EECS 126 Lecture Notes

Dun-Ming Huang

Contents

1 Introduction, Probability Spaces, Conditional Probability Theorem **3**

1.1 Set Operation 3

1.2 Probabilistic Model 5

1.3 Conditional Probability 6

2 Law of Total Probability, Bayes' Rule, Independence **8**

2.1 Partial Information 8

2.2 Total Probability and Bayes' Rule 9

2.3 Independence 10

Chapter 1

Introduction, Probability Spaces, Conditional Probability Theorem

Course Policies

As always, first lecture is mostly about course policies.

- Look over syllabus and course website here
- Check Ed and Gradescope
- Industrial revolution of generative AI :0

Textbook doesn't cover all lecture contents.

Concept of Probability

The concept of probability theory began with the concept of frequency, where, a specific outcome occurs at a certain frequency that denotes the probability of that outcome in some sequence of experiments. For many applications, this stays true. However, in other applications where the notion of frequency does not apply (eg., if a presidential candidate wins at a probability of 50%, does that candidate win on the frequency of 0.5 as well?). Consequently, we find a different mathematical formalism to illustrate the notion of probability. And, with the diversity of interpretations, many frameworks exist for the study of probability. In this course, via exploring numerous examples at which probability theories are applied and investigated, we develop our intuition toward the content of probability theory.

1.1 Set Operation

1.1.1 Definition of Sets

Let us begin with the definition of sets.

Definition 1.1.1. Set

A **set** is a collection of objects. The objects contained in a set is called the **element** of that set.

Sets are usually denoted using capital letters; for example, $\mathcal{A}$.

1.1.2 Element-wise Operations

Suppose that an element x belongs to a set $\mathcal{X}$, we denote it as follows:

Symbol 1.1.1. Invlusion

The expression:

$$x \in \mathcal{X}$$

denotes that the element x belongs to a set $\mathcal{X}$.
Meanwhile, the expression:

$$x \notin \mathcal{X}$$

denotes that the elemnt x does not belong to $\mathcal{X}$.

Usually, sets are written in the following format:

$$\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$$

In the case where the set contains infinite elements, we may denote as follows:

$$\mathcal{X} = \{x_1, x_2, \ldots\}$$

To build a set conditionally, we may use a set-builder notation:

Symbol 1.1.2. Set-Builder Notation

$$\{x | x \text{ satisfies predicate } p\}$$

For example, the set of all integers that are divisible is denoted as:

$$\{2k | k \in \mathbb{Z}\}$$

1.1.3 Set-Set Relations

Sets can be affiliated with each other:

Definition 1.1.2. Subsets

If every element of some set $\mathcal{X}$ is also an element of set $\mathcal{Y}$, we state that $\mathcal{X}$ is a subset of $\mathcal{Y}$:

$$\mathcal{X} \subseteq \mathcal{Y}$$

Definition 1.1.3. Equality of Sets

Sets $\mathcal{X}$ and $\mathcal{Y}$ are equal if $\mathcal{X} \subseteq \mathcal{Y} \wedge \mathcal{Y} \subseteq \mathcal{X}$.

Last but not least, let us discuss the notion of universality in set theory:

Definition 1.1.4. Universal Set

A universal set Ω is the set of all elements that will be related to the application of discussion in mathematics.

On the other hand, empty sets denotes an empty set with no elements. By definition, the empty set is a subset of every set.

Sets can also be related to each other based on their elemental differences.

Definition 1.1.5. Complements

The complement of a set $\mathcal{X}$ is defined as follows:

$$\mathcal{X}^C = \{x | x \in \omega \wedge x \notin \mathcal{X}\}$$

On the other hand, the notations “union” and “intersection” are defined as taught before.

$$\mathcal{X} \cup \mathcal{Y} = \{x | x \in \mathcal{X} \vee x \in \mathcal{Y}\} \qquad \text{(Union)}$$
$$\mathcal{X} \cap \mathcal{Y} = \{x | x \in \mathcal{X} \wedge x \in \mathcal{Y}\} \qquad \text{(Intersection)}$$

By definition, then,

$$(S^C)^C = S$$
$$S \cup \Omega = \Omega$$

1.1.4 Algebraic Rules of Set Arithmetics

There exists commutativeness in the algebra of sets:

$$\forall S, T : S \cup T = T \cup S$$

There also exists associativeness in the algebra of sets:

$$\forall S, T, M : S \cap (T \cup M) = (S \cap T) \cup (S \cap M)$$

And, there’s the famous DeMorgan’s Law:

$$\Big(\bigcup_n S_n\Big)^C = \bigcap_n S_n^C$$
$$\Big(\bigcap_n S_n\Big)^C = \bigcup_n S_n^C$$

1.2 Probabilistic Model

There are two components to a probabilistic model:

1. A sample space Ω that refers to the set of all possible outcomes of an experiment.
2. A probability law that assigns nonegative number $\mathbb{P}(A)$ to each subset $A \subseteq \Omega$ that satisfies Kolmogorov’s axioms.

Let us proceed with a probabilistic model of dice rolls:

Example 1.2.1: Probabilistic Model of Dice Rolls

The sample space of a dice roll may be:

$$\Omega = \{1, 2, \ldots, 6\}$$

where the probability law for a fair dice works to be, for example,

$$\mathbb{P}(\{1\}) = \frac{1}{6}$$

All probabilistic models follow the following Kolmogorov’s axioms:

Theorem 1.2.1. Kolmogorov's Axioms

1. $\mathbb{P}(A) \geq 0$
2. Countable-Additivity: If A and B are disjoint events, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
3. $\mathbb{P}(\Omega) = 1$

Among which, we may expand countable-additivity into the following condition:

For a sequence of disjoint, countably infinite events $A_1, A_2, \ldots,$

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

With these axioms in mind, let us consider an alterantive example regarding probabilistic model of dices:

Example 1.2.2: Probabilistic Model of Pairs of Dice Rolls

Let all possible results from one single four-sided dice rolls be represented by:

$$\omega = \{1, 2, 3, 4\}$$

The sample space of a pair of four-sided dice rolls may be represented as:

$$\Omega = \{(i, j) | i \in \omega \wedge j \in \omega\}$$

Notably, this discrete sample space entertains a uniform probability for all possible outcome. In such case, the discrete uniform probbaility law states that, for some event A:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

We further derive the following proerties of probability laws as the consequence of Kolmogorov's Axioms:

1. $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
3. As implied from the previous rule: $\mathbb{P}(A) + \mathbb{P}(B) \geq \mathbb{P}(A \cap B)$
4. $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(A^C \cap B) + \mathbb{P}(A^C \cap B^C \cap C)$

1.3 Conditional Probability

Conditioning is the essence of many probabilistic applications, and primary method of constructing complicated probabilistic measures.

Definition 1.3.1. Conditional Probability

The probability of an event A provided the transpiration of event B is denoted as:

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

We may observe that if the event A is not independent from event B, then the probability we discuss now is fundamentally different from the sole probability of event A's occurrence.

For demonstration, let us state that A is the event where a fair dice throw presents value 3, and B is where an even value is presented. Then, the event A is dependent upon B such that, with the occurrence of B, event A shall enver occur. Therefore, $\mathbb{P}(A|B) = 0$.
On the other hand, let C be the event where value 4 is presented, then $\mathbb{P}(A|B) = \frac{1}{3}$.

Chapter 2

Law of Total Probability, Bayes' Rule, Independence

2.1 Partial Information

Partial information is the circumstance at which we determine the probability of an event provided the occurrence of another event. Recall from last lecture this example:

Definition 2.1.1. Conditional Probability

The probability of an event A provided the transpiration of event B is denoted as:

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

We may observe that if the event A is not independent from event B, then the probability we discuss now is fundamentally different from the sole probability of event A's occurrence.

For demonstration, let us state that A is the event where a fair dice throw presents value 3, and B is where an even value is presented. Then, the event A is dependent upon B such that, with the occurrence of B, event A shall enver occur. Therefore, $\mathbb{P}(A|B) = 0$.
On the other hand, let C be the event where value 4 is presented, then $\mathbb{P}(A|B) = \frac{1}{3}$.

In addition, provided the above example where we entertain a uniform discrete probability space, we find that,

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

We may further formulate the probability laws of a conditional probability circumstance:

Theorem 2.1.1. Conditional Probability Rules

The following three rules dictate the mechanisms of conditional probability:

1. $\forall A \in \Omega, \mathbb{P}(A|B) \geq 0$
2. Provided two disjoint events A_1, A_2:

$$\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)$$

3. $\forall B \in \Omega, \mathbb{P}(\Omega|B) = 1$

The derivation for rule 2 follows:

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2|B) &= \frac{\mathbb{P}((A_1 \cup A_2) \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}((A_1 \cap B) \cup (A_2 \cap B))}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}((A_1 \cap B)) + \mathbb{P}((A_2 \cap B))}{\mathbb{P}(B)} = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)\end{aligned}$$

The derivation for rule 3 follows:

$$\begin{aligned}\mathbb{P}(\Omega|B) &= \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1\end{aligned}$$

Meanwhile, from the formualtion of conditional probability, we also develop the Multiplication Rule:

Theorem 2.1.2. Multiplication Rule

The multiplication rule states that:

$$\begin{aligned}\mathbb{P}(\bigcap_{i\geq 1} A_i) &= \mathbb{P}(A_1) \prod_{i\geq 1} \mathbb{P}(A_{i+1}| \cap_{j=1}^{i} A_j) \\ &= \mathbb{P}(A_1) \prod_{i\geq 1} \frac{\mathbb{P}(\cap_{k=1}^{i+1} A_k)}{\mathbb{P}(\cap_{j=1}^{i} A_j)}\end{aligned}$$

2.2 Total Probability and Bayes' Rule

Let us begin our discussion with the Total Probability Theorem, used for handling an approach of divide and conquer computation for probabilities:

Theorem 2.2.1. Total Probbaility Theorem

Let us have disjoint events $A_1, A_2, \dots, A_m$ that form a partition of the sample space Ω. Assume that $\forall i \in \{1, \dots, m\} : \mathbb{P}(A_i) \geq 0$. Then, the Total Probability Theorem states that, for any event $B \in \Omega$:

$$\begin{aligned}\mathbb{P}(B) &= \sum_{i=1}^{m} \mathbb{P}(A_i \cap B) \\ &= \sum_{i=1}^{m} \mathbb{P}(A_i)\mathbb{P}(B|A_i)\end{aligned}$$

And, let us also introduce the famous probability rule:

Theorem 2.2.2. Bayes' Rule

The most primitive form of this theorem follows as

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

while the form of this theorem under intervention of Total Probability Theorem is phrased as:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\sum_{i=1}^{m} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}$$

2.3 Independence

The concept may be illustrated by the following context:

> For two same experiments held across different laboratories, the result in lab A should not influence the result in lab B: they are independent, without influence of each other.

In probability theory, there is a simple definition for independence of events:

Definition 2.3.1. Independence

Two events A, B are independent if

$$mathbbP(A|B) = \mathbb{P}(A)$$

There is a rigorous definition for independence that we will discuss later, which concerns the independence of sigma algebras isntead of independence of events.

What does independence contribute to probability theory? Or, what are some interesting properties of independence? For one simple instance, consider our prior definition of conditional probability:

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

Such that at independence of events,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

This shows that, the probabilities of A and B do not affect each other, as the probability of joint occurrences $A \cap B$ and $B \cap A$ are equal. In such sense, we state that independence is **symmetric** for events A and B.

As the concept is more sophisticated, let us discuss independence further with an example

Example 2.3.1: Demonstration of Independence in Dice Throws

Let us work with 2 rolls of a 4-sided die. Let events A_i denote the result of the first rolls, events B_i denote ther esult of second rolls. The probability $\mathbb{P}(A_i \cap B_j)$ is the probability at which the first roll results in value i, and second roll j, and the probability is simply $\frac{1}{4} \times \frac{1}{4}$.
More precisely,

$$\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j) = \frac{1}{16}$$

Now, let us work with a slightly different example:

$$\begin{cases} \text{Event A} & : \text{ The first roll of the dice is } 1 \\ \text{Event B} & : \text{ The sum of two rolls of the dice is } 5 \end{cases}$$

are the two events independent?

Let us inspect the probabilities $\mathbb{P}(A)$ and $\mathbb{P}(B)$:

$$\mathbb{P}(A) = \frac{1}{4}$$
$$\mathbb{P}(B) = \frac{4}{16} = \frac{1}{4}$$
$$\mathbb{P}(A \cap B) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

We arrive at the conclusion that

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

Therefore, events A and B are independent, even if the intuitions would say event B and A "*might*" be independent just because the sum of rolls involve the result of one roll.

Now, let us discuss the independence of many events.

Definition 2.3.2. Independence of Mant Events

Events $A_1, \ldots, A_m$ are independent if:

$$\forall S \subseteq \{1, \ldots, m\} : \mathbb{P}(\bigcap_{i \in S} A_i) = \prod_{i \in S} \mathbb{P}(A_i)$$