

# 1 Conditional Probability

*Chapter abstract:* This chapter is a foundational concept in probability theory that allows us to assess the likelihood of one event occurring given that another event has already happened. This chapter delves into the key principles and applications of conditional probability. We begin by exploring the fundamental difference between ordinary probability and conditional probability, emphasizing the importance of contextual dependence.

## 1.1 CONDITIONAL PROBABILITY

In our previous lecture, we explored the fundamental concepts of probability theory. We established the core axioms that govern probabilities and their applications in calculating the likelihood of events. However, the world around us is rarely as simple as independent events happening in isolation. Often, the likelihood of one event occurring is influenced by the knowledge of another event having already happened

### Example 1.

Consider a sample space that includes three disjoint events,  $A$ ,  $B$ , and  $C$ , as illustrated in Figure 1.1. Suppose we receive new information confirming that event  $A$  has indeed occurred, but we do not have any updates about events  $B$  and  $C$ . How does this additional information about event  $A$  affect our original probability model?

Since events  $A$ ,  $B$ , and  $C$  are disjoint (as shown<sup>1</sup> in Figure 1.1), they cannot happen at the same time. This means that if event  $A$  occurs, events  $B$  and  $C$  cannot occur. Events  $B$  and  $C$  together form the complement of event  $A$ , denoted as  $B \cup C = A^c$ . Therefore, if  $A$  is observed,  $A^c$  cannot be observed, i.e.

$$\mathbb{P}(A^c) = \mathbb{P}(B \cup C) \stackrel{(A3)}{=} \mathbb{P}(B) + \mathbb{P}(C) = 0$$

If and only if  $\mathbb{P}(B) = \mathbb{P}(C) = 0$ .

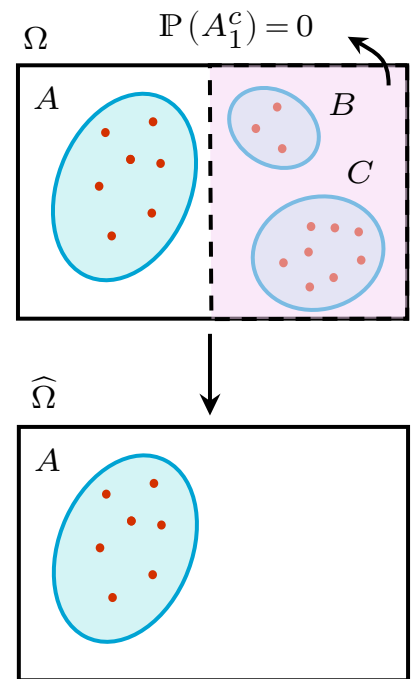
In this modified sample space, the probability distribution is also updated. The probability of event  $A$  occurring is now 1, which makes it a new sample space by axiom (A2), i.e.  $\hat{\Omega} \triangleq A$ , since it is the only event in the new sample space.

### Example 2.

Consider a sample space that includes three events,  $A$ ,  $B$ , and  $C$ , of which event  $C$  is disjoint from events  $A$  and  $B$ , but events  $A$  and  $B$  are not disjoint as illustrated in Figure 1.2. Suppose we receive new information confirming that event  $A$  has indeed occurred, but we do not have any updates about events  $B$  and  $C$ . How does this additional information about event  $A$  affect our original probability model?

### Outline

1.1	Conditional Probability . .	1
1.2	Die Roll Example . . . . .	3
1.3	Properties of Conditional Probability . . . . .	3



**Figure 1.1:** Visual representation of the conditional probability of event  $B$  and  $C$  given event  $A$ . When  $A$  occurs, the probability of  $B$  and  $C$  also occurring is zero, since the three events are mutually exclusive.

When we confirm that event  $A$  has occurred, we must revise our probability model to reflect this new information as done in the previous example. Since event  $C$  is mutually exclusive with events  $A$  and  $B$ , we can immediately conclude that event  $C$  can not occur, i.e.  $\mathbb{P}(C) = 0$ .

Event<sup>2</sup>  $B$  can be broken down into two parts: the region where  $B$  intersects with  $A$  (the green region in Figure 1.2), and the region where  $B$  is observed without  $A$ .

$$B = (A \cap B) \cup (A^c \cap B)$$

However, since event  $A$  is assumed to be observed, the latter region is no longer possible, i.e.  $\mathbb{P}(A^c \cap B) = 0$ . Therefore, the probability of event  $B$  is reduced to

$$\mathbb{P}(B) \stackrel{(A3)}{=} \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \mathbb{P}(A \cap B)$$

Now, we need to establish the relationship between the probability of event  $A$  and the probability of the intersection of events  $A$  and  $B$ , denoted as  $A \cap B$ . In the revised model, event  $A$  becomes the new sample space, and accordingly we denote it by  $\hat{\Omega}$ . Moreover, we denote the intersection of events  $A$  and  $B$  as  $\hat{B}$  in the revised model. The problem is now equivalent to calculating the probability of  $\hat{B}$  within the new sample space  $\hat{\Omega}$ . This is nothing more than a classic demonstration of the discrete uniform law:

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

This ratio represents the probability of event  $B$  occurring given that event  $A$  has occurred. We define this as the conditional probability of  $B$  conditioned on  $A$ .

#### Definition 1: Conditional Probability

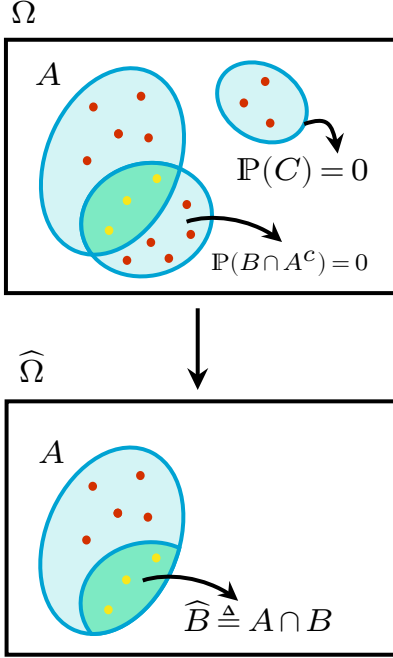
Let  $A, B \subseteq \Omega$ , the conditional probability of an event  $B$  conditioned on an event  $A$  is given by

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

Referring to Figure 1.2, we can observe that  $|A \cap B| = 3$  and  $|A| = 8$ . Using these values, we can calculate the conditional probability of  $B$  given  $A$  as:

$$\begin{aligned} \mathbb{P}(B | A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \\ &= \frac{|A \cap B|}{|A|} \\ &= \frac{3}{8} \end{aligned}$$

2



**Figure 1.2:** Visual representation of the conditional probability of event  $B$  given event  $A$ . When  $A$  occurs, the probability of  $B$  also occurring is equivalent to the intersection between  $A$  and  $B$ .

## 1.2 DIE ROLL EXAMPLE

We consider a simple example where we utilize the conditional probability.

### Example 3.

Consider two four-faced dice. Let  $B$  denote the event that one of the two die resulted in a value of 2, and the other die resulted in a value of at least 2. Moreover, define  $M \triangleq \max(X, Y)$ . What is the probability that  $M = 3$ , where one of the dices has an attained value of 2?

In this example, there are 16 possible outcomes, all equally likely. We define event  $B$  as

$$B \triangleq \{(x, y) \in X \times Y \mid \min(x, y) = 2\}$$

meaning one die shows a value of 2 and the other shows a value of at least 2. Figure 1.3 illustrates<sup>3</sup> these scenarios, we can see that event  $B$  consists of the outcomes

$$B = \{(2, 2), (2, 3), (2, 4), (3, 2), (4, 2)\}$$

Now, we want to find the probability that  $M = 3$  (i.e., the maximum value is 3) given that event  $B$  has occurred. Out of the 5 scenarios for event  $B$ , only 2 of them result in an observation for  $M$ :

$$M = \{(2, 3), (3, 2)\}$$

Therefore, the conditional probability is calculated as

$$\mathbb{P}(M = 3 \mid B) = \frac{2}{5}.$$

Alternatively, we can directly apply the conditional probability formula to reach this result. In this case, we have:

$$\mathbb{P}(M = 3 \mid B) = \frac{\mathbb{P}((M = 3) \cap B)}{\mathbb{P}(B)} = \frac{2/16}{5/16} = \frac{2}{5}$$

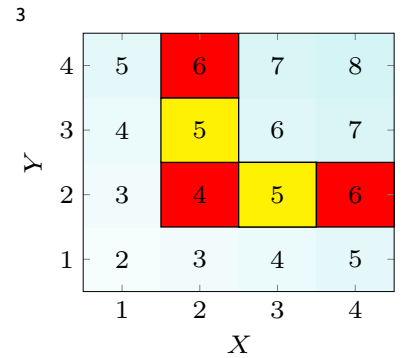
### Remark 1.

Conditional probabilities are same as ordinary probabilities, but applied to different situations. Intrinsically speaking, they obey the standard probability axioms.

The only crucial difference lies in the consideration of additional information. Ordinary probability treats all events independently, while conditional probability takes into account the knowledge that another event has already transpired. This context-dependence allows conditional probability to capture more nuanced relationships between events.

## 1.3 PROPERTIES OF CONDITIONAL PROBABILITY

Conditional probability is a fundamental concept in probability theory that allows us to update our beliefs about the likelihood of events based on new information. In this section, we will expand our knowledge on conditional probabilities by developing some basic properties.



**Figure 1.3:** The sample space in this problem is reduced upon the occurrence of the event  $B$  (highlighted in red).

We begin by emphasizing that conditional probability adheres to the same basic principles as ordinary probability, namely the probability axioms.

**Proposition 1.**

Conditional probability satisfies the probability axioms.

**Proof.** We shall demonstrate this for each axiom:

(A1)  $\mathbb{P}(A | B) \geq 0$ , since

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B) \geq 0}{\mathbb{P}(B) > 0} \geq 0$$

(A2)  $\mathbb{P}(\Omega | B) = 1$ , since

$$\mathbb{P}(\Omega | B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

(A3) Assume  $A \cap C = \emptyset$ , then

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

Note that the pair of  $A$  that belongs in  $B$  is also disjoint from the part of  $C$  that belongs to  $B$ .

Thus, this ensures that our calculations and interpretations remain consistent within the framework of probability theory.