EC5.102 Information and Communication **Spring 2020**

Lecture 5: Probability, Conditional Probability and Bayes' Theorem

Instructor: Dr. Lalitha Vadlamani

A probabilistic model is a mathematical description of an uncertain situation. 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. An event space $\mathcal{F}$ is a collection of events satisfying the following non-empty and closure properties:
  - $\mathcal{F}$ is non-empty. In particular, it definitely contains the null set ϕ.
  - **Closure under complements:** If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
  - **Closure under countable unions:** If $A_1, A_2, \ldots \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.
- The probability law, which assigns to an event in the event space, a nonnegative number P(A) (called the probability of A) that encodes our knowledge or belief about the collective likelihood of the elements of A (See Fig. **??**).

Two examples of event space $\mathcal{F}$ for the die-throwing experiment are as follows:

- $\mathcal{F} =$ Power set of $\{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{F} = \{\phi, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$.

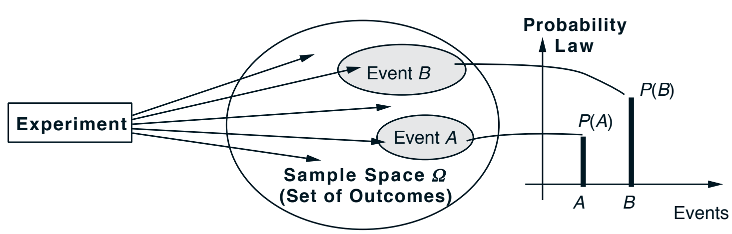

Figure 5.1: Illustrating an experiment, sample space, events and probability law.

There is no restriction on what constitutes an experiment. For example, it could be a single toss of a coin, or three tosses, or an infinite sequence of tosses. However, it is important to note that in our formulation of a probabilistic model, there is only one experiment. So, three tosses of a coin constitute a single experiment, rather than three experiments. The sample space of an experiment may consist of a finite or an infinite number of possible outcomes. An example of a sample space which has infinite outcomes is as follows: Consider throwing a dart on a square target and viewing the point of impact as the outcome.

5.1 Axioms of Probability

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

$$P(A \cup B) = P(A) + P(B). \tag{5.1}$$

   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

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots. \tag{5.2}$$

3. **Normalization:** The probability of the entire sample space is equal to 1, i.e., $P(\Omega) = 1$.

5.2 Properties Derived from Axioms of Probability

1. $P(A^c) = 1 - P(A)$. Follows because A and A^c are disjoint and $A \cup A^c = \Omega$ and $P(\Omega) = 1$.
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Writing $A \cup B$ as the union of disjoint events $A \cap B$, $A^c \cap B$ and $A \cap B^c$, we have

$$P(A \cup B) = P(A \cap B) + P(A^c \cap B) + P(A \cap B^c). \tag{5.3}$$

   Also, A can be written as union of disjoint events $A \cap B$ and $A \cap B^c$. Hence, we have

$$P(A) = P(A \cap B) + P(A \cap B^c). \tag{5.4}$$

   Similarly, we have

$$P(B) = P(A \cap B) + P(A^c \cap B). \tag{5.5}$$

   Combining the above three equations, we have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. From the above property, we have that $P(A \cup B) \leq P(A) + P(B)$. Repeatedly applying this result, we also have that

$$P(A_1 \cup A_2 \cup \ldots A_n) \leq \sum_{i=1}^{n} P(A_i). \tag{5.6}$$

5.3 Conditional Probability

Given an experiment, a corresponding sample space, and a probability law, suppose that we know that the outcome is within some given event B. 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 $P(A|B)$. We would like the conditional probabilities $P(A|B)$ of different events A to constitute a valid probability law, that satisfies the probability axioms.

The conditional probability of an event A given B is defined as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{5.7}$$

where we assume that $P(B) > 0$.

For a fixed event B, it can be verified that the conditional probabilities $P(A|B)$ form a valid probability law that satisfies the three axioms. Nonnegativity is clear. Furthermore,

$$P(\Omega|B) = \frac{P(B)}{P(B)} = 1. \tag{5.8}$$

To verify the additivity axiom, we write for any two disjoint events A_1 and A_2,

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

where we used the fact that $A_1 \cap B$ and $A_2 \cap B$ are disjoint events.

5.3.1 Independent Events

Two events A and B are said to be independent if

$$P(A \cap B) = P(A)P(B). \tag{5.9}$$

In terms of conditional probabilities the above condition can be rewritten as

$$P(A|B) = P(A). \tag{5.10}$$

Example 5.1. *Suppose we toss two fair dice. Let A denote the event that the sum of the dice is 6 and B denote the event that first die equals 4. Then*

$$P(A \cap B) = P(\{4, 2\}) = \frac{1}{36}.$$

while

$$P(A)P(B) = \frac{5}{36}\frac{1}{6} = \frac{5}{216}.$$

Hence, A and B are not independent. Intutitively, the reason is if the first die is a 4, then we still have a possibility of getting a total of 6 whereas if the first die landed 6, then we cannot have a sum of 6. Hence, the probability that the sum is 6 depends on the outcome of the first die and hence the two events are not independent.

Now, let C denote the event that the sum of the dice is 7. Then,

$$P(B \cap C) = P(\{4, 3\}) = \frac{1}{36}.$$

And

$$P(B)P(C) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}.$$

Hence, B and C are independent. Please try to give intuitive reasoning for the same.

5.4 Total Probability Theorem and Bayes' Theorem

Theorem 5.2 (Total Probability Theorem)**.** *Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $A_1, \ldots, A_n$) and assume that $P(A_i) > 0$, for all $i = 1, \ldots, n$. Then, for any event B, we have*

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \tag{5.11}$$

This is because

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

Theorem 5.3 (Bayes' Theorem)**.** *Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $A_1, \ldots, A_n$) and assume that $P(A_i) > 0$, for all $i = 1, \ldots, n$. Then, for any event B such that $P(B) > 0$, we have*

$$\begin{aligned} P(A_i|B) &= \frac{P(B|A_i)P(A_i)}{P(B)} \\ &= \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}. \end{aligned}$$

To verify Bayes theorem, note that $P(A_i)P(B|A_i)$ and $P(A_i|B)P(B)$ are equal, because they are both equal to $P(A_i \cap B)$. This yields the first equality. The second equality follows from the first by using the total probability theorem to rewrite $P(B)$.

Example 5.4. *A diagnostic test has a probability).95 of giving a positive result when applied to a person suffering from a certain disease, and a probability* 0.10 *of giving a (false) positive when applied to a non-sufferer. It is estimated that* 0.5% *of the population are sufferers. Suppose that the test is now administered to a person about whom we have no relevant information relating to the disease (part from the fact that he/she comes from this population). Calculate the following probabilities:*

1. *that the test result will be positive*
2. *that, given a positive result, the person is a sufferer*
3. *that, given a negative result, the person is a sufferer*
4. *that the person will be misclassified.*

Let T denote the event that test is positive and S denote the event that administered person is a sufferer. Then the probabilities specified are $P(T|S) = 0.95, P(T|S^c) = 0.1, P(S) = 0.005$. The following probabilities need to be calculated in each of the parts above: 1. $P(T)$ 2. $P(S|T)$ 3. $P(S^c|T^c)$ 4. $P(T \cap S^c) + P(T^c \cap S)$.

Example 5.5 (Monty Hall Problem)**.** *In a TV game show, a contestant selects one of three doors; behind one of the doors there is a prize, and behind the other two there are no prizes. After the contestant selects a door, the game-show host opens one of the remaining doors, and reveals that there is no prize behind it.*

The host then asks the contestant whether they want to SWITCH their choice to the other unopened door, or STICK to their original choice. Is it probabilistically advantageous for the contestant to SWITCH doors, or is the probability of winning the prize the same whether they STICK or SWITCH? (Assume that the host selects a door to open, from those available, with equal probability).

Solution: Assume without loss of generality that contestant picked door 1. Let C_i be the event that prize is behind door i. Also, let H_i denote the event that host opens door i. Then the following probabilities can be calculated:

$$P(H_2|C_1) = \frac{1}{2}, P(H_2|C_2) = 0, P(H_2|C_3) = 1$$

Hence, based on total probability theorem (assuming $P(C_i) = \frac{1}{3}$), we have $P(H_2) = \frac{1}{2}$. Hence, by applying Bayes theorem we have, $P(C_1|H_2) = \frac{1}{3}$ and $P(C_3|H_2) = \frac{2}{3}$. Hence, it is advantageous to SWITCH.