3.1 Introduction

We often wish to determine probabilities of events when we have some partial information about the outcome of an experiment. For example, if we know that a person has tested positive for a certain virus, what is the probability that they actually do have the virus? Such probabilities are called *conditional probabilities*.

In fact, conditional probabilities can sometimes be used to determine probabilities of interest even when no partial information is available. We will study a technique called *conditioning*, which is used extensively in probability and stochastic processes.

We also introduce the concept of *independence* in this chapter. Roughly speaking, two events are independent if knowledge that one of them has occurred does not change the probability that the other occurs. It is common for people to mistakenly believe that events are independent when they are not (or vice versa), and this often leads to spurious claims. Fortunately, our mathematical approach will make the distinction clear!

3.2 Conditional Probabilities

Consider the experiment of flipping a fair coin five times.

- What is the probability that the last flip comes up heads? $\frac{1}{2}$
- Would your answer change if you were told that the first four flips came up heads?

No, the last flip should not be affected by what happened in the first four flips.

- What if you were told instead that four of the five flips came up heads?

In this case, there are five possible outcomes: THHHH, HTHHH, HHTHH, HHHTH, or HHHHT. These outcomes are all equally likely, so the probability that the last flip came up heads (given this info) is $\frac{4}{5}$.

If we let E denote the event that the last flip results in heads, and F denote the event that four of the five flips result in heads, then the last question asked for...

$$P(E \mid F)$$

vertical bar is read "given"

The conditional probability of E given F.

Example 3.2.1. Consider the experiment of rolling two fair dice.

- Let E be the event that the sum of the numbers on the dice is 6.
- Let F be the event that the number on the first die is a 2.

Find the following conditional probabilities.

(a) $P(E \mid F)$

First Die

Second Die

	⚀	⚁	⚂	⚃	⚄	⚅
⚀		3				
⚁		4				
⚂		5				
⚃		6				
⚄		7				
⚅		8				

Since we are assuming that F occurred, the outcome is one of the six highlighted squares. These six outcomes are equally likely, and the sum of the dice is 6 in only one of them, so

$$P(E \mid F) = \frac{1}{6}$$

(b) $P(F \mid E)$

	⚀	⚁	⚂	⚃	⚄	⚅
⚀					6	
⚁				6		
⚂			6			
⚃		6				
⚄	6					
⚅						

There are 5 equally likely outcomes in E, and only one of these is in F, so

$$P(F \mid E) = \frac{1}{5}$$

In the examples above, we determined conditional probabilities by considering a *reduced sample space*. Essentially, if each outcome of a finite sample space S is equally likely, and the event F has occurred, then we can treat F as the new sample space, and each outcome in F is equally likely.

We now develop a different way to calculate the conditional probability $P(E \mid F)$, for any events E and F. Let's think about how this conditional probability relates to some other probabilities.

- The event F occurs with probability $P(F) = \frac{1}{2}$
- Given that the event F occurs, the event E will occur with probability $P(E \mid F) = \frac{1}{4}$
- So it stands to reason that *both* events will occur with probability $P(F) \cdot P(E \mid F)$
- But the probability that both events occur can also be written as $P(E \cap F)$

Therefore, we should have: $P(E \cap F) = P(F) \cdot P(E \mid F)$

In fact, we use this equation to define conditional probability. (Some authors consider this a fourth axiom of probability instead of a definition.)

Definition 3.2.2. If $P(F) > 0$, then

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

Example 3.2.3. Use Definition 3.2.2 to verify your answers to Example 3.2.1. Remember that

- E is the event that the numbers on the dice sum to 6; and
- F is the event that the number on the first die is a 2.

(a) $P(E \mid F) = \dfrac{P(E \cap F)}{P(F)} = \dfrac{\frac{1}{36}}{\frac{1}{6}} = \dfrac{6}{36} = \dfrac{1}{6}$

(b) $P(F \mid E) = \dfrac{P(F \cap E)}{P(E)} = \dfrac{\frac{1}{36}}{\frac{5}{36}} = \dfrac{1}{5}$

In some situations, conditional probabilities are easy to find directly – especially when a sequence of experiments is performed, and the outcome of each successive experiment determines the probabilities in the next experiment. In such situations, the equation

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

can be rearranged to find the probability of an intersection of events, as follows:

$$P(E \cap F) = P(F) \cdot P(E \mid F)$$

Example 3.2.4. An urn initially contains 1 red ball and 3 black balls. Each time a ball is selected, its colour is noted and it is replaced in the urn along with 2 more balls of the same colour.

(a) What is the probability that the first two balls selected are both red?

Let R_i be the event that the i^{th} ball drawn is red (for $i = 1, 2$)

Then we want

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2 \mid R_1)$$

$$= \frac{1}{4} \cdot \frac{3}{6}$$

we drew a red ball first, then replaced it and added two more red balls.

$$= \frac{1}{8}$$

(b) What is the probability that the first two balls selected are different colours?

First find the probability that the first two balls are black:

$$P(R_1^c \cap R_2^c) = P(R_1^c) \cdot P(R_2^c \mid R_1^c)$$

$$= \frac{3}{4} \cdot \frac{5}{6}$$

$$= \frac{5}{8}$$

Therefore, the probability that the first two balls are different colours is

$$1 - P(R_1 \cap R_2) - P(R_1^c \cap R_2^c)$$

$$= 1 - \frac{1}{8} - \frac{5}{8} = \frac{1}{4}$$

The equation

$$P(E \cap F) = P(E)\,P(F \mid E) \quad \text{or} \quad P(F)\,P(E \mid F)$$

can be applied repeatedly to yield the following result.

Proposition 3.2.5 (The multiplication rule)**.**

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_1 \cap E_2) \cdots P(E_n \mid E_1 \cap \cdots \cap E_{n-1})$$

Example 3.2.6. An urn contains 5 red balls and 7 black balls. Four balls are randomly selected from the urn, one after another, without replacement.

(a) What is the probability that the first two balls are red and the last two are black?

Let R_i be the event that the i^{th} ball is red (for $i = 1, 2, 3, 4$). Then we want

$$P(R_1 \cap R_2 \cap R_3^c \cap R_4^c) = P(R_1) \cdot P(R_2 \mid R_1) \cdot P(R_3^c \mid R_1 \cap R_2) \cdot P(R_4^c \mid R_1 \cap R_2 \cap R_3^c)$$

$$= \frac{5}{12} \cdot \frac{4}{11} \cdot \frac{7}{10} \cdot \frac{6}{9}$$

$$= \frac{7}{99}$$

(b) What is the probability that the first two are black and the last two are red?

$$P(R_1^c \cap R_2^c \cap R_3 \cap R_4) = \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{4}{9}$$

$$= \frac{7}{99}$$

(c) What is the probability that exactly two of the four balls are red?

This probability is

$$\frac{\binom{5}{2}\binom{7}{2}}{\binom{12}{4}} = \frac{42}{99}$$

Example 3.2.7. Ernie has started doggy agility training, and in one competition, he needs to leap over three increasingly difficult barriers. If he fails to make it over any of the three barriers, then he is immediately eliminated from the competition.

- The probability that he makes it over the first barrier is 0.8.
- Given that he makes it over the first barrier, the probability that he makes it over the second barrier is 0.6.
- Given that he makes it over the first two barriers, the probability that he makes it over the third barrier is 0.5.

(a) What is the probability that he makes it over all three barriers?

Let E_i be the event that he makes it over the first i barriers. (So $E_3 \subseteq E_2 \subseteq E_1$.)

Then $$P(\underbrace{E_1 \cap E_2 \cap E_3}_{E_3}) = P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid \underbrace{E_1 \cap E_2}_{E_2})$$

$$= 0.8 \cdot 0.6 \cdot 0.5$$

$$= 0.24$$

(b) Given that he is eliminated at some point, what is the probability that he was eliminated trying to leap over the second barrier?

We want

$$P(E_1 \cap E_2^c \mid E_3^c) = \frac{P(E_1 \cap E_2^c \cap E_3^c)}{P(E_3^c)}$$

$$= \frac{P(E_1 \cap E_2^c)}{1 - P(E_3)}$$

$$= \frac{P(E_1) \cdot P(E_2^c \mid E_1)}{1 - P(E_3)}$$

$$= \frac{0.8 \cdot (1 - 0.6)}{1 - 0.24}$$

$$= \frac{0.32}{0.76} = \frac{8}{19}$$

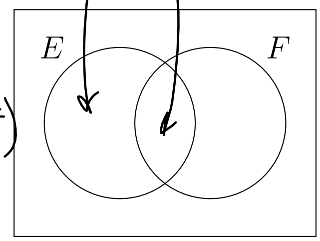
3.3 Bayes' Formula

Example 3.3.1. In order to decide where to take his small dog Ernie for a walk, Lucas flips a coin. If it lands on heads, then they will go to Battle Bluff, where the chance of rain is 20%. Otherwise, they will go to Buse Hill, where the chance of rain is 40%.

- Let E be the event that Ernie gets rained on.
- Let F be the event that the coin comes up heads.

Find the probability that Ernie gets rained on if...

$E = (E \cap F^c) \cup (E \cap F)$



S

(a) The coin is fair.

We have
$$P(E) = P(E \cap F) + P(E \cap F^c)$$
$$= P(F)P(E|F) + P(F^c)\cdot P(E|F^c)$$
$$= \frac{1}{2}(0.2) + \frac{1}{2}(0.4)$$
$$= 0.3$$

(b) The coin comes up heads with probability 3/4.

$$P(E) = P(F)\cdot P(E|F) + P(F^c)\cdot P(E|F^c)$$
$$= \frac{3}{4}\cdot(0.2) + \frac{1}{4}(0.4)$$
$$= 0.25$$

Notice that we are given two conditional probabilities in this problem statement – we are told the chance of rain given either outcome of the coin flip. It seems reasonable that the probability that Ernie gets rained on should be a weighted average of these two conditional probabilities, with the weight on each conditional probability being the probability of the corresponding outcome of the coin flip.

We have essentially proven the following proposition. When we use this formula to compute the probability of E, we say that we are *conditioning* on F.

Proposition 3.3.2. *Let E and F be events. Then*

$$P(E) = P(E|F) \cdot P(F) + P(E|F^c) \cdot P(F^c)$$

Example 3.3.3. On snowy days, Lucas is late with probability 0.4. On snow-free days, he is late with probability 0.2. Suppose that it will snow tomorrow with probability 0.7.

(a) Find the probability that Lucas is late tomorrow.

Let E be the event that Lucas is late tomorrow.
Let F be the event that it snows tomorrow.

$$\begin{aligned} P(E) &= P(E|F) \cdot P(F) + P(E|F^c) \cdot P(F^c) \\ &= (0.4)(0.7) + (0.2)(1-0.7) \\ &= 0.28 + 0.06 \\ &= 0.34 \end{aligned}$$

(b) Given that Lucas was late, what is the conditional probability that it was snowy?

We want $P(F|E) = \dfrac{P(F \cap E)}{P(E)} = \dfrac{P(F) \cdot P(E|F)}{P(E)}$

$$= \frac{(0.7)(0.4)}{0.34} = \frac{14}{17}$$

Notice that we can use the definition of conditional probability twice to "trade off" a conditional probability for the opposite conditional probability:

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

This is a very useful observation!

Example 3.3.4. A test correctly detects the presence of a disease 98% of the time, and returns a false positive result (i.e., incorrectly detects the presence of the disease in a healthy patient) only 1% of the time.

(a) Suppose that 3% of the population has the disease. What is the probability that a person has the disease, given that they test positive?

Let D be the event that the person has the disease.
Let E be the event That the person tests positive.

We want
$$P(D|E) = \frac{P(D \cap E)}{P(E)} = \frac{P(E|D) \cdot P(D)}{P(E)}$$
$$= \frac{P(E|D) \cdot P(D)}{P(E|D) \cdot P(D) + P(E|D^c) \cdot P(D^c)}$$
$$= \frac{0.98 \cdot 0.03}{0.98 \cdot 0.03 + 0.01\,(0.97)}$$
$$\approx 0.752$$

(b) What if only 1 in 500 people have the disease?

Now
$P(D) = \frac{1}{500} = 0.002$
$$P(D|E) = \frac{P(E|D) \cdot P(D)}{P(E|D) \cdot P(D) + P(E|D^c) \cdot P(D^c)}$$
$$= \frac{0.98\,(0.002)}{0.98\,(0.002) + 0.01\,(0.998)} \approx 0.164$$

Example 3.3.5. Twins can either be identical (when a single egg is fertilized and then splits into two genetically identical parts) or fraternal (when two different eggs are fertilized separately). If 64% of all twins born have the same natal sex, then approximately what percentage of twins are identical?

sex at birth
we'll assume 50% male, 50% female

Let I be the event that the twins are identical.

Let S be the event that the twins have the same sex.

We have
$$P(S) = P(S|I) \cdot P(I) + P(S|I^c) \cdot P(I^c)$$
$$\implies 0.64 = 1 \cdot P(I) + \frac{1}{2} \cdot P(I^c)$$
$$\implies 0.64 = P(I) + \frac{1}{2}\left[1 - P(I)\right]$$
$$\implies 0.64 = \frac{1}{2} P(I) + \frac{1}{2}$$
$$\implies 0.14 = \frac{1}{2} P(I) \implies P(I) = 0.28$$

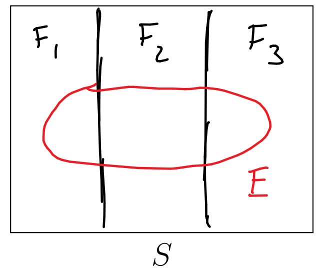
Example 3.3.6. There are three urns on a table.

- The first urn contains one red ball and two black balls.
- The second urn contains one red ball and three black balls.
- The third urn contains three red balls and one black ball.

Suppose that one of the three urns is randomly selected, and then a ball is randomly selected from the chosen urn.

(a) What is the probability that a red ball is selected?

Let F_i be the event that the ith urn is selected, for $i = 1, 2, 3$.
Let E be the event that the selected ball is red.



We have $P(E) = P(E \cap F_1) + P(E \cap F_2) + P(E \cap F_3)$ by Axiom 3

$$= P(E|F_1) \cdot P(F_1) + P(E|F_2) \cdot P(F_2) + P(E|F_3) \cdot P(F_3)$$

$$= \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{1}{3}$$

$$= \frac{1}{9} + \frac{1}{12} + \frac{1}{4}$$

$$= \frac{4}{9}$$

(b) Given that a red ball was selected, what is the conditional probability that the third urn was selected?

$$P(F_3|E) = \frac{P(F_3 \cap E)}{P(E)} = \frac{P(E|F_3) \cdot P(F_3)}{P(E)}$$

$$= \frac{\frac{3}{4} \cdot \frac{1}{3}}{\frac{4}{9}}$$

$$= \frac{1}{4} \cdot \frac{9}{4} = \frac{9}{16}$$

Let $F_1, F_2, \ldots, F_n$ be events.

- Recall that $F_1, F_2, \ldots, F_n$ are *mutually exclusive* if...

$$F_i \cap F_j = \emptyset \quad \text{for all} \quad i \neq j$$

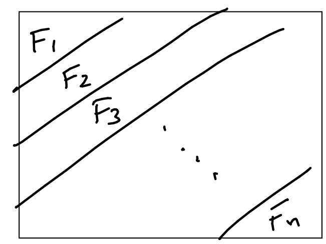
- We say that $F_1, F_2, \ldots, F_n$ are *exhaustive* if...

$$\bigcup_{i=1}^{n} F_i = S$$

- So $F_1, F_2, \ldots, F_n$ are both mutually exclusive and exhaustive if and only if...

exactly one of them occurs.

(Or in other words, they partition the sample space.)



S

Proposition 3.3.7 (The law of total probability)**.** *Let $F_1, F_2, \ldots, F_n$ be mutually exclusive and exhaustive events. Then*

$$P(E) = P(E \mid F_1) \cdot P(F_1) + P(E \mid F_2) \cdot P(F_2) + \cdots + P(E \mid F_n) \cdot P(F_n)$$

$$= \sum_{i=1}^{n} P(E \mid F_i) \cdot P(F_i)$$

Remark: Note that for any event F, the events $\underline{\quad F \quad}$ and $\underline{\quad F^c \quad}$ are mutually exclusive and exhaustive. So Proposition 3.3.7 is a direct generalization of Proposition 3.3.2.

Corollary 3.3.8 (Bayes' formula)**.** *Let $F_1, F_2, \ldots, F_n$ be mutually exclusive and exhaustive events. Then*

$$P(F_j \mid E) = \frac{P(E \mid F_j) \cdot P(F_j)}{P(E)}$$

$$= \frac{P(E \mid F_j) \cdot P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i) \cdot P(F_i)}$$

coin 1

Example 3.3.9. Suppose that we have three coins that are identical except that one is fair, one is two-headed, and one is two-tailed. One of the three coins is randomly selected and flipped. If it shows heads, what is the probability that the other side is tails?

coin 2 coin 3

Let C_i be the event that coin i is chosen, and let H be the event that the flip shows heads.

We want
$$P(C_1 \mid H) = \frac{P(C_1 \cap H)}{P(H)} = \frac{P(H\mid C_1)\cdot P(C_1)}{P(H\mid C_1)\cdot P(C_1) + P(H\mid C_2)\cdot P(C_2) + P(H\mid C_3)\cdot P(C_3)}$$
$$= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 0\cdot\frac{1}{3}}$$
$$= \frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}} = \frac{1}{3}$$

Example 3.3.10. An urn contains 5 red balls and 7 black balls. A fair die is rolled, and then that number of balls is randomly chosen from the urn, without replacement.

(a) What is the probability that all of the chosen balls are red?

Let D_i be the event that we roll an i, for $i = 1,2,3,4,5,6$.
Let R " " " " all of the selected balls are red.

Note that D_i's are mutually exclusive & exhaustive; exactly one of them must occur!

$$P(R) = \sum_{i=1}^{6} P(R \mid D_i)\cdot P(D_i) = \frac{5}{12}\cdot\frac{1}{6} + \frac{\binom{5}{2}}{\binom{12}{2}}\cdot\frac{1}{6}$$
$$+ \frac{\binom{5}{3}}{\binom{12}{3}}\cdot\frac{1}{6} + \frac{\binom{5}{4}}{\binom{12}{4}}\cdot\frac{1}{6}$$
$$+ \frac{1}{\binom{12}{5}}\cdot\frac{1}{6} + 0\cdot\frac{1}{6} = ??$$

(b) Given that all of the chosen balls are red, what is the conditional probability that the die landed on 2?

$$P(D_2 \mid R) = \frac{P(D_2 \cap R)}{P(R)} = \frac{P(R\mid D_2)\cdot P(D_2)}{P(R)}$$
$$= \frac{\frac{\binom{5}{2}}{\binom{12}{2}}\cdot\frac{1}{6}}{P(R)}$$

know from above!

$$= ??$$

3.4 Independent Events

Definition 3.4.1. Two events E and F are said to be *independent* if

this is how the book writes intersection

$$P(EF) = P(E \cap F) = P(E) \cdot P(F)$$

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

Observation 3.4.2. If E and F are independent, then

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E) \cdot \cancel{P(F)}}{\cancel{P(F)}} = P(E)$$

In other words, if E and F are independent, then

knowledge that F occurs does not affect the prob. that E occurs.

By a symmetric argument, if E and F are independent, then $P(F \mid E) =$ $P(F)$, and knowledge that E occurs does not affect the prob. that F occurs.

Example 3.4.3. Suppose that a fair coin is flipped 3 times. Determine whether the following pairs of events are independent or dependent.

(a) The event E that the first flip comes up heads and the event F that the second flip comes up heads.

They seem independent...

$P(E) = \frac{1}{2}$, $P(F) = \frac{1}{2}$, and $P(E \cap F) = \frac{1}{4}$.

So $P(E \cap F) = P(E) \cdot P(F)$ ✓

(b) The event E that no flips come up heads and the event F that all three flips come up heads.

They seem dependent...

$P(E) = \frac{1}{8}$, $P(F) = \frac{1}{8}$, and $P(E \cap F) = 0$

So $P(E \cap F) \neq P(E) \cdot P(F)$ ✗

(c) The event E that at least one flip comes up heads and the event F that all three flips come up heads.

They are dependent.

$P(E) = \frac{7}{8}$, $P(F) = \frac{1}{8}$, $P(E \cap F) = \frac{1}{8}$

So $P(E \cap F) \neq P(E) \cdot P(F)$ ✗

Example 3.4.4. Let

- E be the event that the Toronto Maple Leafs win the Stanley Cup next year;
- F be the event that Auston Matthews (the Maple Leafs' current star player) gets seriously injured next year;
- G be the event that the NHL increases the team salary cap by \$10 million next year.

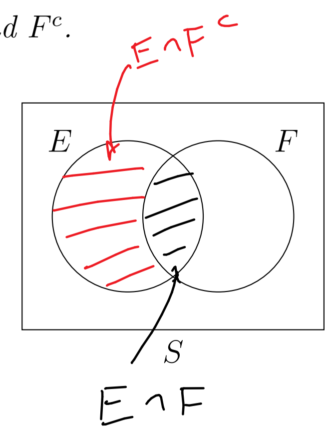
Do you think that the following pairs of events are independent or dependent?

(a) E and F probably dependent

(b) E and G probably dependent

(c) F and G probably independent

Proposition 3.4.5. *If E and F are independent, then so are E and F^c.*



Proof. Suppose E and F are independent,

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

We have

$E = (E \cap F) \cup (E \cap F^c)$, and $E \cap F$ and $E \cap F^c$ are mutually exclusive,

so $P(E) = P(E \cap F) + P(E \cap F^c)$, by Axiom 3.

$\therefore$ We have

$$\begin{aligned} P(E \cap F^c) &= P(E) - P(E \cap F) \\ &= P(E) - P(E) \cdot P(F) \\ &= P(E)\left[1 - P(F)\right] \\ &= P(E)\, P(F^c) \end{aligned}$$

$\therefore$ By the def$^{\underline{n}}$ of independence, we see that E and F^c are independent. $\square$

Definition 3.4.6. Three events E, F, and G are said to be *independent* if

- $P(E \cap F) = P(E) \cdot P(F)$
- $P(E \cap G) = P(E) \cdot P(G)$
- $P(F \cap G) = P(F) \cdot P(G)$
- $P(E \cap F \cap G) = P(E) \cdot P(F) \cdot P(G)$

More generally, events $E_1, E_2, \ldots, E_n$ are said to be *independent* if

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \prod_{j=1}^{r} P(E_{i_j}) \quad \text{for any } i_1 < i_2 < \ldots < i_r$$

Finally, an infinite set of events is said to be *independent* if

any finite subset of these events is independent.

Example 3.4.7. Suppose that we have a coin that comes up heads with some fixed probability p, where $p \in [0,1]$. Suppose that we flip the coin n times.

(a) What is the probability that at least one flip comes up tails?

Let E_i be the event that the i^{th} flip comes up heads. It seems reasonable to assume that the events $E_1, E_2, \ldots, E_n$ are independent. So the probability that all n flips come up heads is

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \ldots \cdot P(E_n) = p^n$$

So the probability that at least one flip comes up tails is $1 - p^n$.

(b) What is the probability that exactly k out of n flips come up heads, where $0 \leq k \leq n$?

There are $\binom{n}{k}$ sequences of flips where k flips come up heads and $n-k$ flips come up tails. For example, if $n = 5$ and $k=3$, then one possible sequence is HTHHT.

k fixed

By the assumed indepence of the events $E_1, E_2, \ldots, E_n$, the probability that any one specific sequence of this type comes up is $p^k(1-p)^{n-k}$. Since there are $\binom{n}{k}$ sequences with k heads, the desired probability is $\binom{n}{k} p^k (1-p)^{n-k}$.

We often consider experiments that consist of a sequence of simple subexperiments, such as flipping a coin repeatedly, rolling a pair of dice repeatedly, a basketball player shooting free throws repeatedly, etc.

- If each subexperiment has the same set of outcomes, then the subexperiments are often called: trials
- We say that a sequence of trials is *independent* if $E_1, E_2, \ldots, E_n, \ldots$ is an independent sequence of events whenever E_i is an event whose occurrence is completely determined by the outcome of the i^{th} trial.

It is often reasonable to assume that a sequence of trials is independent.

Example 3.4.8. Suppose that a pair of dice are rolled repeatedly. When a pair of dice are rolled, the probability p_i of rolling the number i can be determined from the following table.

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

$$p_4 = \frac{3}{36} = \frac{1}{12}$$

$$p_7 = \frac{6}{36} = \frac{1}{6}$$

Let W be the event that the number 4 is rolled before the number 7.

(a) Find $P(\overset{W}{\cancel{E}})$ by first computing the probability of the event $\overset{W}{\cancel{E}}_n$ that neither a 4 nor a 7 is rolled on the first $n-1$ rolls, and a 4 is rolled on the nth roll.

It seems reasonable to assume that the trials are independent. So we have

$$P(W_n) = (1 - p_4 - p_7)^{n-1} \cdot p_4 = \left(\frac{27}{36}\right)^{n-1} \cdot \frac{3}{36} = \left(\frac{3}{4}\right)^{n-1} \cdot \frac{1}{12}$$

Since $W = W_1 \cup W_2 \cup \ldots \cup W_n \cup \ldots = \bigcup_{n=1}^{\infty} W_n$,

and the events $W_1, W_2, \ldots, W_n, \ldots$ are mutually exclusive, we have

$$P(W) = P\left(\bigcup_{n=1}^{\infty} W_n\right) = \sum_{n=1}^{\infty} P(W_n) = \sum_{n=1}^{\infty} \left(\frac{3}{4}\right)^{n-1} \cdot \frac{1}{12}$$

geometric series with initial term a and common ratio r has sum $\frac{a}{1-r}$

$$= \frac{\frac{1}{12}}{1 - \frac{3}{4}}$$

$$= \frac{1}{3}$$

(b) Find $P(\overset{W}{\cancel{A}})$ by conditioning on the outcome of the first roll.

Let F_4 denote the event that we roll a 4 on the first roll.
" F_7 " " " " " " " 7 on the first roll.
" F_{other} " " " " " " neither a 4 nor a 7 on the first roll.

$= F_4^c \cap F_7^c$

Since F_4, F_7, and F_{other} are mutually exclusive and exhaustive, we can condition on their outcome:

$$P(W) = P(W \mid F_4) \cdot P(F_4) + P(W \mid F_7) \cdot P(F_7) + P(W \mid F_{other}) \cdot P(F_{other})$$

$$= 1 \cdot \frac{1}{12} + 0 \cdot \frac{1}{6} + P(W) \cdot \frac{3}{4}$$

Since the trials are independent, the outcome of the first roll does not affect the outcome of any other rolls. So it is like the experiment is restarting!

$$\Rightarrow \quad P(W) = \frac{1}{12} + \frac{3}{4} P(W)$$

$$\Rightarrow \quad \frac{1}{4} P(W) = \frac{1}{12}$$

$$\Rightarrow \quad P(W) = \frac{1}{3}$$

We just found $W_{4,7}$

(c) Let $W_{i,j}$ be the event that the number i is rolled before the number j, where $i \neq j$. Find $P(W_{i,j})$. (Let p_i be the probability of rolling a sum of i.)

By the same argument as above,

$$P(W_{i,j}) = 1 \cdot p_i + P(W_{i,j}) \cdot (1 - p_i - p_j)$$

$$\Rightarrow \quad P(W_{i,j}) = \frac{p_i}{p_i + p_j}$$

3.5 P (· | F) Is a Probability

The main idea of this section is that all of the statements that we've made for probabilities still hold if we make all of the probabilities conditional on some event F (as long as $P(F) > 0$). For instance, we have

$$P(E^c \mid F) = 1 - P(E \mid F)$$

$$P(E_1 \cup E_2 \mid F) = P(E_1 \mid F) + P(E_2 \mid F) - P(E_1 \cap E_2 \mid F)$$

The reason for this is that the function $P(\cdot \mid F)$ satisfies the three axioms of probability.

Proposition 3.5.1. *Consider an experiment with sample space S and probability function P, and let F be an event with $P(F) > 0$. Then*

(a) $0 \le P(E \mid F) \le 1$ (where E is any event)

(b) $P(S \mid F) = 1$

(c) For any infinite sequence $E_1, E_2, \ldots$ of mutually exclusive events, we have $P\left(\bigcup_{i=1}^{\infty} E_i \,\middle|\, F\right) = \sum_{i=1}^{\infty} P(E_i \mid F)$

Recall Proposition 3.3.2, which describes how to compute the probability of an event E_1 by conditioning on a second event E_2:

$$P(E_1) = P(E_1 \mid E_2) \cdot P(E_2) + P(E_1 \mid E_2^c) \cdot P(E_2^c)$$

By Proposition 3.5.1, this equation still holds if we make all probabilities conditional on F. This will give us a way to compute $P(E_1 \mid F)$ by conditioning on E_2. For ease of notation, we write $Q(E) = P(E \mid F)$ for every event E. Then we have

$$P(E_1 \mid F) = Q(E_1) = Q(E_1 \mid E_2) \cdot Q(E_2) + Q(E_1 \mid E_2^c) \cdot Q(E_2^c)$$

Note that

$$Q(E_1 \mid E_2) = \frac{Q(E_1 \cap E_2)}{Q(E_2)} = \frac{P(E_1 \cap E_2 \mid F)}{P(E_2 \mid F)} = \frac{\frac{P(E_1 \cap E_2 \cap F)}{P(F)}}{\frac{P(E_2 \cap F)}{P(F)}}$$

$$= \frac{P(E_1 \cap (E_2 \cap F))}{P(E_2 \cap F)}$$

$$= P(E_1 \mid E_2 \cap F)$$

So from above, we get

$$P(E_1 \mid F) =$$

Again, we are essentially just making all probabilities conditional on F in a statement that we already had.

$$P(E_1 \mid F) = P(E_1 \mid E_2 \cap F) \cdot P(E_2 \mid F) + P(E_1 \mid E_2^c \cap F) \cdot P(E_2^c \mid F)$$

An insurance company believes that people can be divided into two classes: those who are accident-prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas this probability is 0.2 for a person who is not accident prone. Assume that 30% of the population is accident prone.

(a) What is the probability that a new policyholder will have an accident within a year of purchasing the policy?

Let F be the event that the person is accident-prone.
Let A_i be the event that the person has an accident in their ith year of holding the policy.

$$P(A_1) = P(A_1 \mid F) \cdot P(F) + P(A_1 \mid F^c) \cdot P(F^c) = 0.4\,(0.3) + 0.2\,(0.7)$$
$$= 0.26$$

(b) If a new policyholder has an accident within a year of purchasing a policy, what is the probability that they are accident prone?

$$P(F \mid A_1) = \frac{P(F \cap A_1)}{P(A_1)}$$
$$= \frac{P(A_1 \mid F) \cdot P(F)}{P(A_1)}$$
$$= \frac{0.4\,(0.3)}{0.26} = \frac{12}{26} = \frac{6}{13}$$

(c) What is the conditional probability that a new policyholder will have an accident in their second year of policy ownership, given that they had an accident in the first year?

$$P(A_2 \mid A_1) = P(A_2 \mid A_1 \cap F) \cdot P(F \mid A_1) + P(A_2 \mid A_1 \cap F^c) \cdot P(F^c \mid A_1)$$
$$= 0.4 \quad \frac{6}{13} + 0.2 \quad \frac{1}{13}$$
$$= \frac{19}{65}$$

Definition 3.5.2. We say that events E_1 and E_2 are *conditionally independent* given F if

$$P(E_1 \cap E_2 \mid F) = P(E_1 \mid F) \cdot P(E_2 \mid F).$$

This means that the conditional probability that E_1 occurs given F is unchanged by information as to whether or not E_2 occurs. The notion of conditional independence can be extended to more than two events in the same manner as regular independence.

Example 3.5.3. There are $k+1$ coins in a box. When flipped, the ith coin will turn up heads with probability i/k, for $i = 0, 1, 2, \ldots, k$. A coin is randomly selected from the box and is then repeatedly flipped.

If the first n flips all result in heads, what is the probability that the $(n+1)$st flip results in heads?

Let H_n be the event that the first n flips result in heads.
Let C_i be the event that we select coin i, which comes up heads with probability i/k.

$$P(H_{n+1} \mid H_n) = \frac{P(H_{n+1} \cap H_n)}{P(H_n)} = \frac{P(H_{n+1})}{P(H_n)}$$

To find $P(H_n)$ we condition on the outcome of the selected coin, and use the fact that the coin flips are conditionally independent given that coin i is selected:

We have

$$P(H_n) = \sum_{i=0}^{k} P(H_n \mid C_i) \cdot P(C_i)$$

$$= \sum_{i=0}^{k} \left(\frac{i}{k}\right)^n \cdot \frac{1}{k+1}$$

$$= \frac{1}{k+1} \cdot \sum_{i=0}^{k} \left(\frac{i}{k}\right)^n$$

$$\therefore \quad P(H_{n+1} \mid H_n) = \frac{P(H_{n+1})}{P(H_n)} = \frac{\cancel{\frac{1}{k+1}} \cdot \sum_{i=0}^{k} \left(\frac{i}{k}\right)^{n+1}}{\cancel{\frac{1}{k+1}} \cdot \sum_{i=0}^{k} \left(\frac{i}{k}\right)^{n}} = \frac{\sum_{i=0}^{k} \left(\frac{i}{k}\right)^{n+1}}{\sum_{i=0}^{k} \left(\frac{i}{k}\right)^{n}}$$

Note that $\displaystyle \frac{1}{k} \sum_{i=0}^{k} \left(\frac{i}{k}\right)^{n+1} \approx \int_0^1 x^{n+1}\,dx = \left[\frac{x^{n+2}}{n+2}\right]_0^1 = \frac{1}{n+2}$

and $\displaystyle \frac{1}{k} \cdot \sum_{i=0}^{k} \left(\frac{i}{k}\right)^{n} \approx \int_0^1 x^{n}\,dx = \left[\frac{x^{n+1}}{n+1}\right]_0^1 = \frac{1}{n+1}$

$\therefore$ We obtain $\displaystyle P(H_{n+1} \mid H_n) \approx \frac{n+1}{n+2}$ (when k is large).

Summary

- The conditional probability that E occurs given that F occurs is

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

- The above equation can be rearranged to obtain an expression for the probability of the intersection of E and F:

$$P(E \cap F) = P(F) \cdot P(E \mid F)$$

- We can "trade off" one conditional probability for the opposite conditional probability as follows:

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

- The *multiplication rule* says that

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_2 \cap E_1) \cdot \cdots \cdot P(E_n \mid E_{n-1} \cap \ldots \cap E_1)$$

- We can compute $P(E)$ by *conditioning* on whether or not F occurs:

$$P(E) = P(E \mid F) \cdot P(F) + P(E \mid F^c) \cdot P(F^c)$$

- We say that $F_1, F_2, \ldots, F_n$ are
  - *mutually exclusive* if: $F_i \cap F_j = \phi$ for all $i \neq j$
  - *exhaustive* if: $\bigcup_{i=1}^{n} F_i = S$

- If $F_1, F_2, \ldots, F_n$ are mutually exclusive and exhaustive, then we can compute $P(E)$ by *conditioning* on which event occurs:

$$P(E) = \sum_{i=1}^{n} P(E \mid F_i) \cdot P(F_i)$$

  This is called the law of total probability

- If $F_1, F_2, \ldots, F_n$ are mutually exclusive and exhaustive, then

$$P(F_j \mid E) = \frac{P(E \mid F_j) \cdot P(F_j)}{P(E)} = \frac{P(E \mid F_j) \cdot P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i) \cdot P(F_i)}$$

  This is called Bayes' Theorem.

- In words, two events are *independent* if knowledge of the occurrence of one of them does not affect the probability of the other.

- Mathematically, the events E and F are *independent* if $P(E \cap F) = P(E) \cdot P(F)$

- For a fixed event F, the function $P(\cdot \mid F)$ satisfies the three axioms of a probability. So any statement that holds for probabilities still holds if we make all probabilities conditional on F.