DSC511: Statistical Foundations for Data Science

Basic Concepts

Bakkyaraj T
Indian Institute of Information Technology Kottayam
Email:bakkyaraj@iiitkottayam.ac.in

What is Probability Theory?

- Randomness and uncertainty exist in our daily lives and in science, engineering and technology.
- We require a mathematical framework that allows us to analyze random phenomena.
- **Probability theory** provides us such a framework.
- But what do we mean by **random phenomena** and **probability**? How can we express **randomness**?

Randomness

- We define **random phenomena** as events and experiments whose outcomes we cannot predict with certainty.
- For example: Think about flipping a fair coin. We cannot predict whether the outcome would be heads or tails.

Application in a Communication System

- Communication systems transfer information from one place to another as a sequence of 1's and 0's called bits.
- This transmission is often affected by noise, and the information corrupted.
- The figure shows that the transmitted and received bits are different.

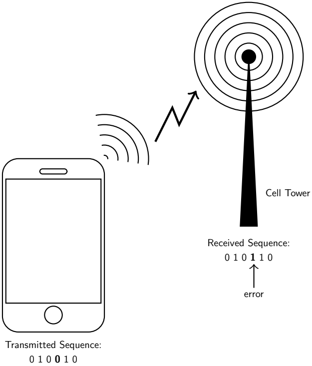


Figure: Transmission of data from a cell phone to a cell tower.

Application in a Communication System

- Such errors affect the quality of our transmission and need to be minimized.
- Noise is a random phenomena and we do not know which bits will be affected before transmission.
- Probability theory is used extensively in the design of communication systems to
  a. Understand the behavior of noise.
  b. Take measures in the system to correct errors.

Set Theory

- A **set** is a well defined collection of objects/things called **elements**.
- A set is denoted in capital letters and defined by simply listing its elements in curly brackets. Example: $A = \{b, c\}$.
- Can also be defined as $A = \{$x:x satisfies some property$\}$.
- <u>Ordering does not matter in sets</u>. Thus $\{1, 2, 3, 4\}$ and $\{3, 2, 1, 4\}$ are the same set.
- $b \in A$ read as b belongs to A where $\in$ means belongs to.
- And $d \notin A$, where $\notin$ means does not belong.

Important Sets

- The set of natural numbers, $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$
- The set of integers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- The set of rational numbers $\mathbb{Q}$.
- The set of real numbers $\mathbb{R}$ and the set of complex numbers $\mathbb{C}$.
- Closed intervals on the real line. Example: [2,3] is set of real numbers such that $2 \leq x \leq 3$.
- Open intervals on the real line. Example: (1,2) is the set of real numbers such that $1 < x < 2$.

More on Sets

- Set A is a **subset** of set B if every member of A is also a member of B. We write $A \subset B$, where $\subset$ indicates subset.
- Equivalently B is the **superset** of A, $B \supset A$.
- Two sets are **equal** $A = B$, if they contain the same elements, that is $A \subset B$ and $B \subset A$
- The **universal set** S or Ω is the set of all things that we could possibly consider in the context we are studying.
- The universal set in probability is also called the **sample space**.
- The set with no elements is called the **empty** or **null set** $\emptyset = \{\}$.

Venn Diagrams

- Venn Diagrams are very useful in visualizing relations between sets.
- In Venn Diagrams, a set is depicted by a closed region.

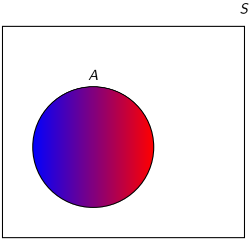


Figure: Venn Diagram

Venn Diagrams

- The figure below shows two sets, A and B, where $B \subset A$.
- Both A and B are subsets of the universal set S.

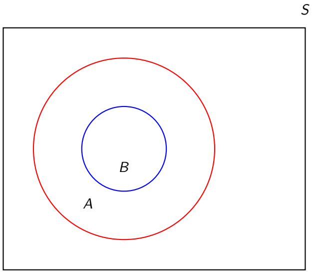


Figure: Venn Diagram for two sets A and B, where $B \subset A$.

Set Operations: Union

- The union of two sets is a set containing all elements that are in A or in B.
- Example: $\{1,2\} \cup \{2,3\} = \{1,2,3\}$.
- In general the union of n sets $A_1, A_2, ..., A_n$ is represented as $\bigcup_{i=1}^{n} A_i$.

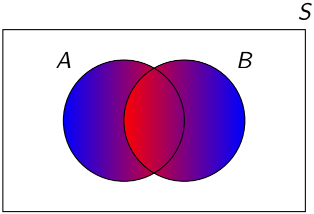


Figure: The shaded area shows the set $B \cup A$.

Set Operations: Intersection

- The intersection of two sets A and B is a set containing all elements that are in A and B.
- Example: $\{1, 2\} \cap \{2, 3\} = \{2\}$.
- In general, the intersection of n sets $\bigcap_{i=1}^{n} A_i$ is the set consisting of elements that are in all n sets.

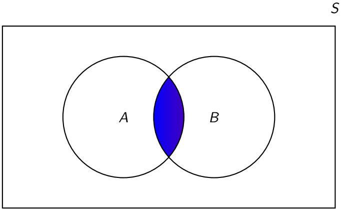


Figure: The shaded area shows the set $B \cap A$.

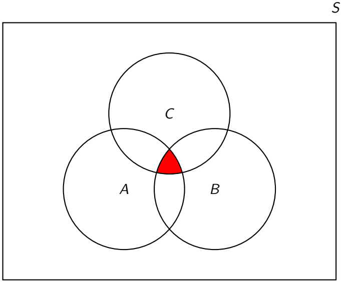


Figure: The shaded area shows the set $A \cap B \cap C$.

Set Operations: Complement

- The complement of a set A is the set of all elements that are in the universal set S but not in A.

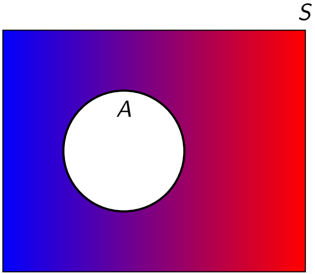


Figure: The shaded area shows the set $\bar{A} = A^c$.

Mutually Exclusive or Disjoint Sets

- Sets A and B are mutually exclusive or disjoint if they do not have any shared elements.
- The intersection of two sets that are disjoint is the empty set i.e. $A \cap B = \emptyset$.
- In general, several sets are disjoint if they are pairwise disjoint.

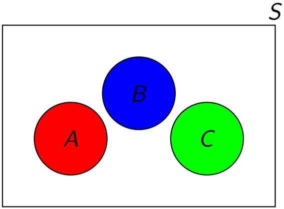


Figure: Sets A, B, and C are disjoint.

Partition of Sets

- A collection of non-empty set $A_1, A_2, \ldots$ is a **partition** of A if they are disjoint and their union is A.

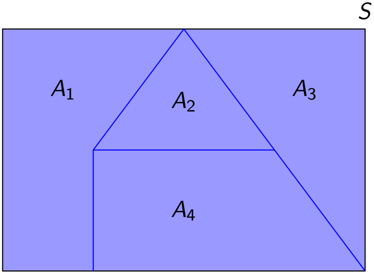


Figure: The collection of sets A_1, A_2, A_3 and A_4 is a partition of S.

Important Theorems

- De Morgan's Law:

  For any two sets A_1, A_2, we have:
  - $(A_1 \cup A_2)^c = {A_1}^c \cap {A_2}^c$
  - $(A_1 \cap A_2)^c = {A_1}^c \cup {A_2}^c$

- Distributive Law

  For any sets
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Cardinality in Finite sets

- **Cardinality** is basically the size of the set.
- If set A only has a finite number of elements, its cardinality is simply the number of elements in A.
- For example, if $A = \{2, 4, 6, 8, 10\}$, then $\mid A \mid = 5$.

Inclusion-Exclusion Principle

- The inclusion-exclusion principle states that for two finite sets A, B and C.
  - $|A \cup B| = |A| + |B| - |A \cap B|$,
  - $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$.

Cardinality in Infinite Sets

- There are two kinds of infinite sets: countable sets and uncountable sets.
- The difference between the two is that you can list elements in a countable set, so $A = \{a_1, a_2, \ldots\}$, but you cannot list elements in an uncountable set.
- The set $\mathbb{R}$ is uncountable and much *larger* than countably infinite sets $\mathbb{N}$ and $\mathbb{Z}$.

Countable vs. Uncountable Sets

- A more rigorous definition of a countable set A is
  - if it is a finite set, $|A| < \infty$; or
  - it can be put in one-to-one correspondence with natural numbers $\mathbb{N}$, in which case the set is said to be countably infinite.
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ and any of their subsets are countable.
- Any set containing an interval on the real line such as $[a, b], (a, b], [a, b)$ and (a, b), where $a < b$ is uncountable.

Functions

- A **function** maps elements from the **domain** set to elements in another set called the **co-domain**.
- Each input in the domain is mapped to exactly one output in the co-domain.
- It is denoted as $f : A \to B$.
- The **range** of a function is the set of all possible values of $f(x)$ and is a subset of the co-domain.

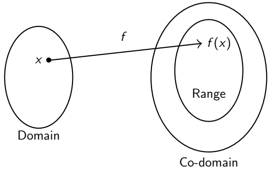


Figure: Function $f : A \to B$, the range is always a subset of the co-domain.

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.
- An **outcome** is a result of a random experiment.
- The set of all possible outcomes is called the **sample space** and in this context the universal set.
- When we repeat a random experiment several times, we call each one a trial.
- An **event** is a subset of the sample space.

Probability

- We assign a probability measure $P(A)$ to an event A.
- This is a value set between 0 and 1 that shows how likely the event is and is such that
  - If $P(A)$ is close to 0, the event A is very unlikely to occur.
  - If $P(A)$ is close to 1, the event A is very likely to occur.
- Probability theory is based on the following axioms that act as the foundation for the theory.

Axioms of Probability

- **Axiom 1**: For any event A, $P(A) \geq 0$
- **Axiom 2**: Probability of the sample space S is $P(S) = 1$
- **Axiom 3**: If $A_1, A_2, A_3 \ldots$ are disjoint events, then $P(A_1 \cup A_2 \cup A_3 \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$.

- It is important to note that *union* means *or* and *intersection* means *and*.
  a. $P(A \cap B) = P(\text{A and B}) = P(A, B)$.
  b. $P(A \cup B) = P(\text{A or B})$.

Finding Probabilities

- To find the probability of an event, we usually follow these two steps
  a. We use the specific information that we have about the random experiment.
  b. We then use the probability axioms seen in the previous slide.
- We shall employ these steps in discrete and continuous probability models.

Inclusion Exclusion Principle and Other Useful Results

- Inclusion-Exclusion Principle

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

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

- $P(A^C) = 1 - P(A)$.
- Probability of the empty set is zero $P(\phi) = 0$.
- For any event $P(A) \leq 1$.

Discrete Probability Models

- Consider a sample space S. If S is a countable set, this refers to a discrete probability model. Since S is countable, we can list all the elements in S as $S = \{s_1, s_2, \dots\}$.
- If A $\subset$ S is an event, then A is also countable, and by the 3rd axiom of probability, we can say that
$$P(A) = P(\bigcup_{s_j \in A} \{s_j\}) = \sum_{s_j \in A} P(s_j)$$
- We sum the probability of individual elements in that set to find the probability of an event.

Finite Sample Space with Equally Likely Outcomes

- A special case of discrete probability model is a finite sample space where each outcome is equally likely that is $S = \{s_1, s_2, \ldots, s_N\}$ where $P(s_i) = P(s_j)$ for all $i, j \in \{1, 2, \ldots, N\}$.
- Since all outcomes are equally likely we have $P(s_i) = \frac{1}{N}$ for all $i \in \{1, 2, ..., N\}$.
- If A is an event with cardinality $|A| = M$, we have

$$P(A) = \sum_{s_j \in A} P(s_j) = \sum_{s_j \in A} \frac{1}{N} = \frac{M}{N} = \frac{|A|}{|S|}.$$

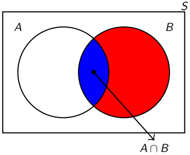
- Finding probability of A reduces to a *counting* problem.

Conditional Probability

- How should you update probabilities of events given more information?
- For example, say we know the probability of the event R, rain in a certain city on a random day is 23%. Thus $P(R) = 0.23$.
- Suppose I pick a random day, but also say that it is cloudy on that given day. Now what is the probability that it rains **given that** it is cloudy?
- If we call the event cloudy as C, we want the conditional probability $P(R|C)$ i.e. probability of R given that C has already happened.
- We see that $P(R|C)$ will be greater than $P(R)$, which is called the **prior** probability.

Calculation of Conditional Probability

- If A and B are events in the sample space S, the conditional probability of A given B, when $P(B) > 0$ is:
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
- The intuition behind the formula is that once we know B has occurred, *our sample space reduces to the set B*.
- $P(A|B)$ is undefined when $P(B) = 0$.



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

Figure: Venn diagram for conditional probability, $P(A|B)$.

Axioms of Probability Revisited

- Conditional probability is a probability measure and must satisfy probability axioms

  - Axiom 1: For any event A, $P(A|B) \geq 0$.
  - Axiom 2: Conditional probability of B given B is 1, i.e. $P(B|B) = 1$.
  - Axiom 3: If $A_1, A_2, A_3, \cdots$ are disjoint events, then $P(A_1 \cup A_2 \cup A_3 \cdots |B) = P(A_1|B) + P(A_2|B) + P(A_3|B) + \cdots$.

Useful Results for Conditional Probability

- $P(A^C|C) = 1 - P(A|C)$
- $P(\phi|C) = 0$.
- $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.
- When A and B are disjoint, $P(A|B) = 0$.
- When B is a subset of A, $P(A|B) = 1$.
- When A is a subset of B, $P(A|B) = \frac{P(A)}{P(B)}$.

Chain Rule for Conditional Probability

- The chain rule is given as
$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$
- Useful in situations when we know the conditional probability, but we are interested in the probability of the intersection.
- A generalization: $P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1) \ldots P(A_n|A_{n-1}, A_{n-2}, \ldots, A_1)$

Independence

- Events A and B are independent if
$$P(A \cap B) = P(A)P(B).$$

- Independence is often confused with disjointness, but they are not the same thing, i.e. $P(A \cap B) = 0 \neq P(A)P(B)$.

- Three events A, B and C are independent if **all** of the following hold

$$P(A \cap B) = P(A)P(B),$$
$$P(A \cap C) = P(A)P(C),$$
$$P(B \cap C) = P(B)P(C),$$
$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Differences between Disjointedness and Independence

Concept	Meaning	Formulas
Disjoint	A and B cannot occur at the same time	$A \cap B = \emptyset$
Independent	A gives no information about B	$P(A\|B) = P(A)$, $P(B\|A) = P(B)$

Useful Results for Independent Events

- If A and B are independent then
  - A and B^C are independent.
  - A^C and B are independent.
  - A^C and B^C are independent.
- If $A_1, A_2, \ldots, A_n$ are independent, then $P(A_1 \cup A_2 \cup \ldots \cup A_n) = 1 - (1 - P(A_1))(1 - P(A_2)) \ldots (1 - P(A_n))$.

Law of Total Probability

- If $B_1, B_2, B_3 \ldots$ is a partition of sample space S, for any event A, we have

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$$

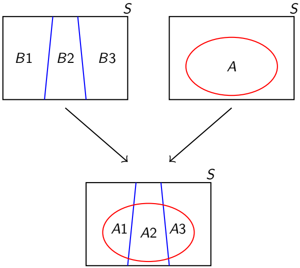


Figure: Law of total probability

Bayes Rule

- This rule allows us to calculate $P(B|A)$ from $P(A|B)$.
- For two events A and B, we have
$$P(B|A)P(A) = P(A \cap B) = P(A|B)P(B).$$
- Dividing by $P(A)$ gives Bayes rule
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
- If $B_1, B_2, B_3, \ldots$ form a partition of the sample space S, and A is any event with $P(A) \neq 0$, we have
$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$