

Chapter 4

Conditional Probability

Conditional probability provides a way to compute the likelihood of an event based on *partial information*. This is a powerful concept that is used extensively throughout engineering with applications to decision making, networks, communications and many other fields.

4.1 Conditioning on Events

We begin our description of conditional probability with illustrative examples. The intuition gained through this exercise is then generalized by introducing a formal definition for this important concept.

Example 22. *The rolling of a fair die is an experiment with six equally likely outcomes. As such, the probability of obtaining any of the outcomes is 1/6. However, if we are told that the upper face features an odd number, then only three possibilities remain, namely $\{1, 3, 5\}$. These three outcomes had equal probabilities before the additional information was revealed. It then seems natural to assume that they remain equally likely afterwards. In particular, it is reasonable to assign a probability of 1/3 to each of the three outcomes that remain possible candidates after receiving the side information. We can express the probability of getting a three given that the outcome is an odd number as*

$$\frac{\Pr(3 \cap \{1, 3, 5\})}{\Pr(\{1, 3, 5\})} = \frac{\Pr(3)}{\Pr(\{1, 3, 5\})} = \frac{1}{3}.$$

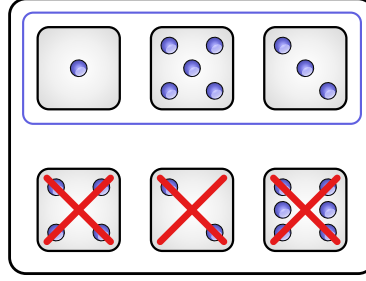


Figure 4.1: Partial information about the outcome of an experiment may change the likelihood of events. The resulting values are known as conditional probabilities.

Example 23. Let A and B be events associated with a random experiment, and assume that $\Pr(B) > 0$. To gain additional insight into conditional probability, we consider the scenario where this experiment is repeated N times. Let N_{AB} be the number of trials for which $A \cap B$ occurs, $N_{A\bar{B}}$ be the number of times where only A occurs, $N_{\bar{A}B}$ be the number of times where only B occurs, and $N_{\bar{A}\bar{B}}$ be the number of trials for which neither takes place. From these definitions, we gather that A is observed exactly $N_A = N_{AB} + N_{A\bar{B}}$ times, and B is seen $N_B = N_{AB} + N_{\bar{A}B}$ times.

The frequentist view of probability is based on the fact that, as N becomes large, one can approximate the probability of an event by taking the ratio of the number of times this event occurs over the total number of trials. For instance, we can write

$$\Pr(A \cap B) \approx \frac{N_{AB}}{N} \qquad \Pr(B) \approx \frac{N_B}{N}.$$

Likewise, the conditional probability of A given knowledge that the outcome lies in B can be computed using

$$\Pr(A|B) \approx \frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} \approx \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (4.1)$$

As N approaches infinity, these approximations become exact and (4.1) unveils the formula for conditional probability.

Having considered intuitive arguments, we turn to the mathematical definition of conditional probability. Let B be an event such that $\Pr(B) > 0$.

A conditional probability law assigns to every event A a number $\Pr(A|B)$, termed the *conditional probability of A given B* , such that

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (4.2)$$

We can show that the collection of conditional probabilities $\{\Pr(A|B)\}$ specifies a valid probability law, as defined in Section 3.2. For every event A , we have

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \geq 0$$

and, hence, $\Pr(A|B)$ is nonnegative. The probability of the entire sample space Ω is equal to

$$\Pr(\Omega|B) = \frac{\Pr(\Omega \cap B)}{\Pr(B)} = \frac{\Pr(B)}{\Pr(B)} = 1.$$

If A_1, A_2, \dots is a sequence of disjoint events, then

$$A_1 \cap B, A_2 \cap B, \dots$$

is also a sequence of disjoint events and

$$\begin{aligned} \Pr\left(\bigcup_{k=1}^{\infty} A_k \middle| B\right) &= \frac{\Pr((\bigcup_{k=1}^{\infty} A_k) \cap B)}{\Pr(B)} = \frac{\Pr(\bigcup_{k=1}^{\infty} (A_k \cap B))}{\Pr(B)} \\ &= \sum_{k=1}^{\infty} \frac{\Pr(A_k \cap B)}{\Pr(B)} = \sum_{k=1}^{\infty} \Pr(A_k|B), \end{aligned}$$

where the third equality follows from the third axiom of probability applied to the set $\bigcup_{k=1}^{\infty} (A_k \cap B)$. Thus, the conditional probability law defined by (4.2) satisfies the three axioms of probability.

Example 24. *A fair coin is tossed repetitively until heads is observed. In Example 20, we found that the probability of observing heads for the first time on trial k is 2^{-k} . We now wish to compute the probability that heads occurred for the first time on the second trial given that it took an even number of tosses to observe heads. In this example, $A = \{2\}$ and B is the set of even numbers. The probability that the outcome is two, given that the number of tosses is even, is equal to*

$$\Pr(2|B) = \frac{\Pr(2 \cap B)}{\Pr(B)} = \frac{\Pr(2)}{\Pr(B)} = \frac{1/4}{1/3} = \frac{3}{4}.$$

In the above computation, we have used the fact that the probability of flipping the coin an even number of times is equal to $1/3$. This fact was established in Example 20.

The definition of conditional probability can be employed to compute the probability of several events occurring simultaneously. Let A_1, A_2, \dots, A_n be a collection of events. The probability of events A_1 through A_n taking place at the same time is given by

$$\Pr\left(\bigcap_{k=1}^n A_k\right) = \Pr(A_1) \Pr(A_2|A_1) \Pr(A_3|A_1 \cap A_2) \cdots \Pr\left(A_n \middle| \bigcap_{k=1}^{n-1} A_k\right). \quad (4.3)$$

This formula is known as the *chain rule of probability*, and it can be verified by expanding each of the conditional probabilities using (4.2),

$$\Pr\left(\bigcap_{k=1}^n A_k\right) = \Pr(A_1) \frac{\Pr(A_1 \cap A_2)}{\Pr(A_1)} \frac{\Pr(A_1 \cap A_2 \cap A_3)}{\Pr(A_1 \cap A_2)} \cdots \frac{\Pr(\bigcap_{k=1}^n A_k)}{\Pr(\bigcap_{k=1}^{n-1} A_k)}.$$

This latter expression implicitly assumes that $\Pr(\bigcap_{k=1}^{n-1} A_k) \neq 0$.

Example 25. An urn contains eight blue balls and four green balls. Three balls are drawn from this urn without replacement. We wish to compute the probability that all three balls are blue. The probability of drawing a blue ball the first time is equal to $8/12$. The probability of drawing a blue ball the second time given that the first ball is blue is $7/11$. Finally, the probability of drawing a blue ball the third time given that the first two balls are blue is $6/10$. Using (4.3), we can compute the probability of drawing three blue balls as

$$\Pr(bbb) = \frac{8}{12} \frac{7}{11} \frac{6}{10} = \frac{14}{55}.$$

4.2 The Total Probability Theorem

The probability of events A and B occurring at the same time can be calculated as a special case of (4.3). For two events, this computational formula simplifies to

$$\Pr(A \cap B) = \Pr(A|B) \Pr(B). \quad (4.4)$$

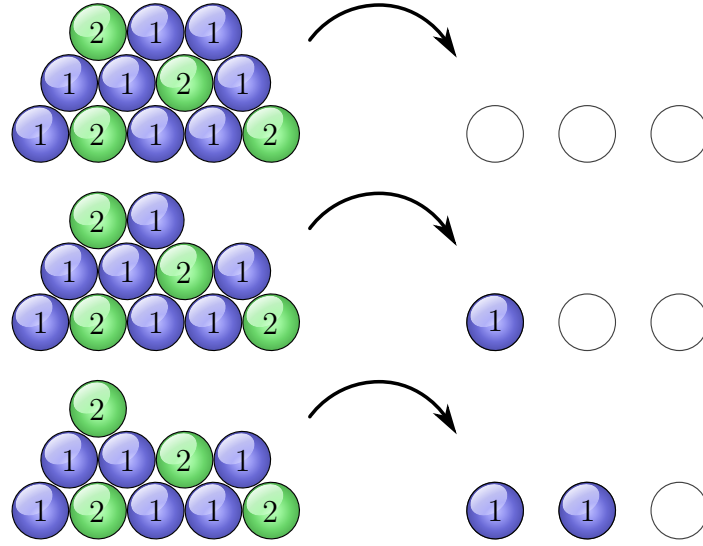


Figure 4.2: Conditional probability can be employed to calculate the probability of multiple events occurring at the same time.

We can also obtain this equation directly from the definition of conditional probability. This property is a key observation that plays a central role in establishing two important results, the *total probability theorem* and *Bayes' rule*. To formulate these two theorems, we need to revisit the notion of a partition. A collection of events A_1, A_2, \dots, A_n is said to be a *partition* of the sample space Ω if these events are disjoint and their union is the entire sample space,

$$\bigcup_{k=1}^n A_k = \Omega.$$

Visually, a partition divides an entire set into disjoint subsets, as exemplified in Figure 4.3.

Theorem 2 (Total Probability Theorem). *Let A_1, A_2, \dots, A_n be a collection of events that forms a partition of the sample space Ω . Suppose that $\Pr(A_k) > 0$ for all k . Then, for any event B , we can write*

$$\begin{aligned} \Pr(B) &= \Pr(A_1 \cap B) + \Pr(A_2 \cap B) + \dots + \Pr(A_n \cap B) \\ &= \Pr(A_1) \Pr(B|A_1) + \Pr(A_2) \Pr(B|A_2) + \dots + \Pr(A_n) \Pr(B|A_n). \end{aligned}$$

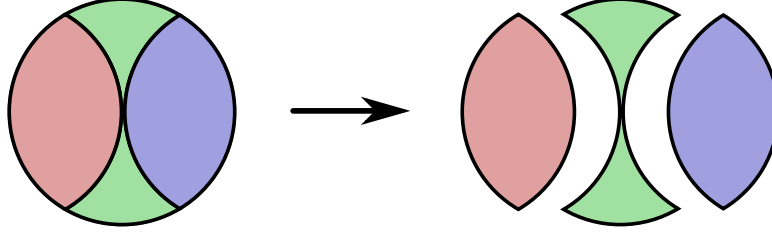


Figure 4.3: A partition of S can be formed by selecting a collection of subsets that are disjoint and whose union is S .

Proof. The collection of events A_1, A_2, \dots, A_n forms a partition of the sample space Ω . We can therefore write

$$B = B \cap \Omega = B \cap \left(\bigcup_{k=1}^n A_k \right).$$

Since A_1, A_2, \dots, A_n are disjoint sets, the events $A_1 \cap B, A_2 \cap B, \dots, A_n \cap B$ are also disjoint. Combining these two facts, we get

$$\begin{aligned} \Pr(B) &= \Pr \left(B \cap \left(\bigcup_{k=1}^n A_k \right) \right) = \Pr \left(\bigcup_{k=1}^n (B \cap A_k) \right) \\ &= \sum_{k=1}^n \Pr(B \cap A_k) = \sum_{k=1}^n \Pr(A_k) \Pr(B|A_k), \end{aligned}$$

where the fourth equality follows from the third axiom of probability. \square

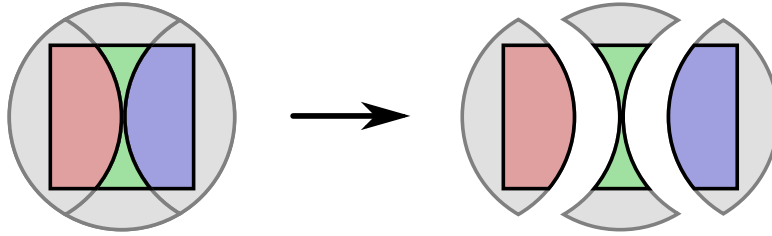


Figure 4.4: The total probability theorem states that the probability of event B can be computed by summing $\Pr(A_i \cap B)$ over all members of the partition A_1, A_2, \dots, A_n .

A graphical interpretation of Theorem 2 is illustrated in Figure 4.4. Event B can be decomposed into the disjoint union of $A_1 \cap B, A_2 \cap B, \dots, A_n \cap B$.

The probability of event B can then be computed by adding the corresponding summands

$$\Pr(A_1 \cap B), \Pr(A_2 \cap B), \dots, \Pr(A_n \cap B).$$

Example 26. *An urn contains five green balls and three red balls. A second urn contains three green balls and nine red balls. One of the two urns is picked at random, with equal probabilities, and a ball is drawn from the selected urn. We wish to compute the probability of obtaining a green ball.*

In this problem, using a divide and conquer approach seems appropriate; we therefore utilize the total probability theorem. If the first urn is chosen, then the ensuing probability of getting a green ball is $5/8$. On the other hand, if a ball is drawn from the second urn, the probability that it is green reduces to $3/12$. Since the probability of selecting either urn is $1/2$, we can write the overall probability of getting a green ball as

$$\begin{aligned} \Pr(g) &= \Pr(g \cap U_1) + \Pr(g \cap U_2) \\ &= \Pr(g|U_1) \Pr(U_1) + \Pr(g|U_2) \Pr(U_2) \\ &= \frac{5}{8} \cdot \frac{1}{2} + \frac{3}{12} \cdot \frac{1}{2} = \frac{7}{16}. \end{aligned}$$

4.3 Bayes' Rule

The following result is also very useful. It relates the conditional probability of A given B to the conditional probability of B given A .

Theorem 3 (Bayes' Rule). *Let A_1, A_2, \dots, A_n be a collection of events that forms a partition of the sample space Ω . Suppose that $\Pr(A_k) > 0$ for all k . Then, for any event B such that $\Pr(B) > 0$, we can write*

$$\begin{aligned} \Pr(A_i|B) &= \frac{\Pr(A_i) \Pr(B|A_i)}{\Pr(B)} \\ &= \frac{\Pr(A_i) \Pr(B|A_i)}{\sum_{k=1}^n \Pr(A_k) \Pr(B|A_k)}. \end{aligned} \tag{4.5}$$

Proof. Bayes' rule is easily verified. We expand the probability of $A_i \cap B$ using (4.4) twice, and we get

$$\Pr(A_i \cap B) = \Pr(A_i|B) \Pr(B) = \Pr(B|A_i) \Pr(A_i).$$

Rearranging the terms yields the first equality. The second equality in (4.5) is obtained by applying Theorem 2 to the denominator $\Pr(B)$. \square

Example 27. *April, a biochemist, designs a test for a latent disease. If a subject has the disease, the probability that the test results turn out positive is 0.95. Similarly, if a subject does not have the disease, the probability that the test results come up negative is 0.95. Suppose that one percent of the population is infected by the disease. We wish to find the probability that a person who tested positive has the disease.*

Let D denote the event that a person has the disease, and let P be the event that the test results are positive. Using Bayes' rule, we can compute the probability that a person who tested positive has the disease,

$$\begin{aligned}\Pr(D|P) &= \frac{\Pr(D) \Pr(P|D)}{\Pr(D) \Pr(P|D) + \Pr(D^c) \Pr(P|D^c)} \\ &= \frac{0.01 \cdot 0.95}{0.01 \cdot 0.95 + 0.99 \cdot 0.05} \\ &\approx 0.1610.\end{aligned}$$

Although the test may initially appear fairly accurate, the probability that a person with a positive test carries the disease remains small.

4.4 Independence

Two events A and B are said to be *independent* if $\Pr(A \cap B) = \Pr(A) \Pr(B)$. Interestingly, independence is closely linked to the concept of conditional probability. If $\Pr(B) > 0$ and events A and B are independent, then

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

That is, the *a priori* probability of event A is identical to the *a posteriori* probability of A given B . In other words, if A is independent of B , then partial knowledge of B contains no information about the likely occurrence of A . We note that independence is a symmetric relation; if A is independent of B , then B is also independent of A . It is therefore unambiguous to say that A and B are independent events.

Example 28. Suppose that two dice are rolled at the same time, a red die and a blue die. We observe the numbers that appear on the upper faces of the two dice. The sample space for this experiment is composed of thirty-six equally likely outcomes. Consider the probability of getting a four on the red die given that the blue die shows a six,

$$\begin{aligned}\Pr(\{r = 4\}|\{b = 6\}) &= \frac{\Pr(\{r = 4\} \cap \{b = 6\})}{\Pr(b = 6)} \\ &= \frac{1}{6} = \Pr(r = 4).\end{aligned}$$

From this equation, we gather that

$$\Pr(\{r = 4\} \cap \{b = 6\}) = \Pr(r = 4) \Pr(b = 6).$$

As such, rolling a four on the red die and rolling a six on the blue die are independent events.

Similarly, consider the probability of obtaining a four on the red die given that the sum of the two dice is eleven,

$$\begin{aligned}\Pr(\{r = 4\}|\{r + b = 11\}) &= \frac{\Pr(\{r = 4\} \cap \{r + b = 11\})}{\Pr(r + b = 11)} = 0 \\ &\neq \frac{1}{6} = \Pr(r = 4).\end{aligned}$$

In this case, we conclude that getting a four on the red die and a sum total of eleven are not independent events.

The basic idea of independence seems intuitively clear: if knowledge about the occurrence of event B has no impact on the probability of A , then these two events must be independent. Yet, independent events are not necessarily easy to visualize in terms of their sample space. A common mistake is to assume that two events are independent if they are disjoint. Two mutually exclusive events can hardly be independent: if $\Pr(A) > 0$, $\Pr(B) > 0$, and $\Pr(A \cap B) = 0$ then

$$\Pr(A \cap B) = 0 < \Pr(A) \Pr(B).$$

Hence, A and B cannot be independent if they are disjoint, non-trivial events.

4.4.1 Independence of Multiple Events

The concept of independence can be extended to multiple events. The events A_1, A_2, \dots, A_n are *independent* provided that

$$\Pr\left(\bigcap_{i \in \mathbb{I}} A_i\right) = \prod_{i \in \mathbb{I}} \Pr(A_i), \quad (4.6)$$

for every subset \mathbb{I} of $\{1, 2, \dots, n\}$.

For instance, consider a collection of three events, A , B and C . These events are independent whenever

$$\begin{aligned} \Pr(A \cap B) &= \Pr(A) \Pr(B) \\ \Pr(A \cap C) &= \Pr(A) \Pr(C) \\ \Pr(B \cap C) &= \Pr(B) \Pr(C) \end{aligned} \quad (4.7)$$

and, in addition,

$$\Pr(A \cap B \cap C) = \Pr(A) \Pr(B) \Pr(C).$$

The three equalities in (4.7) assert that A , B and C are *pairwise independent*. Note that the fourth equation does not follow from the first three conditions, nor does it imply any of them. Pairwise independence does not necessarily imply independence. This is illustrated below.

Example 29. *A fair coin is flipped twice. Let A denote the event that heads is observed on the first toss. Let B be the event that heads is obtained on the second toss. Finally, let C be the event that the two coins show distinct sides. These three events each have a probability of $1/2$. Furthermore, we have*

$$\Pr(A \cap B) = \Pr(A \cap C) = \Pr(B \cap C) = \frac{1}{4}$$

and, therefore, these events are pairwise independent. However, we can verify that

$$\Pr(A \cap B \cap C) = 0 \neq \frac{1}{8} = \Pr(A) \Pr(B) \Pr(C).$$

This shows that events A , B and C are not independent.

Example 30. Two dice are rolled at the same time, a red die and a blue die. Let A be the event that the number on the red die is odd. Let B be the event that the number on the red die is either two, three or four. Also, let C be the event that the product of the two dice is twelve. The individual probabilities of these events are

$$\begin{aligned}\Pr(A) &= \Pr(r \in \{1, 3, 5\}) = \frac{1}{2} \\ \Pr(B) &= \Pr(r \in \{2, 3, 4\}) = \frac{1}{2} \\ \Pr(C) &= \Pr(r \times b = 12) = \frac{4}{36}.\end{aligned}$$

We note that these events are not pairwise independent because

$$\begin{aligned}\Pr(A \cap B) &= \frac{1}{6} \neq \frac{1}{4} = \Pr(A) \Pr(B) \\ \Pr(A \cap C) &= \frac{1}{36} \neq \frac{1}{18} = \Pr(A) \Pr(C) \\ \Pr(B \cap C) &= \frac{1}{12} \neq \frac{1}{18} = \Pr(B) \Pr(C).\end{aligned}$$

Consequently, the multiple events A , B and C are not independent. Still, the probability of these three events occurring simultaneously is

$$\Pr(A \cap B \cap C) = \frac{1}{36} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{4}{36} = \Pr(A) \Pr(B) \Pr(C).$$

4.4.2 Conditional Independence

We introduced earlier the meaning of conditional probability, and we showed that the set of conditional probabilities $\{\Pr(A|B)\}$ specifies a valid probability law. It is therefore possible to discuss independence with respect to conditional probability. We say that events A_1 and A_2 are *conditionally independent*, given event B , if

$$\Pr(A_1 \cap A_2 | B) = \Pr(A_1 | B) \Pr(A_2 | B).$$

Note that if A_1 and A_2 are conditionally independent, we can use equation (4.2) to write

$$\begin{aligned}\Pr(A_1 \cap A_2 | B) &= \frac{\Pr(A_1 \cap A_2 \cap B)}{\Pr(B)} \\ &= \frac{\Pr(B) \Pr(A_1 | B) \Pr(A_2 | A_1 \cap B)}{\Pr(B)} \\ &= \Pr(A_1 | B) \Pr(A_2 | A_1 \cap B).\end{aligned}$$

Under the assumption that $\Pr(A_1|B) > 0$, we can combine the previous two expressions and get

$$\Pr(A_2|A_1 \cap B) = \Pr(A_2|B).$$

This latter result asserts that, given event B has taken place, the additional information that A_1 has also occurred does not affect the likelihood of A_2 . It is simple to show that conditional independence is a symmetric relation as well.

Example 31. *Suppose that a fair coin is tossed until heads is observed. The number of trials is recorded as the outcome of this experiment. We denote by B the event that the coin is tossed more than one time. Moreover, we let A_1 be the event that the number of trials is an even number; and A_2 , the event that the number of trials is less than six. The conditional probabilities of A_1 and A_2 , given that the coin is tossed more than once, are*

$$\begin{aligned}\Pr(A_1|B) &= \frac{\Pr(A_1 \cap B)}{\Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3} \\ \Pr(A_2|B) &= \frac{\Pr(A_2 \cap B)}{\Pr(B)} = \frac{15/32}{1/2} = \frac{15}{16}.\end{aligned}$$

The joint probability of events A_1 and A_2 given B is equal to

$$\begin{aligned}\Pr(A_1 \cap A_2|B) &= \frac{\Pr(A_1 \cap A_2 \cap B)}{\Pr(B)} \\ &= \frac{5/16}{1/2} = \frac{5}{8} = \frac{2}{3} \cdot \frac{15}{16} \\ &= \Pr(A_1|B) \Pr(A_2|B).\end{aligned}$$

We conclude that A_1 and A_2 are conditionally independent given B . In particular, we have

$$\begin{aligned}\Pr(A_2|A_1 \cap B) &= \Pr(A_2|B) \\ \Pr(A_1|A_2 \cap B) &= \Pr(A_1|B).\end{aligned}$$

We emphasize that events A_1 and A_2 are not independent with respect to the unconditional probability law.

Two events that are independent with respect to an unconditional probability law may not be conditionally independent.

Example 32. *Two dice are rolled at the same time, a red die and a blue die. We can easily compute the probability of simultaneously getting a two on the red die and a six on the blue die,*

$$\Pr(\{r = 2\} \cap \{b = 6\}) = \frac{1}{36} = \Pr(r = 2) \Pr(b = 6).$$

Clearly, these two events are independent.

Consider the probability of rolling a two on the red die and a six on the blue die given that the sum of the two dice is an odd number. The individual conditional probabilities are given by

$$\Pr(\{r = 2\} | \{r + b \text{ is odd}\}) = \Pr(\{b = 6\} | \{r + b \text{ is odd}\}) = \frac{1}{6},$$

whereas the joint conditional probability is

$$\Pr(\{r = 2\} \cap \{b = 6\} | \{r + b \text{ is odd}\}) = 0.$$

These two events are not conditionally independent.

It is possible to extend the notion of conditional independence to several events. The events A_1, A_2, \dots, A_n are conditionally independent given B if

$$\Pr\left(\bigcap_{i \in \mathbb{I}} A_i \middle| B\right) = \prod_{i \in \mathbb{I}} \Pr(A_i | B)$$

for every subset \mathbb{I} of $\{1, 2, \dots, n\}$. This definition is analogous to (4.6), albeit using the appropriate conditional probability law.

4.5 Equivalent Notations

In the study of probability, we are frequently interested in the probability of multiple events occurring simultaneously. So far, we have expressed the joint probability of events A and B using the notation $\Pr(A \cap B)$. For mathematical convenience, we also represent the probability that two events occur at the same time by

$$\Pr(A, B) = \Pr(A \cap B).$$

This alternate notation easily extends to the joint probability of several events. We denote the joint probability of events A_1, A_2, \dots, A_n by

$$\Pr(A_1, A_2, \dots, A_n) = \Pr \left(\bigcap_{k=1}^n A_k \right).$$

Conditional probabilities can be written using a similar format. The probability of A given events B_1, B_2, \dots, B_n becomes

$$\Pr(A|B_1, B_2, \dots, B_n) = \Pr \left(A \middle| \bigcap_{k=1}^n B_k \right).$$

From this point forward, we use these equivalent notations interchangeably.

Further Reading

1. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Chapter 3.
2. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Sections 1.3–1.5.
3. Miller, S. L., and Childers, D. G., *Probability and Random Processes with Applications to Signal Processing and Communications*, 2004: Sections 2.4–2.6.
4. Gubner, J. A., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006: Sections 1.5–1.6.