CHAPTER 2

CONDITIONAL PROBABILITY AND INDEPENDENCE

INTRODUCTION

This chapter introduces the important concepts of *conditional probability* and *statistical independence*. Conditional probabilities arise when it is known that a certain event has occurred. This knowledge changes the probabilities of events within the sample space of the experiment. Conditioning on an event occurs frequently, and understanding how to work with conditional probabilities and apply them to a particular problem or application is extremely important. In some cases, knowing that a particular event has occurred will not effect the probability of another event, and this leads to the concept of the statistical independence of events that will be developed along with the concept of conditional independence.

2-1 CONDITIONAL PROBABILITY

The three probability axioms introduced in the previous chapter provide the foundation upon which to develop a theory of probability. The next step is to understand how the knowledge that a particular event has occurred will change the probabilities that are assigned to the outcomes of an experiment. The concept of a *conditional probability* is one of the most important concepts in probability and, although a very simple concept, conditional probability is often confusing to students. In situations where a conditional probability is to be used, it is often overlooked or incorrectly applied, thereby leading to an incorrect answer or conclusion. Perhaps the best starting point for the development of conditional probabilities is a simple example that illustrates the context in which they arise. Suppose that we have an electronic device that has a probability p_n of still working after n months of continuous operation, and suppose that the probability is equal to 0.5 that the device will still be working after one year ($n = 12$). The device is then

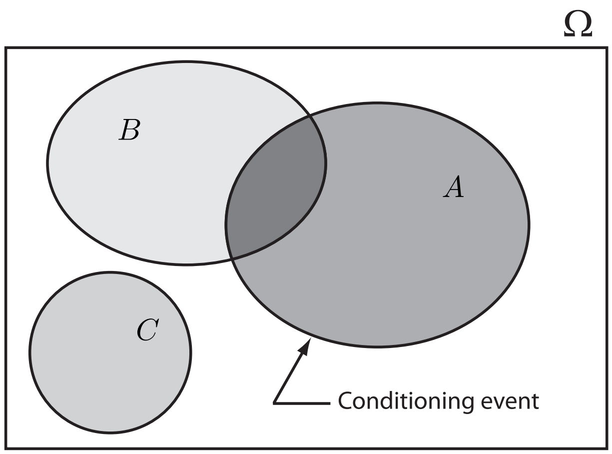

Figure 2-1: Illustration of conditioning by an event A. Any outcome *not in* A and any event C that is mutually exclusive of A becomes an impossible event.

put into operation and after one year it is still working. The question then is "What is the probability that the device will continue working for another n months." These are known as *conditional probabilities* because they are probabilities that are *conditioned* on the event that $n \geq 12$.

With this specific example in mind, let us now look at conditional probability in a more general context. Suppose that we have an experiment with a sample space Ω with probabilities defined on the events in Ω. If it is given that event A has occurred, then the only outcomes that are possible are those that are in A, and any outcomes that are not in A will have a probability of zero. Therefore, it is necessary to adjust or scale the probability of each elementary event within A so that the probability of event A is equal to one. A picture illustrating the effect of conditioning is given in Figure 2-1. There are three observations worth noting at this point:

1. If the probability of an event A is $\mathrm{P}\{A\}$, and if it is given that A has occurred, then the probability of A becomes equal to one (Axiom 2). In other words, since the only outcomes that are possible are those that are in A, then A has effectively become the new sample space or the new certain event.

2. Conditioning by A will not change the relative probabilities between the experimental outcomes in A. For example, if the probability of the elementary event $\omega_i \in A$ is equal to the probability of the elementary event

$\omega_j \in A$, then conditioning by A will not change this relationship. In other words, ω_i and ω_j will still be equally likely outcomes, $\mathrm{P}\{\omega_i\} = \mathrm{P}\{\omega_j\}$.

3. For any event C that is mutually exclusive of the conditioning event, $A \cap C = \varnothing$, the conditional probability of C will be equal to zero. In other words, given that A has occurred, if there are no outcomes in C that are also in A, then $\mathrm{P}\{C\} = 0$.

Important Concept

Conditioning an experiment on an event A effectively changes the sample space from Ω to the conditioning event, A since any outcomes not in A will have a probability of zero.

To make this a bit more concrete, consider the experiment of rolling a fair die. With a sample space consisting of six equally likely events with a probability of 1/6 for each possible outcome, suppose that the experiment is performed and we are told that the roll of the die is even (we know nothing else about the outcome, only that it is even). How does this information (conditioning) change the probabilities of the remaining events in the sample space? It should be clear that the new information (that the outcome of the roll of the die is even) should not change the relative probabilities of the remaining events, so the remaining outcomes should still be equally likely. Since only three possible outcomes remain, then their conditional probabilities should be equal to one third. Note that this also makes the probability of the conditioning event (the new sample space) equal to one. Thus, the probability that a two is rolled given that the roll resulted in an even number is equal to 1/3,

$$\mathrm{P}\{\text{roll a two, given that the roll is even}\} = 1/3$$

If we define the event

$$A = \{\text{Roll is even}\}$$

and the event

$$B = \{\text{Roll a two}\}$$

then this conditional probability of B given A is written as follows

$$\mathrm{P}\{B|A\} = 1/3$$

Note that this is *not* the same as the probability that we roll a two *and* that the roll is even, which we know is equal to one sixth.

A more interesting example of conditional probability is given in the (in)famous Monte Hall problem that may be stated as follows. Monte Hall, a famous game show host, invites you to the stage and explains that behind one of the three large doors behind you there is an expensive sports car, and behind the other two there are small consolation prizes of little value. He tells you that if you select the door that hides the sports car, it is yours to keep. After selecting one of the doors, Monte proceeds to open one of the two remaining doors to show you that the car is not behind that door, and tells you that the car is either behind the door that you selected or behind the other remaining door. Monte then gives you the option to change your selection and choose the other door. The question is *Would your chances of winning the car increase, decrease, or remain the same if you were to change your mind, and switch doors?* Before your selection was made, it is clear that the car is equally likely to be behind any one of the three doors, so the probability that the car is behind the door that you selected is initially equal to 1/3. So now the question is: *What is the probability that the car is behind the door that you selected given that it is not behind the door that was opened by Monte?*[1] For now, you are asked to think about this problem, and see if you can come up with the correct strategy to maximize your odds of winning the car. The Monte Hall problem is developed in one of the problems at the end of the chapter, which you should be able to solve once you become familiar with conditional probabilities.

Having introduced the concept of conditional probability, we now look at how conditional probabilities are found. Let Ω be a sample space with events A and B, and suppose that the probability of event B is to be determined when it is given that event A has occurred, i.e., $\mathrm{P}\{B|A\}$. Given A, all outcomes in Ω that are not in A become impossible events and will have a probability of zero, and the probability of each outcome in A must be scaled. The scaling factor is $1/\mathrm{P}\{A\}$ since this will make the probability of event A equal to one, as it must be since it is given that A has occurred. To find the probability of the event B given A, we first find the set of all outcomes that are in ***both B and A***, $B \cap A$, because any outcome not in $B \cap A$ will be equal to zero. The probability of this event, $\mathrm{P}\{B \cap A\}$, after it is scaled by $1/\mathrm{P}\{A\}$, is the conditional probability.

[1]Note that this problem would be different if Monte eliminated one of the doors *before* you make your choice.

Conditional Probability

Let A be any event with nonzero probability, $\mathrm{P}\{A\} > 0$. For any event B, the conditional probability of B given A, denoted by $\mathrm{P}\{B|A\}$, is

$$\mathrm{P}\{B|A\} = \frac{\mathrm{P}\{B \cap A\}}{\mathrm{P}\{A\}} \tag{2.1}$$

Although it will not be done here, Eq. (2.1) may be derived as a logical consequence of the axioms of probability (see [2], p. 78).

Conditional probabilities are valid probabilities in the sense that they satisfy the three probability axioms given in Sect. 1-4.1. For example, it is clear that Axiom 1 is satisfied,

$$\mathrm{P}\{B|A\} \geq 0$$

since both $\mathrm{P}\{A \cap B\}$ and $\mathrm{P}\{A\}$ are non-negative. It is also clear that $\mathrm{P}\{\Omega|A\} = 1$ since

$$\mathrm{P}\{\Omega|A\} = \frac{\mathrm{P}\{\Omega \cap A\}}{\mathrm{P}\{A\}} = \frac{\mathrm{P}\{A\}}{\mathrm{P}\{A\}} = 1$$

Finally, it is easily verified that for two mutually exclusive events B_1 and B_2,

$$\mathrm{P}\{B_1 \cup B_2|A\} = \mathrm{P}\{B_1|A\} + \mathrm{P}\{B_2|A\}$$

Specifically, note that

$$\begin{aligned}\mathrm{P}\{B_1 \cup B_2|A\} &= \frac{\mathrm{P}\{(B_1 \cup B_2) \cap A\}}{\mathrm{P}\{A\}} \\ &= \frac{\mathrm{P}\{(B_1 \cap A) \cup (B_2 \cap A)\}}{\mathrm{P}\{A\}}\end{aligned}$$

Since B_1 and B_2 are mutually exclusive, then so are the events $B_1 \cap A$ and $B_2 \cap A$. Therefore,

$$\mathrm{P}\{(A \cap B_1) \cup (A \cap B_2)\} = \mathrm{P}\{A \cap B_1\} + \mathrm{P}\{A \cap B_2\}$$

and the result follows.

A special case of conditioning occurs when A and B are mutually exclusive as illustrated in Fig. 2-2(a). Intuitively, since there are no outcomes in B that are also

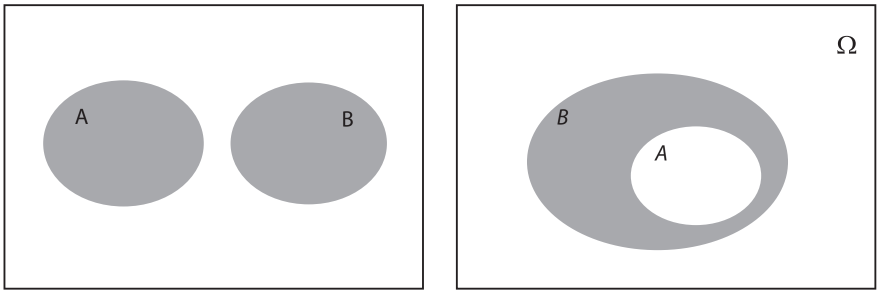

Figure 2-2: Special cases of conditioning on an event A. (a) The sets A and B are mutually exclusive, (b) One set, B, is a subset of another set, A

in A, if it is given that A has occurred then the probability of event B should be zero. To show this more formally, note that

$$\mathrm{P}\{B|A\} = \frac{\mathrm{P}\{B \cap A\}}{\mathrm{P}\{A\}} = \frac{\mathrm{P}\{\varnothing\}}{\mathrm{P}\{A\}} = 0$$

where the last equality follows from the fact that $\mathrm{P}\{\varnothing\} = 1 - \mathrm{P}\{\Omega\} = 0$. It similarly follows that $\mathrm{P}\{A|B\} = 0$ when A and B are mutually exclusive. As a specific example, consider the experiment of rolling a die, and let A and B be the following two events:

$$A = \{\text{Roll a one}\} \quad ; \quad B = \{\text{Roll an even number}\}$$

These two events are clearly disjoint, and it is clear that the probability of A given B is zero as is the probability of B given A.

Another special case occurs when A as a subset of B as illustrated in Fig. 2-2(b). In this case, since $B \cap A = A$, then

$$\mathrm{P}\{B|A\} = \frac{\mathrm{P}\{B \cap A\}}{\mathrm{P}\{B\}} = \frac{\mathrm{P}\{A\}}{\mathrm{P}\{A\}} = 1$$

This, of course, is an intuitive result since, if it is given that A has occurred, then any outcome in A will necessarily be an outcome in B and, therefore, event B also must have occurred. For example, when rolling a die, if

$$A = \{\text{Roll a one}\} \quad ; \quad B = \{\text{Roll an odd number}\}$$

then event A is a subset of event B, and the probability that an odd number is rolled (event B) is equal to one if it is given that a one was rolled (event A). If, on the other hand, the conditioning event is B, then the probability of A given B is

$$\mathrm{P}\{A|B\} = \frac{\mathrm{P}\{A \cap B\}}{\mathrm{P}\{B\}} = \frac{\mathrm{P}\{A\}}{\mathrm{P}\{B\}}$$

so the probability of event A is scaled by the probability of event B.

Example 2-1: **GEOMETRIC PROBABILITY LAW**

Consider an experiment that has a sample space consisting of the set of all positive integers,

$$\Omega = \{1, 2, 3, \ldots\}$$

and let N denote the outcome of an experiment defined on Ω. Suppose that the following probabilities are assigned to N,

$$\mathrm{P}\{N = k\} = (\tfrac{1}{2})^k \quad ; \quad k = 1, 2, 3, \ldots \tag{2.2}$$

This probability assignment is called a ***geometric probability law*** and is one that arises in ***arrival time*** problems as will be seen later.It is easy to show that this is a valid probability assignment since $\mathrm{P}\{N = k\} \geq 0$ for all k, and[2]

$$\mathrm{P}\{\Omega\} = \mathrm{P}\{N \geq 1\} = \sum_{k=1}^{\infty} (\tfrac{1}{2})^k = \tfrac{1}{2} \sum_{k=0}^{\infty} (\tfrac{1}{2})^k = \tfrac{1}{2} \frac{1}{1 - \frac{1}{2}} = 1 \tag{2.3}$$

The third axiom is satisfied automatically since probabilities are assigned individually to each elementary outcome in Ω.

Now let's find the probability that $N > N_1$ given that $N > N_0$ assuming that N_1 is greater than N_0 and both are positive integers. Using the definition of conditional probability, we have

$$\mathrm{P}\{N > N_1 | N > N_0\} = \frac{\mathrm{P}\{(N > N_1) \cap (N > N_0)\}}{\mathrm{P}\{N > N_0\}} = \frac{\mathrm{P}\{N > N_1\}}{\mathrm{P}\{N > N_0\}} \tag{2.4}$$

The probability in the numerator is

$$\mathrm{P}\{N > N_1\} = \sum_{k=N_1+1}^{\infty} (\tfrac{1}{2})^k = (\tfrac{1}{2})^{N_1} \sum_{k=1}^{\infty} (\tfrac{1}{2})^k = (\tfrac{1}{2})^{N_1} \tag{2.5}$$

where the last equality followed by using Eq. (2.3). Similarly, it follows that the probability in the denominator is $\mathrm{P}\{N > N_0\} = (\tfrac{1}{2})^{N_0}$. Therefore, the conditional probability that N is greater than N_1 given that N is greater than N_0 is

$$\mathrm{P}\{N > N_1 | N > N_0\} = \frac{(\frac{1}{2})^{N_1}}{(\frac{1}{2})^{N_0}} = (\tfrac{1}{2})^{N_1 - N_0}$$

[2] In the evaluation of this probability, the geometric series is used (See Appendix 1).

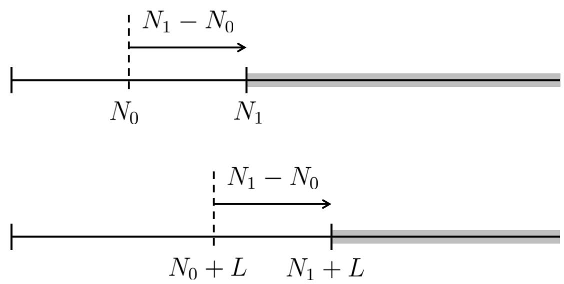

Figure 2-3: The memoryless property. The probability that $N > N_1$ given that $N > N_0$ is the same as the probability that $N > N_1 + L$ given that $N > N_0 + L$.

What is interesting is that this conditional probability depends only on the difference between N_1 and N_0. In other words,

$$\mathrm{P}\{N > N_1 | N > N_0\} = \mathrm{P}\{N > N_1 + L | N > N_0 + L\}$$

for any $L \geq 0$ as illustrated graphically in Fig. 2-1. This is known as the ***memoryless property***. ■

There will be instances in which it will be necessary to work with probabilities that are conditioned on two events, $\mathrm{P}\{A|B \cap C\}$, and express this a form similar to Eq. (2.1) that maintains the conditioning on C. To see how this is done, recall that

$$\mathrm{P}\{A|D\} = \frac{\mathrm{P}\{A \cap D\}}{\mathrm{P}\{D\}} \tag{2.6}$$

Now suppose that D is the intersection of two events, B and C,

$$D = B \cap C$$

It then follows that

$$\mathrm{P}\{A|B \cap C\} = \frac{\mathrm{P}\{A \cap B \cap C\}}{\mathrm{P}\{B \cap C\}}$$

However, we know that

$$\mathrm{P}\{A \cap B \cap C\} = \mathrm{P}\{A \cap B|C\}\mathrm{P}\{C\}$$

and

$$\mathrm{P}\{B \cap C\} = \mathrm{P}\{B|C\}\mathrm{P}\{C\}$$

Therefore,

$$P\{A|B \cap C\} = \frac{P\{A \cap B|C\}}{P\{B|C\}}$$

The interpretation is that we first define a new sample space, C, which is the conditioning event, and then we have standard conditional probability given in Eq. (2.6) that is defined on this new space.

Conditioning on Two Events

$$P\{A|B \cap C\} = \frac{P\{A \cap B|C\}}{P\{B|C\}} \tag{2.7}$$

2-2 INDEPENDENCE

In Chapter 1, the terms independent experiments and independent outcomes were used without bothering to define what was meant by *independence*. With an understanding of conditional probability, it is now possible to define and gain an appreciation for what it means for one or more events to be independent, and what is meant by conditional independence.

2-2.1 INDEPENDENCE OF A PAIR OF EVENTS

When it is said that events A and B are independent, our intuition suggests that this means that the outcome of one event should not have any effect or influence on the outcome of the other. It might also suggest that if it is known that one event has occurred, then this should not effect or change the probability that the other event will occur. Consider, for example, the experiment of rolling two fair dice. It is generally assumed (unless one is superstitious) that after the two dice are rolled, knowing what number appears on one of the dice will not help in knowing what number appears on the other. To make this a little more precise, suppose that one of the dice is red and the other is white, and let A be the event that a *one* occurs on the red die and B the event that a *one* occurs on the white die. Independence of these two events is taken to mean that knowing that event A occurred should not change the probability that event B occurs, and vice versa. Stated in terms of conditional probabilities, this may be written as follows,

$$P\{B|A\} = P\{B\} \tag{2.8a}$$

$$\mathrm{P}\{A|B\} = \mathrm{P}\{A\} \tag{2.8b}$$

From the definition of conditional probability, it follows from Eq. (2.8a) that

$$\mathrm{P}\{B|A\} = \frac{\mathrm{P}\{B \cap A\}}{\mathrm{P}\{A\}} = \mathrm{P}\{B\} \tag{2.9}$$

and, therefore, that

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\} \tag{2.10}$$

Eq. (2.10) also implies Eq. (2.10). This leads to the following definition for the ***statistical independence*** of a pair of events, A and B:

Independent Events

Two events A and B are said to be *statistically independent* (or simply independent) when

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\} \tag{2.11}$$

Two events that are not independent are said to be *dependent*.

Independence is a ***reflexive*** property in the sense that if A is independent of B, then B is independent of A. In other words, if the probability of event B does not change when it is given that event A occurs, then the probability of A will not change if it is given that event B occurs.

The concept of independence plays a central role in probability theory and arises frequently in problems and applications. Testing for independence may not always be easy, and sometimes it is necessary to assume that certain events are independent when it is believed that such an assumption is justified.

Example 2-2: **INDEPENDENCE**

Suppose that two switches are arranged in parallel as illustrated in Fig. 2-4(a). Let A_1 be the event that switch 1 is closed and let A_2 be the event that switch 2 is closed. Assume that these events are independent, and that

$$\mathrm{P}\{A_1\} = p_1 \quad ; \quad \mathrm{P}\{A_2\} = p_2$$

A connection exists from point X to point Y if either of the two switches are closed. Therefore, the probability that there is a connection is

$$\mathrm{P}\{\text{Connection}\} = \mathrm{P}\{A_1 \cup A_2\} = \mathrm{P}\{A_1\} + \mathrm{P}\{A_2\} - \mathrm{P}\{A_1 \cap A_2\} = p_1 + p_2 - p_1 p_2$$

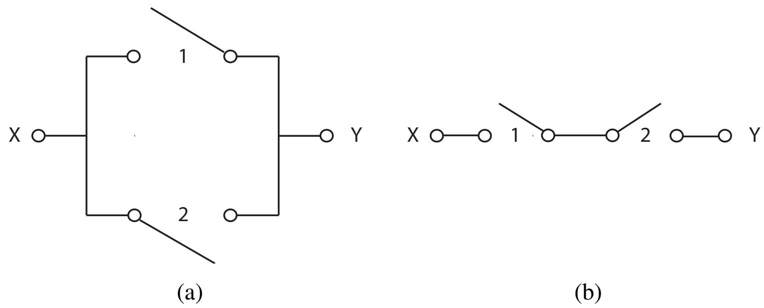

Figure 2-4: Two switches connected in (a) parallel and (b) series.

If the two switches are in series as illustrated in Fig. 2-4(b), then there will be a connection between X and Y only when both switches are closed. Therefore, for the series case,

$$\mathrm{P}\{\text{Connection}\} = \mathrm{P}\{A_1 \cap A_2\} = p_1 p_2$$

■

There are a few properties related to the independence of events that are useful to develop since they will give more insight into what the independence of two events means. The first property is that the sample space Ω is independent of any event $B \neq \Omega$.[3] This follows from

$$\mathrm{P}\{B|\Omega\} = \frac{\mathrm{P}\{B \cap \Omega\}}{\mathrm{P}\{\Omega\}} = \mathrm{P}\{B\}$$

The second property is that if A and B are mutually exclusive events, $A \cap B = \varnothing$, with $\mathrm{P}\{A\} \neq 0$ and $\mathrm{P}\{B\} \neq 0$, then A and B will be ***dependent*** events. To understand this intuitively, note that when A and B are mutually exclusive, if event B occurs then event A cannot occur, and vice versa. Therefore, if it is known that one of these events occurs, this it is known that the other one cannot occur, thereby establishing the dependence between the two events. To show formally, note that if A and B are disjoint events, then

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{\varnothing\} = 0$$

[3]The exclusion of $B \neq \Omega$ is necessary because any set B is always dependent upon itself. More specifically, since $\mathrm{P}\{B|B\} = 1$ then this will not be the same as $\mathrm{P}\{B\}$, which is required for independence, unless $B = \Omega$.

However, in order for A and B to be independent, it is necessary that

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\}$$

With the assumption that both A and B have non-zero probabilities, it then follows that A and B must be dependent.

The next property is that if B is a subset of A, then A and B will be dependent events unless $\mathrm{P}\{A\} = 1$. The fundamental idea here is that if B is a subset of A, then if it is given that event B has occurred, then it is known that event A also must have occurred because any outcome in B is also an outcome in A. To demonstrate this dependence formally, note that if $B \subset A$, then $B \cap A = B$ and

$$\mathrm{P}\{B|A\} = \frac{\mathrm{P}\{B \cap A\}}{\mathrm{P}\{A\}} = \frac{\mathrm{P}\{B\}}{\mathrm{P}\{A\}} \neq \mathrm{P}\{B\}$$

unless $\mathrm{P}\{A\} = 1$, i.e., A is the certain event.

The last property is that if A and B are independent, then A and B^c are also independent. To show this, note that

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

Since B and B^c are mutually exclusive events, then $A \cap B$ and $A \cap B^c$ are also mutually exclusive and

$$\mathrm{P}\{A\} = \mathrm{P}\{A \cap B\} + \mathrm{P}\{A \cap B^c\}$$

Therefore,

$$\mathrm{P}\{A|B^c\} = \frac{\mathrm{P}\{A \cap B^c\}}{\mathrm{P}\{B^c\}} = \frac{\mathrm{P}\{A\} - \mathrm{P}\{A \cap B\}}{1 - \mathrm{P}\{B\}}$$

Since A and B are independent, $\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\}$, and we have

$$\mathrm{P}\{A|B^c\} = \frac{\mathrm{P}\{A\} - \mathrm{P}\{A\}\mathrm{P}\{B\}}{1 - \mathrm{P}\{B\}} = \mathrm{P}\{A\}$$

which establishes the independence of A and B^c.

Properties of Independent Events

1. The events Ω and $\varnothing$ are independent of any event A unless $\mathrm{P}\{A\} = 1$ or $\mathrm{P}\{A\} = 0$.
2. If $A \cap B = \varnothing$, with $\mathrm{P}\{A\} \neq 0$ and $\mathrm{P}\{B\} \neq 0$, then A and B are dependent events.
3. If $B \subset A$ then A and B will be dependent unless $\mathrm{P}\{A\} = 1$.
4. If A and B are independent, then A and B^c are independent,

Example 2-3: **ARRIVAL OF TWO TRAINS**[4]

Trains X and Y arrive at a station at random times between 8:00 A.M. and 8:20 A.M. Train X stops at the station for three minutes and Train Y stops for five minutes. Assuming that the trains arrive at times that are independent of each other, we will find the probabilities of several events that are defined in terms of the train arrival times. First, however, it is necessary that we specify the underlying experiment, draw a picture of the sample space, and make probability assignments on the events that are defined within this sample space.

To begin, let x be the arrival time of train X, and y the arrival time of train Y, with x and y being equal to the amount of time past 8:00 A.M. that the train arrives. It should then be clear that the outcomes of this experiment are all pairs of numbers (x, y) with $0 \leq x \leq 20$ and $0 \leq y \leq 20$. In other words, the sample space Ω consists of all points within the square shown in Fig. 2-5(a).

The next step is to assign probabilities to events within the sample space. It is assumed that the trains arrive at random times between 8:00 A.M. and 8:20 A.M., and that the trains arrive at times that are independent of each other. What it means for a train to arrive at a ***random time*** between 8:00 A.M. and 8:20 A.M. is that a train arrival at any time within this interval is equally likely (equally probable). For example, the probability of a train arriving between 8:00 A.M. and 8:01 A.M. will be the same as the probability that it arrives between 8:10 A.M. and 8:11 A.M. (equal-length time intervals). Since the probability that the train arrives between 8:00 A.M. and 8:20 A.M. is equal to one, this suggests the following probability

[4] From [3], p. 33

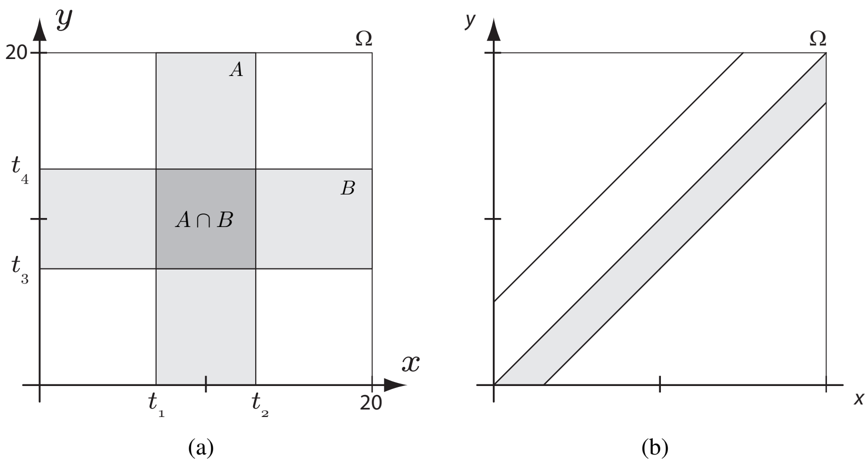

Figure 2-5: The experiment of two train arrivals over a twenty minute time interval. (a) The sample space, Ω, and the events $A = \{t_1 \leq x \leq t_2\}$, $B = \{t_3 \leq y \leq t_4\}$, and $A \cap B$. (b) The event $A = \{y \leq x\}$ and the event B that the trains meet at the station, which is defined by $B = \{-3 < x - y < 5\}$.

measure for the event $A = \{t_1 \leq x \leq t_2\}$

$$\mathrm{P}\{A\} = \frac{t_2 - t_1}{20} \quad ; \quad 0 \leq t_1 \leq t_2 \leq 20$$

Note that the event A corresponds to those outcomes that lie in the vertical strip shown in Fig. 2-5(a), and the probability of event A is equal to the width of the strip divided by 20. Furthermore, the probability of a train arriving over any collection of time intervals will be equal to the total duration of these time intervals divided by 20. For example,

$$\mathrm{P}\{(0 \leq x \leq 5) \cup (12 \leq x \leq 15)\} = \frac{8}{20} = 0.4$$

A similar measure is defined for y, with

$$\mathrm{P}\{t_3 \leq y \leq t_4\} = \frac{t_4 - t_3}{20} \quad ; \quad 0 \leq t_3 \leq t_4 \leq 20$$

Note that the event

$$B = \{t_3 \leq x \leq t_4\}$$

is represented by the horizontal strip of outcomes in Ω shown in Fig. 2-5(a).

To complete the probability specification, it is necessary to determine the probability of the intersection of events A and B. Since it is assumed that the arrival times of the two trains are independent events, then A and B are independent and

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\} = \frac{(t_2 - t_1)(t_4 - t_3)}{20 \times 20}$$

The event $A \cap B$ is the rectangular event shown in Fig. 2-5(a), and we conclude that the probability of any rectangle within Ω is equal to the area of the rectangle divided by 400. More generally, the probability of any general region within the sample space will be equal to the area of the region divided by 400.

Having specified the probabilities on events in Ω, let's find the probability that train X arrives before train Y. This is the event

$$A = \{x \leq y\}$$

which corresponds to those outcomes that are in the triangular region above the line $x = y$ in Fig. 2-5(b). Since the area of this triangle is equal to 200, then

$$\mathrm{P}\{A\} = \frac{200}{400} = 0.5$$

This result makes sense intuitively since each train arrives at a random time and each arrives independently of the other. Therefore, there is nothing that would make one train more likely than the other to arrive first at the station.

Now let's find the probability that the trains meet at the station, i.e., the second train arrives at the station before the first one departs. Since train X is at the station for three minutes, if train X is the first to arrive, then train Y must arrive within three minutes after the arrival of train X, i.e., $x \leq y \leq x + 3$, or

$$0 \leq y - x \leq 3$$

Similarly, if train Y is the first to arrive, since train Y remains at the station for five minutes, then train X must arrive within five minutes after the arrival of train Y, i.e., $y \leq x \leq y + 5$, or

$$0 \leq x - y \leq 5$$

Therefore, the event that the trains meet at the station is

$$C = \{-3 \leq x - y \leq 5\}$$

which corresponds to the shaded region consisting of two trapezoids shown in Fig. 2-5(b). Since the area of these trapezoids is 143, then

$$\mathrm{P}\{C\} = \frac{143}{400}$$

■

2-2.2 Independence of More than Two Events

The definition given in Eq. (2.11) is concerned with the independence of a pair of events, A and B. If there are three events, A, B, and C, then it would be tempting to say that A, B, and C are independent if the following three conditions hold:

$$\begin{aligned} \mathrm{P}\{A \cap B\} &= \mathrm{P}\{A\}\mathrm{P}\{B\} \\ \mathrm{P}\{B \cap C\} &= \mathrm{P}\{B\}\mathrm{P}\{C\} \\ \mathrm{P}\{C \cap A\} &= \mathrm{P}\{C\}\mathrm{P}\{A\} \end{aligned} \tag{2.12}$$

However, when Eq. (2.12) is satisfied, then A, B, and C are said to be ***independent in pairs***, which means that the occurrence of any ***one*** of the three events will not have any effect on the probability that either one of the other events will occur. However, Eq. (2.12) does not necessarily imply that the probability of one of the events will not change if it is given that the other ***two*** events have occurred. In other words, it may not necessarily follow from Eq. (2.12) that

$$\mathrm{P}\{A|B \cap C\} = \mathrm{P}\{A\}$$

nor is it necessarily true that independence in pairs imples that

$$\mathrm{P}\{A \cap B \cap C\} = \mathrm{P}\{A\}\mathrm{P}\{B\}\mathrm{P}\{C\}$$

The following example illustrates this point and shows that some care is needed in dealing with independence of three or more events, and that sometimes our intuition may fail us.

Example 2-4: Independence in Pairs

Consider a digital transmitter that sends two binary digits, b_1 and b_2, with each bit being equally likely to be a zero or a one,

$$\mathrm{P}\{b_i = 0\} = \mathrm{P}\{b_i = 1\} = \tfrac{1}{2} \quad ; \quad i = 1, 2$$

In addition, suppose that the events $\{b_1 = i\}$ is independent of the event $\{b_2 = j\}$ for $i, j = 1, 2$,

$$\mathrm{P}\{(b_1 = i) \cap (b_2 = j)\} = \mathrm{P}\{b_1 = i\}\mathrm{P}\{b_2 = j\} = \tfrac{1}{4} \quad ; \quad i, j = 0, 1$$

The sample space for this experiment consists of four possible outcomes, each corresponding to one of the four possible pairs of binary digits as illustrated in Fig. 2-6(a). Now let A be the event that the first bit is zero,

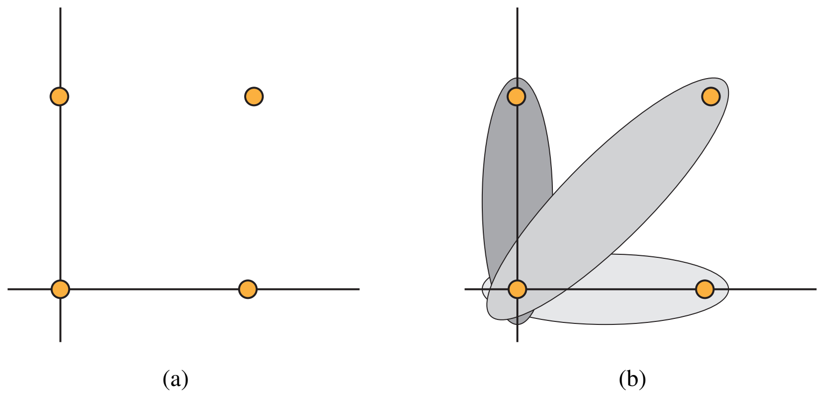

Figure 2-6: Independence in pairs.

$$A = \{b_1 = 0\} = \{00\} \cup \{01\}$$

and B the event that the second bit is zero,

$$B = \{b_2 = 0\} = \{00\} \cup \{10\}$$

and C the event that the two bits are the same,

$$C = \{b_1 = b_2\} = \{00\} \cup \{11\}$$

These events are illustrated in Fig. 2-6(b). Since the probability of each elementary event is equal to 1/4, and since each of the events A, B, and C contain exactly two elementary events, then

$$\mathrm{P}\{A\} = \mathrm{P}\{B\} = \mathrm{P}\{C\} = 1/2$$

It is easy to show that these three events are independent in pairs. For example, since

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{00\} = \tfrac{1}{4} = \mathrm{P}\{A\}\mathrm{P}\{B\}$$

then A and B are independent. It may similarly be shown that A and C are independent and that B and C are independent.

However, consider what happens when one of the events is conditioned on the other two, such as $\mathrm{P}\{A|B \cap C\}$. In this case,

$$\mathrm{P}\{A|B \cap C\} = \frac{\mathrm{P}\{A \cap B \cap C\}}{\mathrm{P}\{B \cap C\}}$$

and since $A \cap B \cap C = \{00\}$ and $B \cap C = \{00\}$ are the same event, then

$$\mathrm{P}\{A|B \cap C\} = 1$$

Therefore,

$$\mathrm{P}\{A|B \cap C\} \neq \mathrm{P}\{A\}$$

and it follows that A is *not independent* of the event $B \cap C$. In addition, note that since

$$\mathrm{P}\{A \cap B \cap C\} = \tfrac{1}{4}$$

and

$$\mathrm{P}\{A\}\mathrm{P}\{B\}\mathrm{P}\{C\} = \tfrac{1}{16}$$

then

$$\mathrm{P}\{A \cap B \cap C\} \neq \mathrm{P}\{A\}\mathrm{P}\{B\}\mathrm{P}\{C\}$$

which would be the obvious generalization of the definition given in Eq. (2.11) for three events. ■

The previous example shows that generalizing the definition for the independence of two events to the independence of three events requires more than pair-wise independence. Therefore, what is required for three events to be said to be statistically independent is given in the following definition:

Independence of Three Events

Three events A, B, and C are said to be *statistically independent* if they are independent in pairs, and

$$\mathrm{P}\{A \cap B \cap C\} = \mathrm{P}\{A\}\mathrm{P}\{B\}\mathrm{P}\{C\} \qquad (2.13)$$

The extension to more than three events follows by induction. For example, four events A, B, C, and D are independent if they are independent in groups of three, and

$$\mathrm{P}\{A \cap B \cap C \cap D\} = \mathrm{P}\{A\}\mathrm{P}\{B\}\mathrm{P}\{C\}\mathrm{P}\{D\}$$

Continuing, events A_i for $i = 1, \ldots, n$ are said to be independent if they are independent in groups of $n-1$ and

$$P\left\{\bigcap_{i=1}^{n} A_i\right\} = \prod_{i=1}^{n} \mathrm{P}\{A_i\}$$

2-2.3 CONDITIONAL INDEPENDENCE

Recall that if A and B are independent events, then event B does not have any influence on event A, and the occurrence of B will not change the probability of event A. Since independence is reflexive, then the converse is also true. Frequently, however, there will be cases in which two events are independent, but this independence will depend (explicitly or implicitly) on some other condition or event. To understand how such a situation might arise, consider the following example.

Example 2-5: **ELECTRICAL COMPONENTS**[5]

Suppose that an electronic system has two components that ***operate independently*** of each other in the sense that the failure of one component is not affected by and does not have any effect on the failure of the other. In addition, let A and B be the following events:

$$A = \{\text{Component 1 operates without failure for one year}\}$$

$$B = \{\text{Component 2 operates without failure for one year}\}$$

It would seem natural to assume that events A and B are statistically independent given the assumption of ***operational independence***. However, this may not be the case since, in some situations, there may be other random factors or events that affect each component in different ways. In these cases, statistical independence will be conditioned (depend upon) these other factors or events. For example, suppose that the operating temperature of the system affects the likelihood of a failure of each component, and it does so in different ways. More specifically, let C be the event that the system is operated within what is considered to be the normal temperature range for at least 90% of the time,

$$C = \{\text{Normal Temperature Range 90\% of the time}\}$$

and suppose that

$$\mathrm{P}\{A|C\} = 0.9 \quad ; \quad \mathrm{P}\{B|C\} = 0.8$$

and

$$\mathrm{P}\{A|C^c\} = 0.8 \quad ; \quad \mathrm{P}\{B|C^c\} = 0.7$$

In addition, let us assume that $\mathrm{P}\{C\} = 0.9$. Since the components operate independently under any given temperature, then it is reasonable to assume that

$$\mathrm{P}\{A|B \cap C\} = \mathrm{P}\{A|C\} \tag{2.14}$$

$$\mathrm{P}\{A|B \cap C^c\} = \mathrm{P}\{A|C^c\} \tag{2.15}$$

[5]From Pfeiffer 4.

In other words, given that the temperature is within the normal range, then failure of one component is not affected by the failure of the other, and the same is true if the temperature is not within the normal range. From Eq. (2.7) we know that the conditional probability $\mathrm{P}\{A|B \cap C\}$ is equal to

$$\mathrm{P}\{A|B \cap C\} = \frac{\mathrm{P}\{A \cap B|C\}}{\mathrm{P}\{B|C\}} = \mathrm{P}\{A|C\}$$

Therefore, it follows from Eq. (2.14) that

$$\mathrm{P}\{A \cap B|C\} = \mathrm{P}\{A|C\}\mathrm{P}\{B|C\} \tag{2.16}$$

which says that A and B are independent when conditioned on event C. Similarly, it follows from Eq. (2.7) and Eq. (2.15) that

$$\mathrm{P}\{A \cap B|C^c\} = \mathrm{P}\{A|C^c\}\mathrm{P}\{B|C^c\} \tag{2.17}$$

However, neither Eq. (2.16) nor Eq. (2.17) necessarily imply that A and B are (unconditionally) independent, since this requires that

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\}$$

To determine whether or not A and B are independent, we may use the special case of the total probability theorem given in Eq. (3.2) to find the probability of event A,

$$\begin{aligned}
\mathrm{P}\{A\} &= \mathrm{P}\{A|C\}\mathrm{P}\{C\} + \mathrm{P}\{A|C^c\}\mathrm{P}\{C^c\} \\
&= (0.9)(0.9) + (0.8)(0.1) = 0.89
\end{aligned}$$

as well as the probability of event B,

$$\begin{aligned}
\mathrm{P}\{B\} &= \mathrm{P}\{B|C\}\mathrm{P}\{C\} + \mathrm{P}\{B|C^c\}\mathrm{P}\{C^c\} \\
&= (0.8)(0.9) + (0.7)(0.1) = 0.79
\end{aligned}$$

Finally, again using Eq. (3.2) along with Eq. (2.16) and Eq. (2.17) we have

$$\begin{aligned}
\mathrm{P}\{A \cap B\} &= \mathrm{P}\{A \cap B|C\}\mathrm{P}\{C\} + \mathrm{P}\{A \cap B|C^c\}\mathrm{P}\{C^c\} \\
&= \mathrm{P}\{A|C\}\mathrm{P}\{B|C\}\mathrm{P}\{C\} + \mathrm{P}\{A|C^c\}\mathrm{P}\{B|C^c\}\mathrm{P}\{C^c\} \\
&= (0.9)(0.8)(0.9) + (0.8)(0.7)(0.1) = 0.704
\end{aligned}$$

Since $\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\}\mathrm{P}\{B\} = 0.7031$ then A and B are not independent.

In order to more clearly understand where the dependency is coming in, note that if Component 1 fails, then it is more likely that the operating temperature is outside the normal range, which increases the probability that the second component will fail. If Component 1 does not fail, then this makes it more likely that the operating temperature is within the normal range and, therefore, it is less likely that Component 2 will fail. ■

As illustrated in the previous example, two events A and B that are not statistically independent may become independent when conditioned on another event C. This leads to the concept of conditional independence, which is defined as follows:

Conditional Independence

Two events A and B are said to be conditionally independent given an event C if

$$\mathrm{P}\{A \cap B|C\} = \mathrm{P}\{A|C\}\mathrm{P}\{B|C\} \tag{2.18}$$

A convenient way to interpret Eq. (2.18) and to view the concept of conditional independence is as follows. When it is given that event C occurs, then C becomes the new sample space, and it is in this new sample space that event B becomes independent of A. Thus, the conditioning event removes the dependencies that exist between A and B. As is the case for independence, conditional independence is ***reflexive*** in the sense that if A is conditionally independent of B given C then B is conditionally independent of A given C.

One might be tempted to conclude that conditional independence is a weaker form of independence in the sense that if A and B are independent, then they will be independent for any conditioning event C. This, however, is not the case as illustrated abstractly in Fig. 2-7, which shows two events A and B with $A \cap B$ not empty and a conditioning set C that includes elements from both A and B. If $\mathrm{P}\{A\} = \mathrm{P}\{B\} = \frac{1}{2}$ and $\mathrm{P}\{A \cap B\} = \frac{1}{4}$, then A and B are independent events. However, note that

$$\mathrm{P}\{A \cap B|C\} = 0$$

while both $\mathrm{P}\{A|C\}$ and $\mathrm{P}\{B|C\}$ are non-zero. Therefore, A and B are not conditionally independent when the conditioning event is C. A more concrete example is given below.

Example 2-6: INDEPENDENT BUT NOT CONDITIONALLY INDEPENDENT

Figure 2-7: Something.

Let $\Omega = \{1, 2, 3, 4\}$ be a set of five equally like outcomes, and let

$$A = \{1, 2\} \quad ; \quad B = \{2, 3\}$$

Clearly, $\mathrm{P}\{A\} = 1/2$ and $\mathrm{P}\{B\} = 1/2$ and $\mathrm{P}\{A \cap B\} = 1/4$. Therefore, A and B are independent. However, if $C = \{1, 4\}$, then $\mathrm{P}\{A|C\} = 1/2$ and $\mathrm{P}\{B|C\} = 1/2$ while $\mathrm{P}\{A \cap B|C\} = 0$. Therefore, although A and B are independent, they are not conditionally independent given the event C. ■

References

1. Alvin W.Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill, New York, 1967.

2. Harold J. Larson and Bruno O. Schubert, *Random Variables and Stochastic Processes, Volume 1*, John Wiley & Sons, 1979.

3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Second Edition, 1984.

4. P. Pfeiffer, *Probability for Applications*, Springer, 1989.