

Chapter 5. Probability

Probability is a mathematical language for quantifying uncertainty. In this chapter we introduce the basic ideas underlying probability theory.

5.1 Sample Spaces and Events

5.1.1 Sample Spaces

We consider a random **experiment** whose range of possible outcomes can be described by a set S , called the **sample space**.

We use S as our universal set (Ω).

Example

- Coin tossing: $S = \{H, T\}$.
- Die rolling: $S = \{\square, \blacksquare, \blacklozenge, \blacktriangle, \blacktriangledown, \blacksquare\}$.
- 2 coins: $S = \{(H, H), (H, T), (T, H), (T, T)\}$.

■

5.1.2 Events

An **event** E is any subset of the sample space, $E \subseteq S$; it is a collection of some of the possible outcomes.

Example

- Coin tossing: $E = \{H\}$, $E = \{T\}$.
- Die rolling: $E = \{\square\}$, $E = \{\text{Even numbered face}\} = \{\blacksquare, \blacklozenge, \blacksquare\}$.
- 2 coins: $E = \{\text{Head on the first toss}\} = \{(H, H), (H, T)\}$.

■

Extreme possible events are \emptyset (the *null* event) or S .

The *singleton* subsets of S , those subsets which contain exactly one element from S , are known as the **elementary events** of S .

Suppose we now perform this random experiment; the outcome will be a single element $s^* \in S$. Then for any event $E \subseteq S$, we will say E has **occurred** if and only if $s^* \in E$.

If E has not occurred, it must be that $s^* \notin E \iff s^* \in \bar{E}$, so \bar{E} has occurred; so \bar{E} can be read as the event *not* E .

First notice that the smallest event which will have occurred will be the singleton $\{s^*\}$. For any other event E , E will occur if and only if $\{s^*\} \subset E$. Thus we can immediately draw two conclusions before the experiment has even been performed.

Remark 1. For any sample space S , the following statements will always be true:

1. the null event \emptyset will never occur;
2. the universal event S will always occur.

Hence it is only for events E in between these extreme events, $\emptyset \subset E \subset S$ for which we have uncertainty about whether E will occur. It is precisely for quantifying this uncertainty over these events that we require the notion of probability.

5.1.3 Combinations of Events

Consider a set of events $\{E_1, E_2, \dots\}$.

- The event $\bigcup_i E_i = \{s \in S \mid \exists i \text{ s.t. } s \in E_i\}$ will occur if and only if *at least one* of the events $\{E_i\}$ occurs. So $E_1 \cup E_2$ can be read as event E_1 or E_2 or both.
- The event $\bigcap_i E_i = \{s \in S \mid \forall i, s \in E_i\}$ will occur if and only if *all* of the events $\{E_i\}$ occur. So $E_1 \cap E_2$ can be read as events E_1 and E_2 .
- The events are said to be **mutually exclusive** if $\forall i, j, E_i \cap E_j = \emptyset$ (i.e. they are disjoint). At most one of the events can occur.

5.2 The σ -algebra

Henceforth, we shall think of events as subsets of the sample space S . Thus events are subsets of S , but need all subsets of S be events? The answer is *no*, for reasons that are beyond the scope of this course. But it suffices to think of the collection of events as a subcollection \mathcal{F} of the sets of all subsets of S . This subcollection should have the following properties:

- a) if $E, F \in \mathcal{F}$ then $E \cup F \in \mathcal{F}$ and $E \cap F \in \mathcal{F}$;
- b) if $E \in \mathcal{F}$ then $\bar{E} \in \mathcal{F}$;
- c) $\emptyset \in \mathcal{F}$.

A collection \mathcal{F} of subsets of S which satisfies these three conditions is called a **field**. It follows from the properties of a field that if $E_1, E_2, \dots, E_k \in \mathcal{F}$, then

$$\bigcup_{i=1}^k E_i \in \mathcal{F}.$$

So, \mathcal{F} is closed under finite unions and hence under finite intersections also. To see this note that if $E_1, E_2 \in \mathcal{F}$, then

$$\bar{E}_1, \bar{E}_2 \in \mathcal{F} \implies \underbrace{\bar{E}_1 \cup \bar{E}_2 \in \mathcal{F}}_{\text{by a)}} \implies \underbrace{\overline{(\bar{E}_1 \cup \bar{E}_2)} \in \mathcal{F}}_{\text{by b)}} \implies \underbrace{E_1 \cap E_2 \in \mathcal{F}}_{\text{by De Morgans Law}}.$$

This is fine when S is a finite set, but we require slightly more to deal with the common situation when S is infinite, as the following example indicates.

Example A coin is tossed repeatedly until the first head turns up; we are concerned with the number of tosses before this happens. The set of all possible outcomes is the set $S = \{0, 1, 2, \dots\}$. We may seek to assign a probability to the event E , that the first head occurs after an even number of tosses, that is $E = \{1, 3, 5, \dots\}$. This is an infinite countable union of members of S and we require that such a set belongs to \mathcal{F} in order that we can discuss its probability. ■

Thus we also require that the collection of events to be closed under the operation of taking countable unions, not just finite unions.

Definition 5.2.1. A collection \mathcal{F} of subsets of S is called a σ -field (or σ -algebra) if it satisfies the following conditions:

- a) $\emptyset \in \mathcal{F}$;
- b) if $E_1, E_2, \dots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$;
- c) if $E \in \mathcal{F}$ then $\bar{E} \in \mathcal{F}$.

Here are some examples of σ -algebras:

Example

- The smallest σ -algebra associated with S is the collection $\mathcal{F} = \{\emptyset, S\}$.
- If E is any subset of S then $\mathcal{F} = \{\emptyset, E, \bar{E}, S\}$ is a σ -algebra. ■

To recap, with any experiment we may associate a pair (S, \mathcal{F}) , where S is the set of all possible outcomes (or elementary events) and \mathcal{F} is a σ -algebra of subsets of S , which contains all the events in whose occurrences we may be interested. So, from now on, to call a set E an event is equivalent to asserting that E belongs to the σ -algebra in question.

5.3 Probability Measure

For an event $E \subseteq S$, the probability that E occurs will be written as $P(E)$.

Interpretation: P is a set-function or measure that assigns “weight” to collections of possible outcomes of an experiment. There are many ways to think about precisely how this assignment is achieved;

CLASSICAL : “Consider equally likely sample outcomes ...”

FREQUENTIST : “Consider long-run *relative frequencies*...”

SUBJECTIVE : “Consider personal degree of belief...”

Formally, we have the following definition:

Definition 5.3.1. A **probability measure** P on (S, \mathcal{F}) is a mapping $P : \mathcal{F} \rightarrow [0, 1]$ satisfying

a) $P(S) = 1$;

b) if E_1, E_2, \dots is a collection of disjoint members of \mathcal{F} , so that $E_i \cap E_j = \emptyset$ for all pairs i, j with $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

The triple (S, \mathcal{F}, P) , consisting of a set S , a σ -algebra \mathcal{F} and probability measure P on (S, \mathcal{F}) is called a **probability space**.

5.3.1 Properties of $P(\cdot)$: The Axioms of Probability

The triple (S, \mathcal{F}, P) denotes a typical probability space. We now give some of its simple but important properties. For events $E, F \subseteq S$,

1. $P(\bar{E}) = 1 - P(E)$
2. if $E \subseteq F$, then $P(E) \leq P(F)$.
3. In general, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.
4. $P(E \cap \bar{F}) = P(E) - P(E \cap F)$.
5. $P(E \cup F) \leq P(E) + P(F)$.
6. $P(E \cap F) \geq P(E) + P(F) - 1$.

5.3.2 Interpretations of Probability

Classical

Suppose that the sample space $S = \{s_1, \dots, s_n\}$ is finite. For example, if we toss a die twice, then S has 36 elements: $S = \{(i, j); i, j \in \{1, \dots, 6\}\}$. If each outcome is equally likely, then $P(E) = |E|/36$ where $|E|$ denotes the number of elements in E . The probability that the sum of the dice is 11 is $2/36$ since there are two outcomes that correspond to this event.

In general, if S is finite and the *elementary events* are considered “equally likely”, then the probability of an event E is the proportion of all outcomes in S in which lie inside E ,

$$P(E) = \frac{|E|}{|S|}.$$

Example Rolling a die: Elementary events are $\{\square\}, \{\square\}, \dots, \{\boxplus\}$.

- $P(\{\square\}) = P(\{\square\}) = \dots = P(\{\boxplus\}) = \frac{1}{6}$.
- $P(\text{Odd number}) = P(\{\square, \square, \boxplus\}) = \frac{3}{6} = \frac{1}{2}$.

■

The “equally likely” (uniform) idea can be extended to infinite spaces, by apportioning probability to sets not by their cardinality but by other standard *measures*, like volume or mass.

Example If a meteorite were to strike Earth, the probability that it will strike land rather than sea would be given by

$$\frac{\text{Total area of land}}{\text{Total area of Earth}}.$$

■

Frequentist

Observation shows that if one takes repeated observations in “identical” random situations, in which event E may or may not occur, then the proportion of times in which E occurs tends to some limiting value – called the probability of E .

Example Proportion of heads in tosses of a coin: $H, H, T, H, T, T, H, T, T, \dots \rightarrow \frac{1}{2}$.

■

Subjective

Probability is a degree of belief held by an individual.

For example, De Finetti (1937/1964) suggested the following: Suppose a random experiment is to be performed, where an event $E \subseteq S$ may or may not happen. Now suppose an individual is entered into a game regarding this experiment where he has two choices, each leading to monetary consequences:

1. Gamble: If E occurs, he wins £1; if \bar{E} occurs, he wins £0;
2. Stick: Regardless of the outcome of the experiment, he receives $\mathcal{E}P(E)$ for some real number $P(E)$.

The critical value of $P(E)$ for which the individual is *indifferent* between options 1 and 2 is defined to be the individual’s probability for the event E occurring.

This procedure can be repeated for all possible events E in S .

Suppose after this process of *elicitation* of the individual’s preferences under the different events, we can simultaneously arrange an arbitrary number of monetary bets with the individual based on the outcome of the experiment.

If it is possible to choose these bets in such a way that the individual is certain to lose money (this is called a “Dutch Book”), then the individual's degrees of belief are said to be **incoherent**.

To be **coherent**, it is easily seen, for example, that we must have $0 \leq P(E) \leq 1$ for all events E , $E \subseteq F \implies P(E) \leq P(F)$, etc.

5.3.3 Independent Events

If we flip a fair coin twice, then the probability of two heads is $\frac{1}{2} \times \frac{1}{2}$. We multiply the probabilities because we regard the two tosses as independent. The definition of independence is as follows:

Definition 5.3.2. Two events E and F are **independent** if and only if

$$P(E \cap F) = P(E)P(F).$$

Extension: The events E_1, \dots, E_k are independent if, for **every** subset of events of size $\ell \leq k$, indexed by $\{i_1, \dots, i_\ell\}$, say,

$$P\left(\bigcap_{j=1}^{\ell} E_{i_j}\right) = \prod_{j=1}^{\ell} P(E_{i_j})$$

Independence can arise in two distinct ways. Sometimes, we explicitly assume that two events are independent. For example, in tossing a coin twice we usually assume the tosses are independent. In other instances, we derive independence by verifying that $P(E \cap F) = P(E)P(F)$ holds true.

Example Toss a fair coin 10 times. Let A be the event {at least one head}. Let T_j be the event {tails occurs on the j th toss}. Then

$$\begin{aligned} P(A) &= 1 - P(\overline{A}) \\ &= 1 - P(\text{all tails}) \\ &= 1 - P(T_1 \cap T_2 \cap \dots \cap T_{10}) \\ &= 1 - P(T_1)P(T_2) \cdots P(T_{10}) \quad \text{by indep.} \\ &= 1 - \left(\frac{1}{2}\right)^{10} \end{aligned}$$

■

Example Suppose that the events E and F are independent. Show that \overline{E} and F are also independent.

$$\text{So } \underbrace{P(\overline{E} \cap F) = P(F) - P(E \cap F)}_{\text{by Axiom (4)}} = \underbrace{P(F) - P(E)P(F)}_{\text{by indep}} = (1 - P(E))P(F) = P(\overline{E})P(F). \quad \blacksquare$$

Example Suppose that E and F are disjoint events, and that $P(E) > 0$ and $P(F) > 0$. Can the events E and F be independent?

No. As $P(E) > 0$ and $P(F) > 0$ we have that $P(E)P(F) > 0$. However, since $E \cap F = \emptyset$ we also have that $P(E \cap F) = P(\emptyset) = 0$. Since $P(E \cap F) \neq P(E)P(F)$ we conclude that the events are not independent. ■

Summary of Independence

1. E and F are independent if and only if $P(E \cap F) = P(E)P(F)$.
2. Independence is sometimes assumed and sometimes derived.
3. Disjoint events with positive probability are not independent.

5.4 More Examples

Example Which of these two events is more likely?

$E = \{4 \text{ rolls of a die yield at least one } 1\}$; or

$F = \{24 \text{ rolls of two dice yield at least one pair of } (1,1)\}$.

We calculate $P(E)$ and $P(F)$.

1. Each roll of the die is independent from the previous rolls, and so there are 6^4 equally likely outcomes. Of these, 5^4 show no 1s.

So the probability of no 1 showing is $\frac{5^4}{6^4} \approx 0.4823$.

So $P(E)$, the probability of at least one 1 showing, is $1 - \frac{5^4}{6^4} \approx 1 - 0.4823 = 0.5177$.

2. There are 36^{24} equally likely outcomes here. Of these, 35^{24} don't show a (1,1).

So the probability of no (1,1) is $\frac{35^{24}}{36^{24}} \approx 0.5086$

So $P(F)$, the probability of at least one (1,1), is $\approx 1 - 0.5086 = 0.4914$

Hence $P(E) \approx 0.5177 > \frac{1}{2} > 0.4914 \approx P(F)$. ■

Example There is a 1% probability for a hard drive to crash. Therefore, it has two backups, each having a 2% probability to crash, and all three components are independent of each other. The stored information is lost only in the event that all three devices crash. What is the probability that the information is saved?

Start by organising and labelling the events. Denote

$H = \{\text{hard drive crashes}\}$

$B_1 = \{\text{first backup crashes}\}$

$B_2 = \{\text{second backup crashes}\}$

In the wording, we are given that H, B_1 and B_2 are independent and

$$P(H) = 0.01, \quad P(B_1) = 0.02, \quad P(B_2) = 0.02$$

Then, applying rules of complements and intersection for independent events we have

$$\begin{aligned} P(\text{saved}) &= 1 - P(\text{lost}) = 1 - P(H \cap B_1 \cap B_2) \\ &= 1 - P(H)P(B_1)P(B_2) \\ &= 1 - (0.01)(0.02)(0.02) = 1 - 0.000004 = 0.999996 \end{aligned}$$







■

5.4.1 Joint events

So far we have only considered a single outcome or event. We can extend the ideas seen so far for joint events.

For example, consider tossing a coin and rolling a die. We would consider each of the 12 possible combinations of Head/Tail and die value as equally likely.

So we can construct a **probability table**:

							
H	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
T	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

From this table we can calculate the probability of *any* event we might be interested in, simply by adding up the probabilities of all the elementary events it contains.

For example, the event of getting a head on the coin

$$\{H\} = \{(H, \text{1 dot}), (H, \text{2 dots}), \dots, (H, \text{6 dots})\}$$

has probability

$$\begin{aligned} P(\{H\}) &= P(\{(H, \text{1 dot})\}) + P(\{(H, \text{2 dots})\}) + \dots + P(\{(H, \text{6 dots})\}) \\ &= \frac{1}{12} + \frac{1}{12} + \dots + \frac{1}{12} \\ &= \frac{1}{2}. \end{aligned}$$

Notice the two experiments satisfy our probability definition of independence, since for example

$$P(\{(H, \text{6 dots})\}) = \frac{1}{12} = \frac{1}{2} \times \frac{1}{6} = P(\{H\}) \times P(\{\text{6 dots}\}).$$

A crooked die called a *top* has the same faces on opposite sides.

Suppose we have two dice, one normal and one which is a top with opposite faces numbered \square , \blacksquare , or \boxtimes .

Now suppose we first flip the coin. If it comes up heads, we roll the normal die; tails, and we roll the top.

To calculate the probability table easily, we notice that this is equivalent to the previous game using one normal die except with the change after tails that a roll of a \square is relabelled as a \boxtimes , $\blacksquare \rightarrow \square$, $\boxtimes \rightarrow \square$. So we can just merge those probabilities in the tails row.

	\square	\blacksquare	\boxtimes	\boxtimes	\blacksquare	\boxtimes
H	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
T	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0
	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$

The probabilities of the different outcomes of the dice change according to the outcome of the coin toss. And note, for example,

$$P(\{(H, \boxtimes)\}) = \frac{1}{12} \neq \frac{1}{24} = \frac{1}{2} \times \frac{1}{12} = P(\{H\}) \times P(\{\boxtimes\}).$$

So the two experiments are now *dependent*.

5.5 Conditional Probability

Assuming that $P(F) > 0$, we define the conditional probability of E given that F has occurred as follows:

Definition 5.5.1. If $P(F) > 0$ then the **conditional probability** of E given F is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Note If E and F are *independent*, then

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)P(F)}{P(F)} = P(E).$$

Example Suppose a normal die is rolled once.

Questions

Q1) What is the probability of $E = \{\text{the die shows a } \blacksquare\}$?

Q2) What is the probability of $E = \{\text{the die shows a } \blacksquare\}$ given we know $F = \{\text{the die shows an odd number}\}$?

Solutions

$$S1) P(E) = \frac{\text{Number of ways a } \square \text{ can come up}}{\text{Total number of possible outcomes}} = \frac{1}{6}.$$

S2) Now the set of possible outcomes is just $F = \{\square, \square, \square\}$.

$$\text{So } P(E|F) = \frac{\text{Number of ways a } \square \text{ can come up}}{\text{Total number of possible outcomes}} = \frac{1}{3}.$$

$$\text{Note } P(F) = \frac{1}{2} \text{ and } E \cap F = E, \text{ and hence we have } P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Now suppose we roll two normal dice, one from each hand. Then the sample space comprises all of the ordered pairs of dice values

$$S = \{(\square, \square), (\square, \square), \dots, (\blacksquare, \blacksquare)\}.$$

Let E be the event that the die thrown from the left hand will show a larger value than the die thrown from the right hand.

$$P(E) = \frac{\# \text{ outcomes with left value} > \text{right}}{\text{total} \# \text{ outcomes}} = \frac{15}{36}.$$

Suppose we are now informed that an event F has occurred, where

$$F = \{\text{the value of the left hand die is } \blacksquare\}$$

How does this change the probability of E occurring?

Well since F has occurred, the only sample space elements which could have possibly occurred are exactly those elements in $F = \{(\blacksquare, \square), (\blacksquare, \square), (\blacksquare, \square), (\blacksquare, \square), (\blacksquare, \square), (\blacksquare, \blacksquare)\}$.

Similarly the only sample space elements in E that could have occurred now must be in $E \cap F = \{(\blacksquare, \square), (\blacksquare, \square), (\blacksquare, \square), (\blacksquare, \square)\}$.

So our revised probability is

$$\frac{\# \text{ outcomes with left value } \blacksquare > \text{right}}{\text{total} \# \text{ outcomes } (\blacksquare, \cdot)} = \frac{4}{6} = \frac{P(E \cap F)}{P(F)} \equiv P(E|F).$$

■

In both examples, we considered the probability of an event E , and then reconsidered what this probability would be if we were given the knowledge that F had occurred. What we did was replace the sample space S by F , and the event E was replaced by $E \cap F$. So originally, we had

$$P(E) = P(E|S) = \frac{P(E \cap S)}{P(S)} \text{ (since } E \cap S = E, \text{ and } P(S) = 1 \text{ by def of a prob measure.)}$$

So we can think of probability conditioning as a shrinking of the sample space, with events replaced by their intersections with the reduced space and a consequent rescaling of probabilities.

Warning! It is generally the case that $P(E|F) \neq P(F|E)$. People are confused by this all the time. For example, the probability of spots given you have measles is 1, but the probability

that you have measles given that you have spots is not 1. In this case, the difference between $P(E|F)$ and $P(F|E)$ is obvious but there are cases where it is less obvious.

Example A medical test for a disease D has outcomes $+$ and $-$. The probabilities are

	D	\bar{D}
$+$	0.009	0.099
$-$	0.001	0.891

Using the definition of conditional probability, we have

$$P(+|D) = \frac{P(+ \cap D)}{P(D)} = \frac{0.009}{0.009 + 0.001} = 0.9$$

and

$$P(-|\bar{D}) = \frac{P(- \cap \bar{D})}{P(\bar{D})} = \frac{0.891}{0.891 + 0.099} \approx 0.9$$

Apparently, the test is fairly accurate. Sick people yield a positive 90% of the time and healthy people yield a negative about 90% of the time.

Now suppose you go for a test and get a positive. What is the probability you have the disease? Most people would answer 0.9. The correct answer is

$$P(D|+) = \frac{P(+ \cap D)}{P(+)} = \frac{0.009}{0.009 + 0.099} \approx 0.08$$

which is much less than 0.9.

The lesson here is that you need to compute the answer numerically. Do not trust your intuition!

Also note that $P(+|D) \neq P(D|+)$ ■

5.5.1 Conditional Independence

Earlier we met the concept of independence of events according to a probability measure P . We can now extend that idea to conditional probabilities since $P(\cdot|F)$ is itself a measure obeying the axioms of probability.

For three events E_1 , E_2 and F , the event pair E_1 and E_2 are said to be **conditionally independent** given F if and only if $P(E_1 \cap E_2|F) = P(E_1|F)P(E_2|F)$.¹

5.5.2 Bayes' Theorem

Before we introduce Bayes' theorem, we require a preliminary result.

¹This is sometimes written $E_1 \perp E_2|F$.

Theorem 5.6 (The Theorem of Total Probability). *Let E_1, \dots, E_k be a partition on S . Then, for any event $F \subseteq S$, we have*

$$P(F) = \sum_{i=1}^k P(F|E_i)P(E_i)$$

Proof. Define $C_j = F \cap E_j$. Note that C_1, \dots, C_k are disjoint and that $F = \bigcup_{j=1}^k C_j$. Hence,

$$P(F) = \underbrace{P\left(\bigcup_{j=1}^k C_j\right)}_{\text{by def. of probability measure}} = \sum_{j=1}^k P(C_j) = \sum_{j=1}^k P(F \cap E_j) = \sum_{j=1}^k P(F|E_j)P(E_j)$$

□

A simple use of this theorem is as follows: for any events E and F in S , note that $\{F, \bar{F}\}$ form a partition of S . So by the law of total probability we have

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap \bar{F}) \\ &= P(E|F)P(F) + P(E|\bar{F})P(\bar{F}). \end{aligned}$$

Theorem 5.7 (Bayes' Theorem). *Let E_1, \dots, E_k be a partition on S such that $P(E_i) > 0$ for each i . If $P(F) > 0$ then, for each $i = 1, \dots, k$, we have*

$$P(E_i|F) = \frac{P(F|E_i)P(E_i)}{P(F)} = \frac{P(F|E_i)P(E_i)}{\sum_{j=1}^k P(F|E_j)P(E_j)}$$

Proof. Starting with the LHS, we have

$$P(E_i|F) = \frac{P(E_i \cap F)}{P(F)} = \frac{P(F|E_i)P(E_i)}{P(F)} = \frac{P(F|E_i)P(E_i)}{\sum_{j=1}^k P(F|E_j)P(E_j)}$$

□

Note We will often deal with probabilities of single events, intersection of events and conditional events. To this end we will refer to

- probabilities of the form $P(E|F)$ as **conditional probabilities**;
- probabilities of the form $P(E \cap F)$ as **joint probabilities**;
- probabilities of the form $P(E)$ as **marginal probabilities**.

5.7.1 More Examples

Example A box contains 5000 VLSI chips, 1000 from company X and 4000 from Y. 10% of the chips made by X are defective and 5% of those made by Y are defective. If a randomly chosen chip is found to be defective, find the probability that it came from company X.

Let E = “chip was made by X”;

let F = “chip is defective”.

First of all, which probabilities have we been given?

The statement

“A box contains 5000 VLSI chips, 1000 from company X and 4000 from Y.”

$$\implies P(E) = \frac{1000}{5000} = 0.2, \quad \text{and} \quad P(\bar{E}) = \frac{4000}{5000} = 0.8.$$

and

“10% of the chips made by X are defective and 5% of those made by Y are defective.”

$$\implies P(F|E) = 10\% = 0.1, \quad \text{and} \quad P(F|\bar{E}) = 5\% = 0.05.$$

We have enough information to construct the probability table

	E	\bar{E}	
F	0.02	0.04	0.06
\bar{F}	0.18	0.76	0.94
	0.2	0.8	

The law of total probability has enabled us to extract the marginal probabilities $P(F)$ and $P(\bar{F})$ as 0.06 and 0.94 respectively.

So by Bayes Theorem we can calculate the conditional probabilities. In particular, we want

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.02}{0.06} = \frac{1}{3}.$$

■

Example Kidney stones are small (< 2cm diam) or large (> 2cm diam). Treatment can succeed or fail. The following data were collected from a sample of 700 patients with kidney stones.

	Success (S)	Failure (\bar{S})	
Large (L)	247	96	343
Small (\bar{L})	315	42	357
Total	562	138	700

For a patient randomly drawn from this sample, what is the probability that the outcome of treatment was successful, given the kidney stones were large?

Clearly we can get the answer directly from the table by ignoring the small stone patients

$$P(S|L) = \frac{247}{343}$$

or we can go the long way round:

$$P(L) = \frac{343}{700}, \quad P(S \cap L) = \frac{247}{700},$$

$$P(S|L) = \frac{P(S \cap L)}{P(L)} = \frac{\frac{247}{700}}{\frac{343}{700}} = \frac{247}{343}.$$

■

Example A multiple choice question has c available choices. Let p be the probability that the student knows the right answer, and $1 - p$ that he does not. When he doesn't know, he chooses an answer at random. Given that the answer the student chooses is correct, what is the probability that the student knew the correct answer?

Let A be the event that the question is answered correctly;

let K be the event that the student knew the correct answer.

Then we require $P(K|A)$.

By Bayes Theorem

$$P(K|A) = \frac{P(A|K)P(K)}{P(A)}$$

and we know $P(A|K) = 1$ and $P(K) = p$, so it remains to find $P(A)$.

By the partition rule, we have

$$P(A) = P(A|K)P(K) + P(A|\bar{K})P(\bar{K})$$

and since $P(A|\bar{K}) = \frac{1}{c}$, this gives

$$P(A) = 1 \times p + \frac{1}{c} \times (1 - p).$$

Hence

$$P(K|A) = \frac{p}{p + \frac{1-p}{c}} = \frac{cp}{cp + 1 - p}.$$

Notice that the larger c is, the greater the probability that the student knew the answer, given that they answered correctly. ■

Example Measurements at the North Carolina Super Computing Center (NCSC) on a certain day showed that 15% of the jobs came from Duke, 35% from UNC, and 50% from NC State

University. Suppose that the probabilities that each of these jobs is a multitasking job is 0.01, 0.05, and 0.02 respectively.

Questions

- Q1) Find the probabilities that a job chosen at random is a multitasking job.
 Q2) Find the probability that a randomly chosen job comes from UNC, given that it is a multitasking job.

Solutions Let

U_i = "job is from university i ", $i = 1, 2, 3$ for Duke, UNC, NC State respectively; and

M = "job uses multitasking".

S1) We want to find $P(M)$. Since U_1, U_2, U_3 form a partition we have

$$\begin{aligned} P(M) &= P(M|U_1)P(U_1) + P(M|U_2)P(U_2) + P(M|U_3)P(U_3) \\ &= 0.01 \times 0.15 + 0.05 \times 0.35 + 0.02 \times 0.5 = 0.029. \end{aligned}$$

S2) We want to find the conditional probability $P(U_2|M)$.

$$P(U_2|M) = \frac{P(M|U_2)P(U_2)}{P(M)} = \frac{0.05 \times 0.35}{0.029} = 0.603.$$

■

Example A new HIV test is claimed to correctly identify 95% of people who are really HIV positive and 98% of people who are really HIV negative. Is this acceptable?

If only 1 in a 1000 of the population are HIV positive, what is the probability that someone who tests positive actually has HIV?

Solution: Let

H = "has the HIV virus"; and

T = "test is positive"

We have been given $P(T|H) = 0.95$, $P(\bar{T}|\bar{H}) = 0.98$ and $P(H) = 0.001$. We wish to find $P(H|T)$.

$$\begin{aligned} P(H|T) &= \frac{P(T|H)P(H)}{P(T|H)P(H) + P(T|\bar{H})P(\bar{H})} \\ &= \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.02 \times 0.999} \\ &= 0.045 \end{aligned}$$

That is, less than 5% of those who test positive really have HIV.

If the HIV test shows a positive result, the individual might wish to retake the test. Suppose that the results of a person retaking the HIV test are conditionally independent given HIV status (clearly two results of the test would certainly not be *unconditionally* independent).

If the test again gives a positive result, what is the probability that the person actually has HIV?

Solution: Let $T_i = "i^{\text{th}} \text{ test is positive}"$.

$$\begin{aligned} P(H|T_1 \cap T_2) &= \frac{P(T_1 \cap T_2|H)P(H)}{P(T_1 \cap T_2)} \\ &= \frac{P(T_1 \cap T_2|H)P(H)}{P(T_1 \cap T_2|H)P(H) + P(T_1 \cap T_2|\bar{H})P(\bar{H})} \\ &= \frac{P(T_1|H)P(T_2|H)P(H)}{P(T_1|H)P(T_2|H)P(H) + P(T_1|\bar{H})P(T_2|\bar{H})P(\bar{H})} \end{aligned}$$

by conditional independence.

Since $P(T_i|H) = 0.95$ and $P(T_i|\bar{H}) = 0.02$,

$$P(H|T_1 \cap T_2) = \frac{0.95 \times 0.95 \times 0.001}{0.95 \times 0.95 \times 0.001 + 0.02 \times 0.02 \times 0.999} \approx 0.693.$$

So almost a 70% chance after taking the test twice and both times showing as positive. For three times, this goes up to 99%. ■

Example Question: I divide my emails into 3 categories: $A_1 = \text{"spam"}$, $A_2 = \text{"reply today"}$ and $A_3 = \text{"reply later"}$. From previous experience I find that $P(A_1) = 0.5$, $P(A_2) = 0.1$ and $P(A_3) = 0.4$. Let B be the event that the email contains the word "trial". From previous experience, I find that $P(B|A_1) = 0.9$, $P(B|A_2) = 0.05$ and $P(B|A_3) = 0.05$. I receive an email with the word "trial". What is the probability that it is spam?

Solution: Using Bayes' theorem we have

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{(0.9)(0.5)}{(0.9)(0.5) + (0.05)(0.1) + (0.05)(0.4)} = \frac{0.45}{0.45 + 0.005 + 0.02} = \frac{0.45}{0.475} \end{aligned}$$

■

Summary of Conditional Probability

1. If $P(F) > 0$ then

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

2. $P(\cdot|F)$ satisfies the axioms of probability, for fixed F . However, in general, $P(E|\cdot)$ does not satisfy the axioms of probability, for fixed E .
3. In general, $P(E|F) \neq P(F|E)$.
4. E and F are independent if and only if $P(E|F) = P(E)$.