac{x}{10}, & \text{if } x = 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases}$$

(a) Verify that $p(x)$ satisfies the properties of a PMF.

(b) Find $P(X = 3)$.

(c) Compute $P(X \leq 3)$.

Solution:

(a) Verify that $\mathbf{p}(\mathbf{x})$ satisfies the properties of a PMF.The properties of a PMF are:
- $p(x) \geq 0$ for all x.
- $\sum_x p(x) = 1$, i.e., the sum of the probabilities of all possible outcomes should be 1.

Given the PMF:

$$\mathbf{p}(\mathbf{x}) = \begin{cases} \frac{x}{10}, & \text{if } x = 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases}$$

Verification:
- **Non-negativity:**

- For $x = 1, 2, 3, 4$:

$$p(1) = \frac{1}{10} > 0, \quad p(2) = \frac{2}{10} > 0, \quad p(3) = \frac{3}{10} > 0, \quad p(4) = \frac{4}{10} > 0.$$

Since $p(x) \geq 0$ for all x, the first property holds.

- **Normalization:**

$$\sum_{x=1}^{4} p(x) = \frac{1}{10} + \frac{2}{10} + \frac{3}{10} + \frac{4}{10} = \frac{10}{10} = 1.$$

This confirms that the PMF is properly normalized.

(b) Find $\mathbf{P(X = 3)}$ From the given PMF:

$$\mathbf{p(3) = \frac{3}{10}}.$$

So,

$$\mathbf{P(X = 3) = \frac{3}{10}}.$$

(c) Compute $\mathbf{P(X \leq 3)}$
To find $\mathbf{P(X \leq 3)}$, sum the probabilities for $\mathbf{x = 1, 2, 3}$:

$$\mathbf{P(X \leq 3) = p(1) + p(2) + p(3)}.$$

Substitute the values:

$$\mathbf{P(X \leq 3) = \frac{1}{10} + \frac{2}{10} + \frac{3}{10} = \frac{6}{10} = \frac{3}{5}}.$$

1. Final Answer:
   (a) The PMF $\mathbf{p(x)}$ satisfies the properties of a PMF.
   (b) $\mathbf{P(X = 3) = \frac{3}{10}}$.
   (c) $\mathbf{P(X \leq 3) = \frac{3}{5}}$.

4.4 Problem -2: Missing PMF Constant:

A PMF $\mathbf{p(x)}$ of a random variable $\mathbf{X}$ is given as:

$$\mathbf{p(x)} = \begin{cases} kx, & \text{if } x = 1, 2, 3 \\ 0, & \text{otherwise} \end{cases}$$

(a) Determine the value of $\mathbf{k}$ such that $\mathbf{p(x)}$ is a valid PMF.

(b) Find $\mathbf{P(X = 2)}$.

4.5 Some Popular Special Discrete Distribution:

1. Bernoulli Trial and Bernoulli Distribution:

A **Bernoulli trial** is a random experiment with exactly two possible outcomes, commonly labeled as "success" and "failure." The probability of success is denoted by p, and the probability of failure is $1 - p$. Each trial is independent, meaning the outcome of one trial does not influence the outcomes of others.

A **Bernoulli distribution** describes the probability distribution of a random variable X that takes the value 1 (success) with probability p and the value 0 (failure) with probability $1 - p$. The probability mass function (PMF) is given by

$$\mathbf{P(X = x) = p^x (1 - p)^{1-x}}, \quad \text{where } \mathbf{x \in \{0, 1\}}.$$

The expectation of $\mathbf{X}$ is:

$$\mathbf{E[X] = p}$$

and the variance is:

$$\texttt{Var}(\mathbf{X}) = \mathbf{p(1 - p)}$$

Bernoulli trials and distributions are foundational concepts in probability theory and are widely used in statistical modeling and machine learning.

2. Binomial Distribution:

The **Binomial distribution** describes the probability distribution of the number of successes in a fixed number of independent Bernoulli trials. If a random variable $\mathbf{X}$ follows a Binomial distribution with $\mathbf{n}$ trials and probability of success $\mathbf{p}$, it is denoted as $\mathbf{X} \sim \text{Bin}(\mathbf{n}, \mathbf{p})$. The probability mass function (PMF) is given by:

$$\mathbf{P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}}, \quad \text{where } \mathbf{k = 0, 1, 2, \ldots, n},$$

and $\mathbf{\binom{n}{k} = \frac{n!}{k!(n-k)!}}$ is the binomial coefficient.
The expectation and variance of $\mathbf{X}$ are given by:

$$\mathbf{E[X] = np}, \quad \text{Var}(\mathbf{X}) = \mathbf{np(1 - p)}.$$

The Binomial distribution is widely used in statistics and probability to model scenarios where a fixed number of repeated independent trials are performed, such as flipping a coin or testing defective products.