Introduction to Probability

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Content Reference Book :

Introduction to Probability by Dimitri P. Bertsekas and John N. Tsitsiklis , MIT

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Background

- Generally concept of probability is used to discuss an uncertain situation
- The first approach is to define probability in terms of frequency of occurrence, as a percentage of successes in a moderately large number of similar situations.
- While there are many situations involving uncertainty in which the frequency interpretation is appropriate, there are other situations in which it is not.
- For example, the probability that my brother will top the coming Board Exam in 2021 is 90%
- This not based on frequency, but it is one time event. It is an expression of my subjective belief.
- One may think that subjective beliefs are not interesting, at least from mathematical or scientific point of view.
- On the other hand, people often have to make choices in the presence of uncertainty, and a systematic way of making use of their beliefs is a prerequisite for successful, or at least consistent, decision
- In fact, the choices and actions of a rational person, can reveal a lot about the inner-held subjective probabilities, even if the person does not make conscious use of probabilistic reasoning

Background

- The objectives behind learning Probability is
- Develop the art of describing uncertainty in terms of probabilistic models,
- As well as the skill of probabilistic reasoning,

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Set Definition

- Probability makes extensive use of set operations

- A set is a collection of objects, which are the elements of the set.
- If **S** is a set and $\boldsymbol{x}$ is an element of **S**, we write $\boldsymbol{x} \in \mathbf{S}$
- If $\boldsymbol{x}$ is not an element of **S**, we write $\boldsymbol{x} \notin S$
- A set can have no elements, in which case it is called the empty set, denoted by **Ø**
- If **S** contains a finite number of elements, say $\boldsymbol{x_1, x_2, \ldots, x_n}$
- We write it as a list of the elements, in braces

$$\boldsymbol{S} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\}$$

- The set of possible outcomes of a die roll is **{1, 2, 3, 4, 5, 6}**, and
- The set of possible outcomes of a coin toss is **{H, T}** , where **H** stands for "heads" and **T** stands for "tails."

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

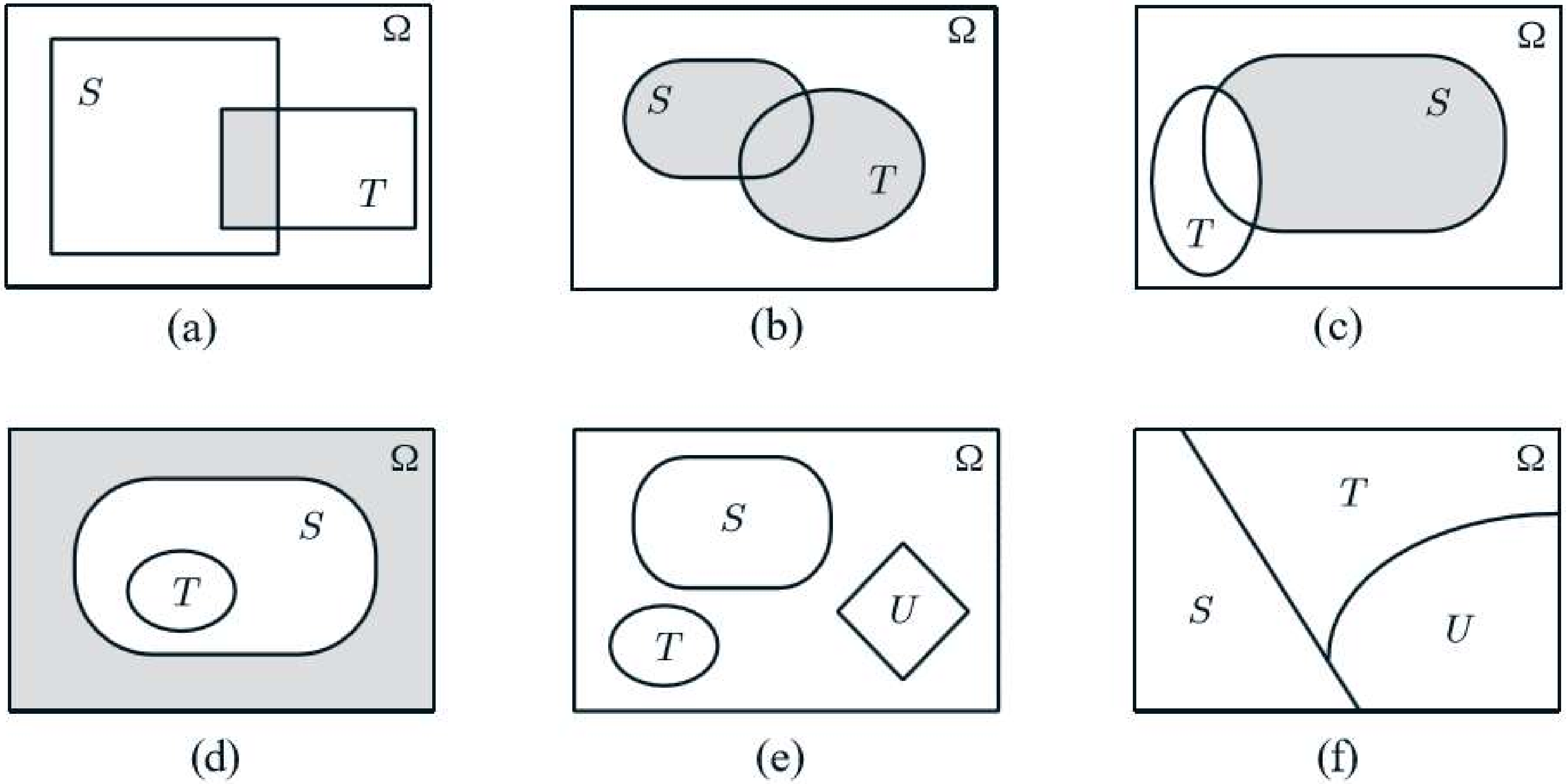
Set Representation

- If **S** contains infinitely many elements $x_1, x_2, \dots,$ which can be enumerated in a list (so that there are as many elements as there are positive integers).
- We write $\mathbf{S} = \{ x_1, x_2, \dots \}$ and we say that **S** is countably infinite
- For example, the set of even integers can be written as $\{0, 2, -2, 4, -4, \dots\}$, and is countably infinite
- Alternatively, we can consider the set of all x that have a certain property **P**, denoted $\{x \mid x \text{ satisfies } P\}$
- For example the set of even integers can be written as $\{k \mid k/2 \text{ is integer}\}$
- Similarly, the set of all scalars x in the interval $[0,1]$ can be written as $\{x \mid 0 \le x \le 1\}$.
- Note that the elements x of the latter set take a continuous range of values, and cannot be written down in a list, such a set is said to be uncountable
- If every element of a set **S** is also an element of a set **T**, we say that **S** is a subset of **T**, and we write $S \subset T$ or $T \supset S$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Set Operations

Examples of Venn diagrams

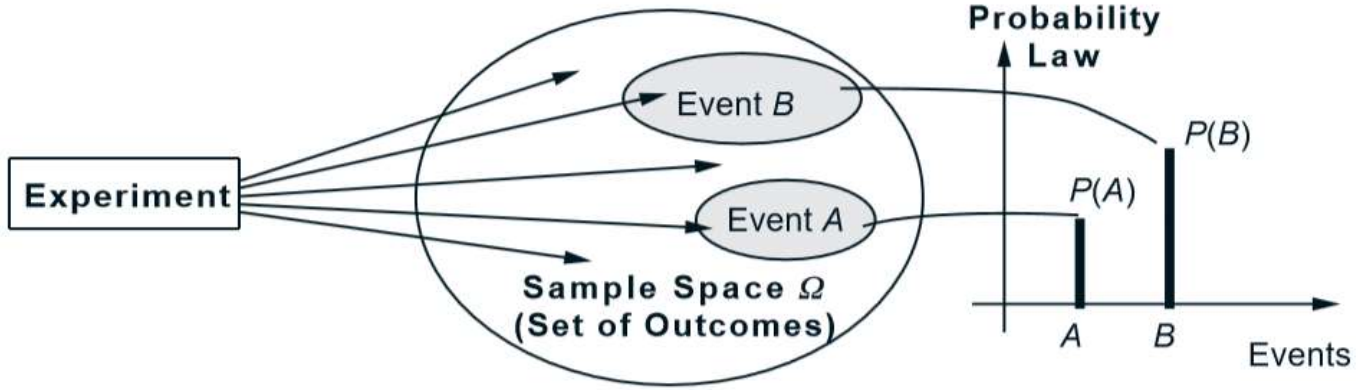


(a) The shaded region is **S** ∩ **T**. (b) The shaded region is **S** ∪ **T**. (c) The shaded region is **S** ∩ **T**^{c}. (d) Here, **T** ⊂ **S**. The shaded region is the complement of **S**. (e) The sets **S**, **T**, and **U** are disjoint. (f) The sets **S**, **T**, and **U** form a partition of the set **Ω**.

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models

- A probabilistic model is a mathematical description of an uncertain situation
- Its two main ingredients are listed below and are visualized



Elements of a Probabilistic Model

- The **sample space Ω**, which is the set of all possible outcomes of an experiment
- The **probability law**, which assigns to a set **A** of possible outcomes (also called an event) a non negative number **P(A)** (called the probability of A) that encodes our knowledge or belief about the collective "likelihood" of the elements of **A**. The probability law must satisfy certain properties.

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models

- Every probabilistic model involves an underlying process, called the experiment, that will produce exactly one out of several possible outcomes
- The **set of all possible outcomes** is called the sample space of the experiment, and is denoted by **Ω**. A subset of the sample space, that is, **a collection of possible outcomes**, is called an event
- Choosing an Appropriate Sample Space is the KEY
- Different elements of the sample space **should be distinct** and **mutually exclusive** so that when the experiment is carried out, there is a unique
- For example, the sample space associated with the roll of a die cannot contain "1 or 3" as a possible outcome and also "1 or 4" as another possible outcome.
- Generally, the sample space chosen for a probabilistic model must be **collectively exhaustive**, in the sense that no matter what happens in the experiment, we always obtain an outcome that has been included in the sample space

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models

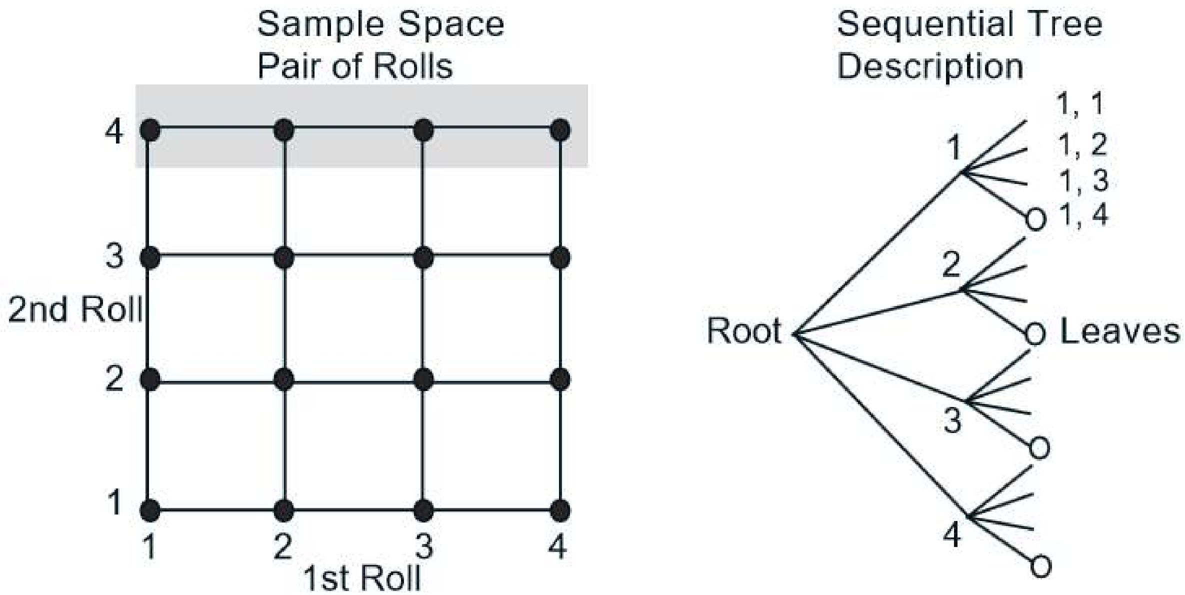
- Consider two alternative games, both involving ten successive coin tosses:
- **Game 1:** We receive **$1** each time a head comes up.
- **Game 2:** We receive **$1** for every coin toss, up to and including the first time a head comes up. Then, we receive **$2** for every coin toss, up to the second time a head comes up. More generally, the dollar amount per toss is doubled each time a head comes up
- In **Game 1**, it is only the total number of heads in the ten-toss sequence that matters, while in game 2, the order of heads and tails is also important.
- Thus, in a probabilistic model for game 1, we can work with a sample space consisting of eleven possible outcomes, namely, 0 , 1 , ... ,10
- In **Game 2**, a finer grain description of the experiment is called for, and it is more appropriate to let the sample space consist of every possible ten-long sequence of heads and tails

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models - Sequential

- Many experiments have an inherently sequential character, such as for example tossing a coin three times, or observing the value of a stock on five successive days, or receiving eight successive digits at a communication receiver
- It is then often useful to describe the experiment and the associated sample space by means of a Tree-based sequential description

Two equivalent descriptions of the sample space of an experiment involving two rolls of a 4-sided die



Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Probability Law

❖ The probability law assigns to every event **A**, a number **P(A)**, called the probability of **A**, satisfying the following axioms

Probability Axioms

1. **(Nonnegativity)** $\mathbf{P}(A) \geq 0$, for every event A.
2. **(Additivity)** If A and B are two disjoint events, then the probability of their union satisfies

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

Furthermore, if the sample space has an infinite number of elements and $A_1, A_2, \ldots$ is a sequence of disjoint events, then the probability of their union satisfies

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots$$

3. **(Normalization)** The probability of the entire sample space Ω is equal to 1, that is, $\mathbf{P}(\Omega) = 1$.

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Discrete

- Here is an illustration of how to construct a probability law starting from some common sense assumptions about a model
- **Coin tosses** - Consider an experiment involving a single coin toss. There are two possible outcomes, heads (**H**) and tails (**T**).
- The sample space is $\Omega = \{H, T\}$, and the events are $\{H,T\}$, $\{H\}$, $\{T\}$, $\emptyset$.
- If the coin is fair, i.e., if we believe that heads and tails are "equally likely," we should assign equal probabilities to the two possible outcomes and specify that $P(\{H\}) = P(\{T\}) = 0.5$
- The additive axiom implies $P(\{H, T\}) = P(\{H\}) + P(\{T\}) = 1$,
- which is consistent with the normalization axiom.
- Thus, the probability law is given by $P(\{H, T\}) = 1$, $P(\{H\}) = 0.5$, $P(\{T\}) = 0.5$, $P(\emptyset) = 0$, and satisfies all three axioms

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Discrete

- Consider another experiment involving three coin tosses. The outcome will now be a 3-long string of heads or tails.

  The sample space is Ω = { **HHH**, **HHT**, **HTH**, **HTT**, **THH**, **THT**, **TTH**, **TTT** }

- We assume that each possible outcome has the same probability of **1/8**. Let us construct a probability law that satisfies the three axioms.
- Consider, as an example, the event **A** = { **exactly 2 heads occur** } = { **HHT**, **HTH**, **THH** }.
- Using additivity, the probability of **A** is the sum of the probabilities of its elements:

$$\mathbf{P}(\{\,\mathbf{HHT}, \mathbf{HTH}, \mathbf{THH}\,\}) = \mathbf{P}(\{\,\mathbf{HHT}\,\}) + \mathbf{P}(\{\,\mathbf{HTH}\,\}) + \mathbf{P}(\{\,\mathbf{THH}\,\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

- Similarly, the probability of any event is equal to 1/8 times the number of possible outcomes contained in the event
- This defines a probability law that satisfies the three axioms

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Discrete

❖ By using the additivity axiom and by generalizing the reasoning in the preceding example, we reach the following conclusion

Discrete Probability Law: 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 $\{s_1, s_2, \ldots, s_n\}$ is the sum of the probabilities of its elements: $\mathbf{P}(\{s_1, s_2, \ldots, s_n\}) = \mathbf{P}(\{s_1\}) + \mathbf{P}(\{s_2\}) + \cdots + \mathbf{P}(\{s_n\})$

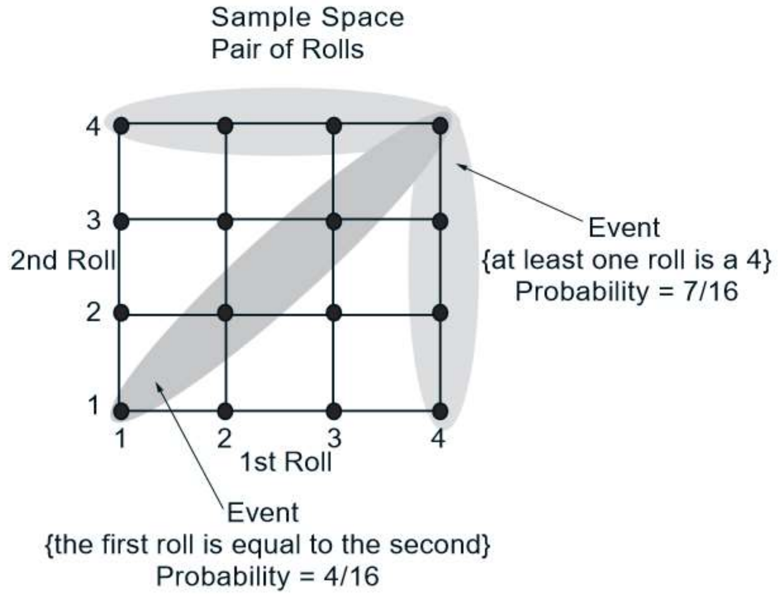
Discrete Uniform Probability Law: If the sample space consists of n possible outcomes which are equally likely (i.e., all single-element events have the same probability),

Then the probability of any event A is given by $$P(A) = \frac{\text{Number of elements of } A}{n}$$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Discrete

- Consider the experiment of rolling **a pair of 4-sided dice** and we assume the dice are fair, and we interpret this assumption to mean
- That each of the sixteen possible outcomes [**ordered pairs (i,j), with i,j = 1, 2, 3, 4**], has the same probability of $\frac{1}{16}$. To calculate the probability of an event, we must count the number of elements of event and divide by 16 (the total number of possible outcomes).
- P({ the sum of the rolls is even })= $\frac{8}{16} = \frac{1}{2}$
- P({ the sum of the rolls is odd }) = $\frac{8}{16} = \frac{1}{2}$
- P({ the first roll is equal to the second }) = $\frac{4}{16} = \frac{1}{4}$
- P({ the first roll is larger than the second })= $\frac{6}{16} = \frac{3}{8}$
- P({ at least one roll is equal to 4 }) = $\frac{7}{16}$



Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Continuous

- Probabilistic models with continuous sample spaces differ from their discrete counterparts in that the probabilities of the single-element events may not be sufficient to characterize the probability law.
- This is illustrated in the following examples, which also illustrate how to generalize the uniform probability law to the case of a continuous sample space

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Continuous

- A wheel of fortune is continuously calibrated from **0** to **1**, so the possible outcomes of an experiment consisting of a single spin are the numbers in the interval $\Omega = [\mathbf{0},\mathbf{1}]$.
- Assuming a fair wheel, it is appropriate to consider all outcomes equally likely, but what is the probability of the event consisting of a single element? It cannot be positive, because then, using the additivity axiom, it would follow that events with a sufficiently large number of elements would have probability larger than 1.
- Therefore, the probability of any event that consists of a single element must be 0. In this example, it makes sense to assign probability $\mathbf{b-a}$ to any subinterval $[\mathbf{a},\mathbf{b}]$ of $[\mathbf{0},\mathbf{1}]$, and to calculate the probability of a more complicated set by evaluating its “length.”
- This assignment satisfies the three probability axioms and qualifies as a legitimate probability law

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Continuous

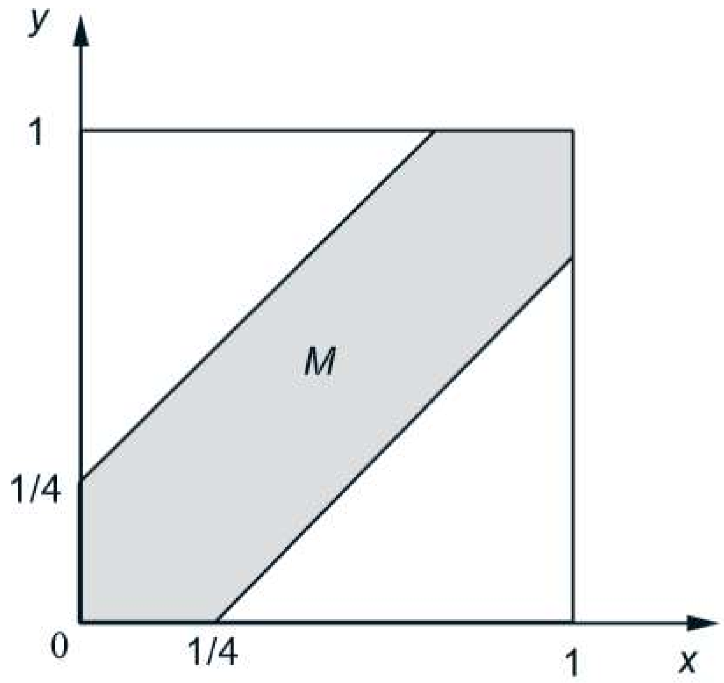
- Romeo and Juliet have a date at a given time, and each will arrive at the meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely.
- The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived. What is the probability that they will meet?
- Let us use as sample space the square $\Omega = [\mathbf{0},\mathbf{1}] \times [\mathbf{0},\mathbf{1}]$, whose elements are the possible pairs of delays for the two of them.
- Our interpretation of "equally likely" pairs of delays is to let the probability of a subset of Ω be equal to its area.
- This probability law satisfies the three probability axioms.
- What is the event that Romeo and Juliet will meet ?

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Continuous

- The event that Romeo and Juliet will meet is the shaded region

Probability is calculated to be $\frac{7}{16}$



- M = { (x,y) | $|x-y| \le \frac{1}{4}$, $0 \le x \le 1$, $0 \le y \le 1$,and is
- The area of **M** is 1 minus the area of the two unshaded triangles, or $1 - (\frac{3}{4}) * (\frac{3}{4}) = \frac{7}{16}$
- Thus, the probability of meeting is $\frac{7}{16}$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Probabilistic Models – Properties of Probability Laws

- Some Properties of Probability Laws Consider a probability law, and let **A**, **B**, and **C** be events.
- (a) If $\mathbf{A} \subset \mathbf{B}$, then $\mathbf{P}(\mathbf{A}) \leq \mathbf{P}(\mathbf{B})$.
- (b) $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{A} \cap \mathbf{B})$.
- (c) $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) \leq \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B})$.
- (d) $\mathbf{P}(\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{A}^c \cap \mathbf{B}) + \mathbf{P}(\mathbf{A}^c \cap \mathbf{B}^c \cap \mathbf{C})$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

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? In more precise terms, given an experiment, a corresponding sample space, and a probability law, suppose that we know that the outcome is within some given event **B**.

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Conditional Probability

- We wish to quantify the likelihood that the outcome also belongs to some other given event A. We thus seek to 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 $\mathbf{P}(\mathbf{A} \mid \mathbf{B})$
- For example, suppose that all six possible outcomes of a fair die roll are equally likely. If we are told that the outcome is even, we are left with only three possible outcomes, namely, 2, 4, and 6. These three outcomes were equally likely to start with, and so they should remain equally likely given the additional knowledge that the outcome was even. Thus, it is reasonable to let

$$\mathbf{P}(\text{the outcome is } 6 \mid \text{the outcome is even}) = \frac{1}{3}$$

$$\mathbf{P}(\mathbf{A} \mid \mathbf{B}) = \frac{\text{number of elements of } \mathbf{A} \cap \mathbf{B}}{\text{number of elements of } \mathbf{B}}$$

We introduce the following definition of conditional probability:

$$\mathrm{P}(\mathbf{A} \mid \mathbf{B}) = \frac{\mathbf{P}(\mathbf{A} \cap \mathbf{B})}{\mathbf{P}(\mathbf{B})}, \text{ where we assume that } \mathbf{P}(\mathbf{B}) > 0$$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Conditional Probability

❖ We toss a fair coin three successive times. We wish to find the conditional probability **P**(**A** | **B**) when **A** and **B** are the events

A = { **more heads than tails come up** }, **B** = { **1st toss is a head** }

The sample space consists of eight sequences Ω = { **HHH**, **HHT**, **HTH**, **HTT**, **THH**, **THT**, **TTH**, **TTT**}
which we assume to be equally likely

The event **B** consists of the four elements **HHH**, **HHT**, **HTH**, **HTT**, so its probability is $\mathbf{P}(B) = \frac{4}{8}$

The event **A∩B** consists of the three elements outcomes HHH, HHT, HTH, so its probability is $\mathbf{P}(\mathbf{A}\cap\mathbf{B}) = \frac{3}{8}$

Thus, the conditional probability **P**(A | B) is $\mathbf{P}(\mathbf{A} \mid \mathbf{B}) = \frac{\mathbf{P}(\mathbf{A}\cap\mathbf{B})}{\mathbf{P}(\mathbf{B})} = \frac{3/8}{4/8} = \frac{3}{4}$

We can bypass the calculation of **P**(**B**) and **P**(**A∩B**), and simply divide the number of elements shared by **A** and **B** (which is 3) with the number of elements of **B** (which is 4), to obtain the same result $\frac{3}{4}$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

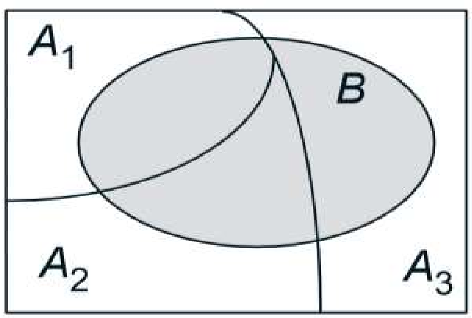
Total Probability Theorem and Bayes' Rule

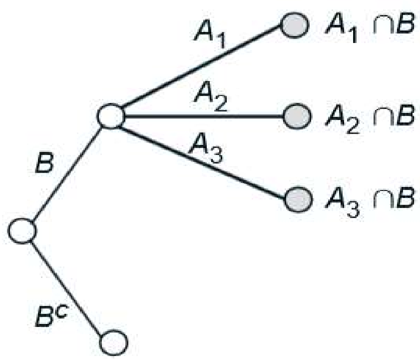
❖ We start with the following theorem, which is often useful for computing the probabilities of various events, using a "divide-and-conquer" approach

Total Probability Theorem: Let $\mathbf{A_1, \ldots, A_n}$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $\mathbf{A_1, \ldots, A_n}$) and assume that $\mathbf{P(A_i)} > 0$, for all $\mathbf{i = 1, \ldots, n}$

Then, for any event B, we have, $\mathbf{P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)}$

$$= \mathbf{P(A_1)\, P(B \mid A_1) + \cdots + P(A_n)\, P(B \mid A_n)}.$$





Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Total Probability Theorem and Bayes' Rule

- You enter a chess tournament where your probability of winning a game is **0.3** against half the players (call them type 1), **0.4** against a quarter of the players (call them type 2), and **0.5** against the remaining quarter of the players (call them type 3).
- You play a game against a randomly chosen opponent. What is the probability of winning? Let $\mathbf{A_i}$ be the event of playing with an opponent of type **i**.

We have $\mathbf{P}(\mathbf{A_1}) = 0.5$ $\mathbf{P}(\mathbf{A_2}) = 0.25$ $\mathbf{P}(\mathbf{A_3}) = 0.25$

Let also **B** be the event of winning. We have $\mathbf{P}(\mathbf{B} \mid \mathbf{A_1}) = 0.3$, $\mathbf{P}(\mathbf{B} \mid \mathbf{A_2}) = 0.4$, $\mathbf{P}(\mathbf{B} \mid \mathbf{A_3}) = 0.5$

Thus, by the **total probability theorem**, the probability of winning is

$$\boxed{\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{A_1})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_1}) + \mathbf{P}(\mathbf{A_2})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_2}) + \mathbf{P}(\mathbf{A_3})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_3})}$$

$$= 0.5*0.3 + 0.25*0.4 + 0.25*0.5 = \mathbf{0.375}$$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Total Probability Theorem and Bayes' Rule

- The total probability theorem is often used in conjunction with the following celebrated theorem, which relates conditional probabilities of the form $\mathbf{P(A\,|\,B)}$ with conditional probabilities of the form $\mathbf{P(B\,|\,A)}$, in which the order of the conditioning is reversed.

Bayes' Rule : Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i.

Then, for any event B such that $P(B) > 0$, we have

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

$$= \frac{P(A_i)\,P(B\,|\,A_i)}{P(A_1)P(B\,|\,A_1) + \cdots + P(A_n)\,P(B\,|\,A_n)}$$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Total Probability Theorem and Bayes' Rule

- You enter a chess tournament where your probability of winning a game is **0.3** against half the players (call them **type 1**), **0.4** against a quarter of the players (call them **type 2**), and **0.5** against the remaining quarter of the players (call them **type 3**).
- You play a game against a randomly chosen opponent. What is the probability of winning? Let $\mathbf{A_i}$ be the event of playing with an opponent of type **i**.

We have $\mathbf{P}(\mathbf{A_1}) = 0.5$ $\mathbf{P}(\mathbf{A_2}) = 0.25$ $\mathbf{P}(\mathbf{A_3}) = 0.25$

Let also **B** be the event of winning. We have $\mathbf{P}(\mathbf{B} \mid \mathbf{A_1}) = 0.3$, $\mathbf{P}(\mathbf{B} \mid \mathbf{A_2}) = 0.4$, $\mathbf{P}(\mathbf{B} \mid \mathbf{A_3}) = 0.5$

Suppose that you win. What is the probability $\mathbf{P}(\mathbf{A_1} \mid \mathbf{B})$ that you had an opponent of type 1 ?

Using Bayes' rule, we have

$$\boxed{\mathbf{P}(\mathbf{A_1} \mid \mathbf{B}) = \frac{\mathbf{P}(\mathbf{A_1})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_1})}{\mathbf{P}(\mathbf{A_1})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_1}) + \mathbf{P}(\mathbf{A_2})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_2}) + \mathbf{P}(\mathbf{A_3})\,\mathbf{P}(\mathbf{B} \mid \mathbf{A_3})}} = \frac{0.5*0.3}{0.5*0.3 + 0.25*0.4 + 0.25*0.5} = 0.4$$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Independence

- We have introduced the conditional probability $\mathbf{P(A \mid B)}$ 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,.

$$\boxed{\text{i.e., } \mathbf{P(A \mid B) = P(A)}}$$

- When the above equality holds, we say that **A** is independent of **B**
- Note that by the definition $\mathbf{P(A \mid B)} = \dfrac{\mathbf{P(A \cap B)}}{\mathbf{P(B)}}$
- This is equivalent to $\mathbf{P(A \cap B) = P(A)\ P(B)}$

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Independence

- Consider an experiment involving two successive rolls of a 4-sided die in which all 16 possible outcomes are equally likely and have probability 1/16
- (a) Are the events $\mathbf{A_i}$ = { 1st roll results in I }, $\mathbf{B_j}$ = { 2nd roll results in j }, independent ?

$$\mathbf{P}(\mathbf{A} \cap \mathbf{B}) = \mathbf{P}(\text{ the result of the two rolls is } (i,j)\ ) = \frac{1}{16}$$

$$\mathbf{P}(\mathbf{A_i}) = \frac{\text{number of elements of } \mathbf{A_i}}{\text{total number of possible outcomes}} = \frac{4}{16}$$

$$\mathbf{P}(\mathbf{B_j}) = \frac{\text{number of elements of } \mathbf{B_i}}{\text{total number of possible outcomes}} = \frac{4}{16}$$

We observe that $\mathbf{P}(\mathbf{A_i} \cap \mathbf{B_j}) = \mathbf{P}(\mathbf{A_i})\ \mathbf{P}(\mathbf{B_j})$, and the independence of $\mathbf{A_i}$ and $\mathbf{B_j}$ is verified

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Independence

- **(b)** Are the events **A** = { 1st roll is a 1 }, **B** = { sum of the two rolls is a 5 }, independent?
- The answer here is not quite obvious.
- We have **P**(**A** ∩ **B**) = **P** the result of the two rolls is (1,4)= $\frac{1}{16}$, and also
- P(A) = number of elements of **A** total number of possible outcomes = $\frac{4}{16}$
- The event **B** consists of the outcomes (1,4), (2,3), (3,2), and (4,1), and

$$\mathbf{P}(\mathbf{B}) = \frac{\text{number of elements of B}}{\text{total number of possible outcomes}} = \frac{4}{16}$$

Thus, we see that **P**(**A** ∩ **B**) = **P**(**A**) * **P**(**B**), and the events **A** and **B** are independent

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Independence

- **(c)** Are the events **A** = { maximum of the two rolls is 2 }, **B** = { minimum of the two rolls is 2 }, independent ?
- Intuitively, the answer is "no" because the minimum of the two rolls tells us something about the maximum. For example, if the minimum is 2, the maximum cannot be 1
- More precisely, to verify that **A** and **B** are not independent, we calculate **P**(**A** ∩ **B**) = **P** the result of the two rolls is (2,2) = $\frac{1}{16}$

$$\mathbf{P}(\mathbf{A}) = \frac{\text{number of elements of } \mathbf{A}}{\text{total number of possible outcomes}} = \frac{3}{16}$$

$$\mathbf{P}(\mathbf{B}) = \frac{\text{number of elements of } \mathbf{B}}{\text{total number of possible outcomes}} = \frac{5}{16}$$

We have **P**(**A**) * **P**(B) = $\frac{15}{(16)^2}$, so that **P**(**A** ∩ **B**) ≠ **P**(**A**) * **P**(**B**), and **A** and **B** are not independent

Dr. Mohammed Javed
Assistant Professor, Department of IT
Indian Institute of Information Technology (IIIT-Allahabad)
Email: javed@iiita.ac.in

Thank You