Topic 6

Conditional Probability and Independence

One of the most important concepts in the theory of probability is based on the question: *How do we modify the probability of an event in light of the fact that something new is known?* What is the chance that we will win the game now that we have taken the first point? What is the chance that I am a carrier of a genetic disease now that my first child does not have the genetic condition? What is the chance that a child smokes if the household has two parents who smoke? This question leads us to the concept of **conditional probability.**

6.1 Restricting the Sample Space - Conditional Probability

Toss a fair coin 3 times. Let winning be "at least two heads out of three"

HHH	HHT	HTH	HTT
THH	THT	TTH	TTT

Figure 6.1: Outcomes on three tosses of a coin, with the winning event indicated.

If we now know that the first coin toss is heads, then only the top row is possible and we would like to say that the probability of winning is

$$\frac{\#(\text{outcomes that result in a win and also have a heads on the first coin toss})}{\#(\text{outcomes with heads on the first coin toss})}$$
$$= \frac{\#\{\text{HHH, HHT, HTH}\}}{\#\{\text{HHH, HHT, HTH, HTT}\}} = \frac{3}{4}.$$

We can take this idea to create a formula in the case of equally likely outcomes for the statement the **conditional probability of A given B**.

$$\begin{aligned} P(A|B) &= \text{the proportion of outcomes in } A \text{ that are also in } B \\ &= \frac{\#(A \cap B)}{\#(B)} \end{aligned}$$

We can turn this into a more general statement using only the probability, P, by dividing both the numerator and the denominator in this fraction by $\#(\Omega)$.

$$P(A|B) = \frac{\#(A \cap B)/\#(\Omega)}{\#(B)/\#(\Omega)} = \frac{P(A \cap B)}{P(B)} \tag{6.1}$$

We thus take this version (6.1) of the identity as the general definition of conditional probability for any pair of events A and B as long as the denominator $P(B) > 0$.

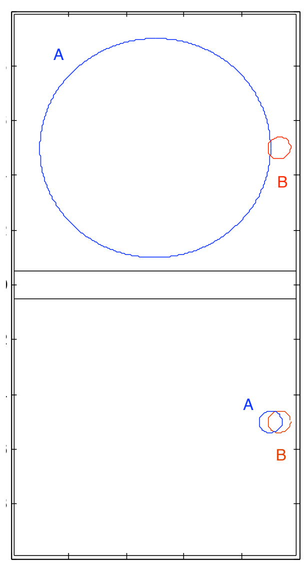

Figure 6.2: Two Venn diagrams to illustrate conditional probability. For the top diagram $P(A)$ is large but $P(A|B)$ is small. For the bottom diagram $P(A)$ is small but $P(A|B)$ is large.

Exercise 6.1. *Pick an event B so that $P(B) > 0$. Define, for every event A,*

$$Q(A) = P(A|B).$$

Show that Q satisfies the three axioms of a probability. In words, a conditional probability is a probability.

Exercise 6.2. *Roll two dice. Find* $P\{\text{sum is 8}|\text{first die shows 3}\}$, *and* $P\{\text{sum is 8}|\text{first die shows 1}\}$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Figure 6.3: Outcomes on the roll of two dice. The event $\{\text{first roll is 3}\}$ is indicated.

Exercise 6.3. *Roll two four-sided dice. With the numbers 1 through 4 on each die, the value of the roll is the number on the side facing downward. Assuming all 16 outcomes are equally likely, find* $P\{\textit{sum is at least 5}\}$, $P\{\textit{first die is 2}\}$ *and* $P\{\textit{sum is at least 5}|\textit{first die is 2}\}$

6.2 The Multiplication Principle

The defining formula (6.1) for conditional probability can be rewritten to obtain **the multiplication principle**,

$$P(A \cap B) = P(A|B)P(B). \tag{6.2}$$

Now, we can complete an earlier problem:

$$\begin{aligned} P\{\text{ace on first two cards}\} &= P\{\text{ace on second card}|\text{ace on first card}\}P\{\text{ace on first card}\} \\ &= \frac{3}{51} \times \frac{4}{52} = \frac{1}{17} \times \frac{1}{13}. \end{aligned}$$

We can continue this process to obtain a **chain rule**:

$$P(A \cap B \cap C) = P(A|B \cap C)P(B \cap C) = P(A|B \cap C)P(B|C)P(C).$$

Thus,

$$\begin{aligned} &P\{\text{ace on first three cards}\} \\ &= P\{\text{ace on third card}|\text{ace on first and second card}\}P\{\text{ace on second card}|\text{ace on first card}\}P\{\text{ace on first card}\} \\ &= \frac{2}{50} \times \frac{3}{51} \times \frac{4}{52} = \frac{1}{25} \times \frac{1}{17} \times \frac{1}{13}. \end{aligned}$$

Extending this to 4 events, we consider the following question:

Example 6.4. *In a urn with b blue balls and g green balls, the probability of green, blue, green, blue (in that order) is*

$$\frac{g}{b+g} \cdot \frac{b}{b+g-1} \cdot \frac{g-1}{b+g-2} \cdot \frac{b-1}{b+g-3} = \frac{(g)_2(b)_2}{(b+g)_4}.$$

Notice that any choice of 2 green and 2 blue would result in the same probability. There are $\binom{4}{2} = 6$ such choices. Thus, with 4 balls chosen without replacement

$$P\{\text{2 blue and 2 green}\} = \binom{4}{2}\frac{(g)_2(b)_2}{(b+g)_4}.$$

Exercise 6.5. *Show that*

$$\binom{4}{2}\frac{(g)_2(b)_2}{(b+g)_4} = \frac{\binom{b}{2}\binom{g}{2}}{\binom{b+g}{4}}.$$

Explain in words why $P\{2 \text{ blue and } 2 \text{ green}\}$ *is the expression on the right.*

We will later extend this idea when we introduce sampling without replacement in the context of the hypergeometric random variable.

6.3 The Law of Total Probability

If we know the fraction of the population in a given state of the United States that has a given attribute - is diabetic, over 65 years of age, has an income of \$100,000, owns their own home, is married - then how do we determine what fraction of the total population of the United States has this attribute? We address this question by introducing a concept - **partitions** - and an identity - **the law of total probability**.

Definition 6.6. *A* **partition** *of the sample space* Ω *is a finite collection of pairwise mutually exclusive events*

$$\{C_1, C_2, \ldots, C_n\}$$

whose union is Ω.

Thus, every outcome $\omega \in \Omega$ belongs to *exactly* one of the C_i. In particular, distinct members of the partition are mutually exclusive. ($C_i \cap C_j = \emptyset$, if $i \neq j$)

If we know the fraction of the population from 18 to 25 that has been infected by the H1N1 influenza A virus in each of the 50 states, then we cannot just average these 50 values to obtain the fraction of this population infected in the whole country. This method fails because it give equal weight to California and Wyoming. The **law of total probability** shows that we should weigh these conditional probabilities by the probability of residence in a given state and then sum over all of the states.

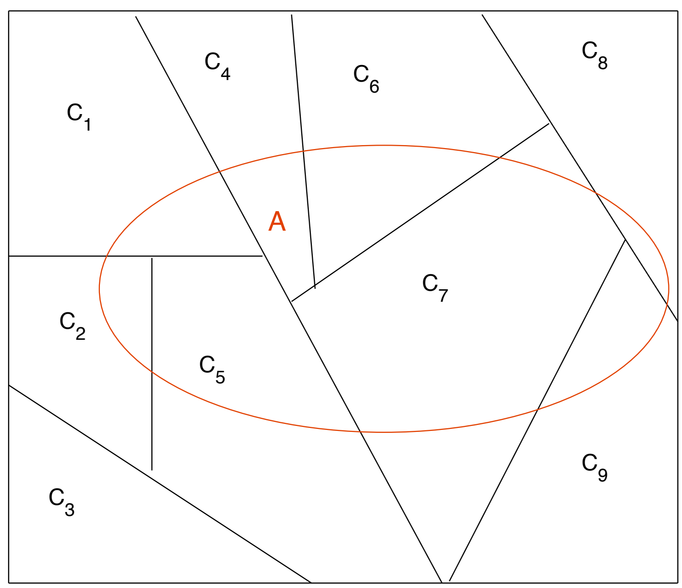

Figure 6.4: A partition $\{C_1 \ldots, C_9\}$ of the sample space Ω. The event A can be written as the union $(A \cap C_1) \cup \cdots \cup (A \cap C_9)$ of mutually exclusive events.

Theorem 6.7 (law of total probability)**.** *Let* P *be a probability on* Ω *and let* $\{C_1, C_2, \ldots, C_n\}$ *be a partition of* Ω *chosen so that* $P(C_i) > 0$ *for all* i. *Then, for any event* $A \subset \Omega$,

$$P(A) = \sum_{i=1}^{n} P(A|C_i)P(C_i). \tag{6.3}$$

Because $\{C_1, C_2, \ldots, C_n\}$ is a partition, $\{A \cap C_1, A \cap C_2, \ldots, A \cap C_n\}$ are pairwise mutually exclusive events. By the distributive property of sets, their union is the event A. (See Figure 6.4.)

To refer the example above the C_i are the residents of state i, $A \cap C_i$ are those residents who are from 18 to 25 years old and have been been infected by the H1N1 influenza A virus. Thus, distinct $A \cap C_i$ are mutually exclusive - individuals cannot reside in 2 different states. Their union is A, all individuals in the United States between the ages of 18 and 25 years old who have been been infected by the H1N1 virus.

Thus,

$$P(A) = \sum_{i=1}^{n} P(A \cap C_i). \qquad (6.4)$$

Finish by using the multiplication identity (6.2),

$$P(A \cap C_i) = P(A|C_i)P(C_i), \quad i = 1, 2, \ldots, n$$

and substituting into (6.4) to obtain the identity in (6.3).

The most frequent use of the law of total probability comes in the case of a partition of the sample space into two events, $\{C, C^c\}$. In this case the law of total probability becomes the identity

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c). \qquad (6.5)$$

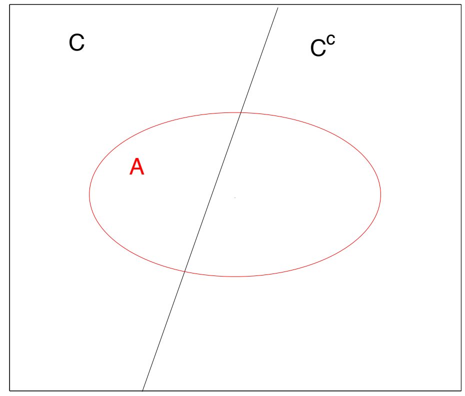

Figure 6.5: A partition into two events C and C^c.

Exercise 6.8. *The problem of points is a classical problem in probability theory. The problem concerns a series of games with two sides who have equal chances of winning each game. The winning side is one that first reaches a given number n of wins. Let $n = 4$ for a* best of seven *playoff. Determine*

$$p_{ij} = P\{\text{winning the playoff after } i \text{ wins vs } j \text{ opponent wins}\}$$

(Hint: $p_{ii} = \frac{1}{2}$ for $i = 0, 1, 2, 3$.)

6.4 Bayes formula

Let A be the event that an individual tests positive for some disease and C be the event that the person actually has the disease. We can perform clinical trials to estimate the probability that a randomly chosen individual tests positive given that they have the disease,

$$P\{\text{tests positive}|\text{has the disease}\} = P(A|C),$$

by taking individuals with the disease and applying the test. However, we would like to use the test as a method of *diagnosis* of the disease. Thus, we would like to be able to give the test and assert the chance that the person has the disease. That is, we want to know the probability with the reverse conditioning

$$P\{\text{has the disease}|\text{tests positive}\} = P(C|A).$$

Example 6.9. *The Public Health Department gives us the following information.*

- *A test for the disease yields a positive result 90% of the time when the disease is present.*
- *A test for the disease yields a positive result 1% of the time when the disease is not present.*
- *One person in 1,000 has the disease.*

Let's first think about this intuitively and then look to a more formal way using Bayes formula to find the probability of

$$P(C|A).$$

- In a city with a population of 1 million people, on average,

  1,000 have the disease and 999,000 do not

- Of the 1,000 that have the disease, on average,

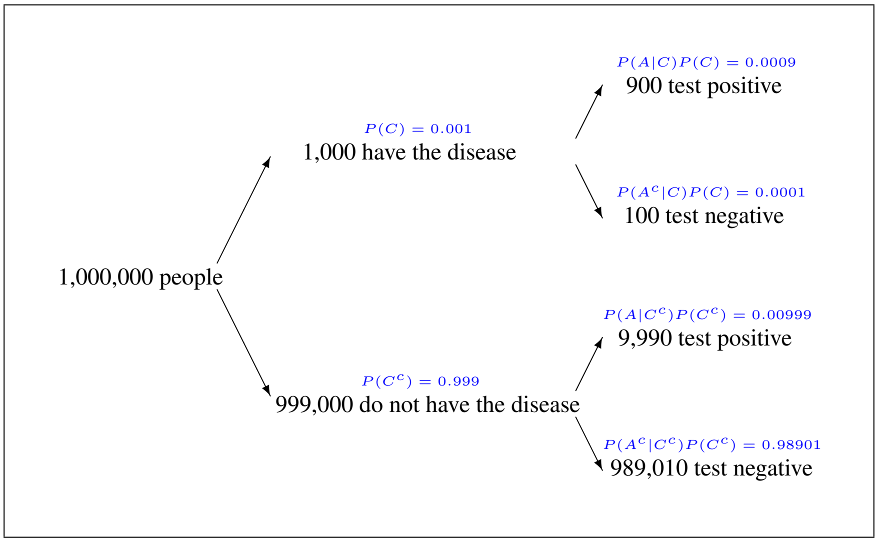

Figure 6.6: Tree diagram. We can use a tree diagram to indicate the number of individuals, on average, in each group (in black) or the probablity (in blue). Notice that in each column the number of individuals adds to give 1,000,000 and the probabilities add to give 1. In addition, each pair of arrows divides an events into two mutually exclusive subevents. Thus, both the numbers and the probabilities at the tip of the arrows add to give the respective values at the head of the arrow.

900 test positive and 100 test negative

- Of the 999,000 that do not have the disease, on average,

$$999{,}000 \times 0.01 = 9990 \text{ test positive and } 989{,}010 \text{ test negative.}$$

Consequently, among those that test positive, the *odds* of having the disease is

#(have the disease):#(does not have the disease)

900:9990

and converting odds to probability we see that

$$P\{\text{have the disease}|\text{test is positive}\} = \frac{900}{900 + 9990} = 0.0826.$$

We now derive Bayes formula. First notice that we can flip the order of conditioning by using the multiplication formula (6.2) **twice**

$$P(A \cap C) = \begin{cases} P(A|C)P(C) \\ \\ P(C|A)P(A) \end{cases}$$

Now we can create a formula for $P(C|A)$ as desired in terms of $P(A|C)$.

$$P(C|A)P(A) = P(A|C)P(C) \quad \text{or} \quad P(C|A) = \frac{P(A|C)P(C)}{P(A)}.$$

Thus, given A, the probability of C changes by the **Bayes factor**

$$\frac{P(A|C)}{P(A)}.$$

researcher			**public health worker**	**clinician**			
	has disease C	does not have disease C^c	$\longrightarrow$		has disease C	does not have disease C^c	sum
tests positive A	$P(A\vert C)$ 0.90	$P(A\vert C^c)$ 0.01	$P(C) = 0.001$ $P(C^c) = 0.999$	tests positive A	$P(C\vert A)$ 0.0826	$P(C^c\vert A)$ 0.9174	1
tests negative A^c	$P(A^c\vert C)$ 0.10	$P(A^c\vert C^c)$ 0.99	$\nwarrow$	tests negative A^c	$P(C\vert A^c)$ 0.0001	$P(C^c\vert A^c)$ 0.9999	1
sum	1	1					

Table I: Using Bayes formula to evaluate a test for a disease. Successful analysis of the results of a clinical test require researchers to provide results on the quality of the test and public health workers to provide information on the prevalence of a disease. The conditional probabilities, provided by the researchers, and the probability of a person having the disease, provided by the public health service (shown by the east arrow), are necessary for the clinician, using Bayes formula (6.6), to give the probability of the conditional probability of having the disease given the test result. Notice, in particular, that the order of the conditioning needed by the clinician is the reverse of that provided by the researcher. If the clinicians provide reliable data to the public health service, then this information can be used to update the probabilities for the prevalence of the disease (indicated by the northeast arrow). The numbers in gray can be computed from the numbers in black by using the complement rule. In particular, the column sums for the researchers and the row sums for the clinicians much be .

Example 6.10. *Both autism A and epilepsy C exists at approximately 1% in human populations. In this case*

$$P(A|C) = P(C|A)$$

Clinical evidence shows that this common value is about 30%. The Bayes factor is

$$\frac{P(A|C)}{P(A)} = \frac{0.3}{0.01} = 30.$$

Thus, the knowledge of one disease increases the chance of the other by a factor of 30.

From this formula we see that in order to determine $P(C|A)$ from $P(A|C)$, we also need to know $P(C)$, the fraction of the population with the disease and $P(A)$. We can find $P(A)$ using the law of total probability in (6.5) and write Bayes formula as

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)}. \tag{6.6}$$

This shows us that we can determine $P(A)$ if, in addition, we collect information from our clinical trials on $P(A|C^c)$, the fraction that test positive who do not have the disease.

Let's now compute $P(C|A)$ using Bayes formula directly and use this opportunity to introduce some terminology. We have that $P(A|C) = 0.90$. If one tests negative for the disease (the outcome is in A^c) given that one has the disease, (the outcome is in C), then we call this a **false negative**. In this case, the false negative probability is $P(A^c|C) = 0.10$

If one tests positive for the disease (the outcome is in A) given that one does not have the disease, (the outcome is in C^c), then we call this a **false positive**. In this case, the false positive probability is $P(A|C^c) = 0.01$.

The probability of having the disease is $P(C) = 0.001$ and so the probability of being disease free is $P(C^c) = 0.999$. Now, we apply the law of total probability (6.5) as the first step in Bayes formula (6.6),

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) = 0.90 \cdot 0.001 + 0.01 \cdot 0.999 = 0.0009 + 0.009999 = 0.01089.$$

Thus, the probability of having the disease given that the test was positive is

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)} = \frac{0.0009}{0.01089} = 0.0826.$$

Notice that the numerator is one of the terms that was summed to compute the denominator.

The answer in the previous example may be surprising. Only 8% of those who test positive actually have the disease. This example underscores the fact that good predictions based on intuition are hard to make in this case. To determine the probability, we must weigh the odds of two terms, each of them itself a product.

- $P(A|C)P(C)$, a big number (the true positive probability) times a small number (the probability of having the disease) **versus**

- $P(A|C^c)P(C^c)$, a small number (the false positive probability) times a large number (the probability of being disease free).

We do not need to restrict Bayes formula to the case of C, *has the disease*, and C^c, *does not have the disease*, as seen in (6.5), but rather to any partition of the sample space. Indeed, Bayes formula can be generalized to the case of a partition $\{C_1, C_2, \ldots, C_n\}$ of Ω chosen so that $P(C_i) > 0$ for all i. Then, for any event $A \subset \Omega$ and any j

$$P(C_j|A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^n P(A|C_i)P(C_i)}. \tag{6.7}$$

To understand why this is true, use the law of total probability to see that the denominator is equal to $P(A)$. By the multiplication identity for conditional probability, the numerator is equal to $P(C_j \cap A)$. Now, make these two substitutions into (6.7) and use one more time the definition of conditional probability.

Example 6.11. *We begin with a simple and seemingly silly example involving fair and two sided coins. However, we shall soon see that this leads us to a question in the vertical transmission of a genetic disease.*

A box has a two-headed coin and a fair coin. It is flipped n times, yielding heads each time. What is the probability that the two-headed coin is chosen?

To solve this, note that

$$P\{\text{two-headed coin}\} = \frac{1}{2}, \qquad P\{\text{fair coin}\} = \frac{1}{2}.$$

and

$$P\{n \text{ heads}|\text{two-headed coin}\} = 1, \qquad P\{n \text{ heads}|\text{fair coin}\} = 2^{-n}.$$

By the law of total probability,

$$\begin{aligned} P\{n \text{ heads}\} &= P\{n \text{ heads}|\text{two-headed coin}\}P\{\text{two-headed coin}\} + P\{n \text{ heads}|\text{fair coin}\}P\{\text{fair coin}\} \\ &= 1 \cdot \frac{1}{2} + 2^{-n} \cdot \frac{1}{2} = \frac{2^n + 1}{2^{n+1}}. \end{aligned}$$

Next, we use Bayes formula.

$$P\{\text{two-headed coin}|n \text{ heads}\} = \frac{P\{n \text{ heads}|\text{two-headed coin}\}P\{\text{two-headed coin}\}}{P\{n \text{ heads}\}} = \frac{1 \cdot (1/2)}{(2^n+1)/2^{n+1}} = \frac{2^n}{2^n+1} < 1.$$

Notice that as n increases, the probability of a two headed coin approaches 1 - with a longer and longer sequence of heads we become increasingly suspicious (but, because the probability remains less than one, are never completely certain) that we have chosen the two headed coin.

This is the related genetics question: Based on the pedigree of her past, a female knows that she has in her history a allele on her X chromosome that indicates a genetic condition. The allele for the condition is recessive. Because she does not have the condition, she knows that she cannot be homozygous for the recessive allele. Consequently, she wants to know her chance of being a carrier (heteorzygous for a recessive allele) or not a carrier (homozygous for the common genetic type) of the condition. The female is a mother with n male offspring, none of which show the recessive allele on their single X chromosome and so do not have the condition. What is the probability that the female is not a carrier?

Let's look at the computation above again, based on her pedigree, the female estimates that

$$P\{\text{mother is not a carrier}\} = p, \qquad P\{\text{mother is a carrier}\} = 1-p.$$

Then, from the law of total probability

$$\begin{aligned} &P\{n \text{ male offspring condition free}\} \\ = &P\{n \text{ male offspring condition free}|\text{mother is not a carrier}\}P\{\text{mother is not a carrier}\} \\ &+P\{n \text{ male offspring condition free}|\text{mother is a carrier}\}P\{\text{mother is a carrier}\} \\ = &1\cdot p + 2^{-n}\cdot(1-p). \end{aligned}$$

and Bayes formula

$$\begin{aligned} &P\{\text{mother is not a carrier}|n \text{ male offspring condition free}\} \\ = &\frac{P\{n \text{ male offspring condition free}|\text{mother is not a carrier}\}P\{\text{mother is not a carrier}\}}{P\{n \text{ male offspring condition free}\}} \\ = &\frac{1\cdot p}{1\cdot p + 2^{-n}\cdot(1-p)} = \frac{p}{p+2^{-n}(1-p)} = \frac{2^n p}{2^n p + (1-p)}. \end{aligned}$$

Again, with more sons who do not have the condition, we become increasingly more certain that the mother is not a carrier.

One way to introduce **Bayesian statistics** *is to consider the situation in which we do not know the value of p and replace it with a probability distribution. Even though we will concentrate on classical approaches to statistics, we will take the time in later sections to explore the Bayesian approach*

6.5 Independence

An event A is **independent of** B if its Bayes factor is 1, i.e.,

$$1 = \frac{P(A|B)}{P(A)}, \quad P(A) = P(A|B).$$

In words, the occurrence of the event B does not alter the probability of the event A. Multiply this equation by $P(B)$ and use the multiplication rule to obtain

$$P(A)P(B) = P(A|B)P(B) = P(A\cap B).$$

The formula

$$P(A)P(B) = P(A\cap B) \tag{6.8}$$

is the usual definition of independence and is symmetric in the events A and B. If A is independent of B, then B is independent of A. Consequently, when equation (6.8) is satisfied, we say that A and B are **independent**.

Example 6.12. *Roll two dice.*

$$\begin{aligned} \frac{1}{36} &= P\{a \text{ on the first die}, b \text{ on the second die}\} \\ &= \frac{1}{6}\times\frac{1}{6} = P\{a \text{ on the first die}\}P\{b \text{ on the second die}\} \end{aligned}$$

and, thus, the outcomes on two rolls of the dice are independent.

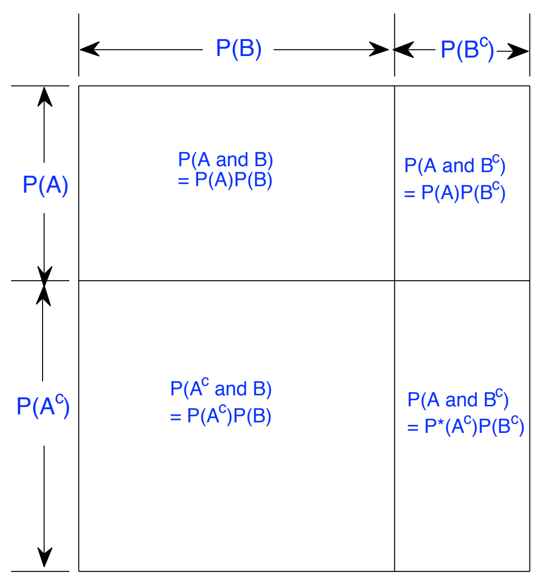

Figure 6.7: The Venn diagram for independent events is represented by the horizontal strip A and the vertical strip B is shown above. The identity $P(A\cap B) = P(A)P(B)$ is now represented as the area of the rectangle. Other aspects of Exercise 6.12 are indicated in this Figure.

Exercise 6.13. *If A and B are independent, then show that A^c and B, A and B^c, A^c and B^c are also independent.*

We can also use this to extend the definition to n independent events:

Definition 6.14. *The events $A_1, \cdots, A_n$ are called* **independent** *if for any choice $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ taken from this collection of n events, then*

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}). \tag{6.9}$$

A similar product formula holds if some of the events are replaced by their complement.

Exercise 6.15. *Flip 10 biased coins. Their outcomes are independent with the i-th coin turning up heads with probability p_i. Find*

$$P\{\text{first coin heads, third coin tails, seventh \& ninth coin heads}\}.$$

Example 6.16. *Mendel studied inheritance by conducting experiments using a garden peas. Mendel's First Law, the* **law of segregation** *states that every diploid individual possesses a pair of alleles for any particular trait and that each parent passes one randomly selected allele to its offspring.*

In Mendel's experiment, each of the 7 traits under study express themselves independently. This is an example of Mendel's Second Law, also known as the **law of independent assortment**. *If the dominant allele was present in the population with probability p, then the recessive allele is expressed in an individual when it receive this allele from both of its parents. If we assume that the presence of the allele is independent for the two parents, then*

$$\begin{aligned} P\{\text{recessive allele expressed}\} &= P\{\text{recessive allele paternally inherited}\} \times P\{\text{recessive allele maternally inherited}\} \\ &= (1-p) \times (1-p) = (1-p)^2. \end{aligned}$$

In Mendel's experimental design, p was set to be 1/2. Consequently,

$$P\{\text{recessive allele expressed}\} = (1-1/2)^2 = 1/4.$$

Using the complement rule,

$$P\{\text{dominant allele expressed}\} = 1 - (1-p)^2 = 1 - (1 - 2p + p^2) = 2p - p^2.$$

This number can also be computed by added the three alternatives shown in the Punnett square in Table 6.1.

$$p^2 + 2p(1-p) = p^2 + 2p - 2p^2 = 2p - p^2.$$

Next, we look at two traits - 1 and 2 - with the dominant alleles present in the population with probabilities p_1 and p_2. If these traits are expressed independently, then, we have, for example, that

$$\begin{aligned} &P\{\text{dominant allele expressed in trait 1, recessive trait expressed in trait 2}\} \\ &= P\{\text{dominant allele expressed in trait 1}\} \times P\{\text{recessive trait expressed in trait 2}\} \\ &= (1 - (1-p_1)^2)(1-p_2)^2. \end{aligned}$$

Exercise 6.17. *Show that if two traits are genetically linked, then the appearance of one increases the probability of the other. Thus,*

$$P\{\text{individual has allele for trait 1}|\text{individual has allele for trait 2}\} > P\{\text{individual has allele for trait 1}\}.$$

implies

$$P\{\text{individual has allele for trait 2}|\text{individual has allele for trait 1}\} > P\{\text{individual has allele for trait 2}\}.$$

More generally, for events A and B,

$$P(A|B) > P(A) \quad \textit{implies} \quad P(B|A) > P(B) \tag{6.10}$$

then we way that A and B are **positively associated**.

Exercise 6.18. *A* **genetic marker** *B for a disease A is one in which $P(A|B) \approx 1$. In this case, approximate $P(B|A)$.*

Definition 6.19. Linkage disequilibrium *is the non-independent association of alleles at two loci on single chromosome. To define linkage disequilibrium, let*

- *A be the event that a given allele is present at the first locus, and*
- *B be the event that a given allele is present at a second locus.*

Then the linkage disequilibrium,

$$D_{A,B} = P(A)P(B) - P(A \cap B).$$

Thus if $D_{A,B} = 0$, the the two events are independent.

Exercise 6.20. *Show that $D_{A,B^c} = -D_{A,B}$*

6.6 Answers to Selected Exercises

	S	s
S	SS p^2	Ss $p(1-p)$
s	sS $(1-p)p$	ss $(1-p)^2$

Table II: Punnett square for a monohybrid cross using a dominant trait S (say spherical seeds) that occurs in the population with probability p and a recessive trait s (wrinkled seeds) that occurs with probability $1-p$. Maternal genotypes are listed on top, paternal genotypes on the left. See Example 6.14. The probabilities of a given genotype are given in the lower right hand corner of the box.

6.1. Let's check the three axioms;

1. For any event A,
$$Q(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$
2. For the sample space Ω,
$$Q(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$
3. For mutually exclusive events, $\{A_j; j \geq 1\}$, we have that $\{A_j \cap B; j \geq 1\}$ are also mutually exclusive and

$$Q\left(\bigcup_{j=1}^\infty A_j\right) = P\left(\bigcup_{j=1}^\infty A_j \Big| B\right) = \frac{P\left(\left(\bigcup_{j=1}^\infty A_j\right) \cap B\right)}{P(B)} = \frac{P(\bigcup_{j=1}^\infty (A_j \cap B))}{P(B)}$$
$$= \frac{\sum_{j=1}^\infty P(A_j \cap B)}{P(B)} = \sum_{j=1}^\infty \frac{P(A_j \cap B)}{P(B)} = \sum_{j=1}^\infty P(A_j|B) = \sum_{j=1}^\infty Q(A_j)$$

6.2. $P\{\text{sum is 8}|\text{first die shows 3}\} = 1/6$, and $P\{\text{sum is 8}|\text{first die shows 1}\} = 0$.

6.3 Here is a table of outcomes. The symbol $\times$ indicates an outcome in the event $\{\text{sum is at least 5}\}$. The rectangle indicates the event $\{\text{first die is 2}\}$. Because there are 10 $\times$'s,

$$P\{\text{sum is at least 5}\} = 10/16 = 5/8.$$

	1	2	3	4
1				$\times$
2			$\times$	$\times$
3		$\times$	$\times$	$\times$
4	$\times$	$\times$	$\times$	$\times$

The rectangle contains 4 outcomes, so

$$P\{\text{first die is 2}\} = 4/16 = 1/4.$$

Inside the event $\{\text{first die is 2}\}$, 2 of the outcomes are also in the event $\{\text{sum is at least 5}\}$. Thus,

$$P\{\text{sum is at least 5}\}|\text{first die is 2}\} = 2/4 = 1/2$$

Using the definition of conditional probability, we also have

$$P\{\text{sum is at least 5}\}|\text{first die is 2}\} = \frac{P\{\text{sum is at least 5 and first die is 2}\}}{P\{\text{first die is 2}\}} = \frac{2/16}{4/16} = \frac{2}{4} = \frac{1}{2}.$$

6.5. We modify both sides of the equation.

$$\binom{4}{2}\frac{(g)_2(b)_2}{(b+g)_4} = \frac{4!}{2!2!}\frac{(g)_2(b)_2}{(b+g)_4}$$

$$\frac{\binom{b}{2}\binom{g}{2}}{\binom{b+g}{4}} = \frac{(b)_2/2!\cdot(g)_2/2!}{(b+g)_4/4!} = \frac{4!}{2!2!}\frac{(g)_2(b)_2}{(b+g)_4}.$$

The sample space Ω is set of collections of 4 balls out of $b+g$. This has $\binom{b+g}{4}$ outcomes. The number of choices of 2 blue out of b is $\binom{b}{2}$. The number of choices of 2 green out of g is $\binom{g}{2}$. Thus, by the fundamental principle of counting, the total number of ways to obtain the event 2 blue and 2 green is $\binom{b}{2}\binom{g}{2}$. For equally likely outcomes, the probability is the ratio of $\binom{b}{2}\binom{g}{2}$, the number of outcomes in the event, and $\binom{b+g}{4}$, the number of outcomes in the sample space.

6.8. Let A_{ij} be the event of winning the series that has i *wins* versus j *wins for the opponent*. Then $p_{ij} = P(A_{ij})$. We know that

$$p_{0,4} = p_{1,4} = p_{2,4} = p_{3,4} = 0$$

because the series is lost when the opponent has won 4 games. Also,

$$p_{4,0} = p_{4,1} = p_{4,2} = p_{4,3} = 1$$

because the series is won with 4 wins in games. For a tied series, the probability of winning the series is 1/2 for both sides.

$$p_{0,0} = p_{1,1} = p_{2,2} = p_{3,3} = \frac{1}{2}.$$

These values are filled in blue in the table below. We can determine the remaining values of p_{ij} iteratively by looking forward one game and using the law of total probability to condition of the outcome of the ($i+j+1$-st) game. Note that $P\{\text{win game } i+j+1\} = P\{\text{lose game } i+j+1\} = \frac{1}{2}$.

$$\begin{aligned} p_{ij} &= P(A_{ij}|\text{win game } i+j+1\}P\{\text{win game } i+j+1\} + P(A_{ij}|\text{lose game } i+j-1\}P\{\text{lose game } i+j+1\} \\ &= \frac{1}{2}(p_{i+1,j} + p_{i,j+1}) \end{aligned}$$

This can be used to fill in the table above the diagonal. For example,

$$p_{23} = \frac{1}{2}(p_{33} + p_{42}) = \frac{1}{2}\left(\frac{1}{2} + 1\right) = \frac{3}{4}.$$

For below the diagonal, note that

$$p_{ij} = 1 - p_{ji}.$$

For example,

$$p_{23} = 1 - p_{32} = 1 - \frac{3}{4} = \frac{1}{4}.$$

Filling in the table, we have:

		i				
		0	1	2	3	4
j	0	1/2	21/32	13/16	15/16	1
	1	11/32	1/2	11/16	7/8	1
	2	3/16	5/16	1/2	3/4	1
	3	1/16	1/8	1/4	1/2	1
	4	0	0	0	0	-

6.13. We take the questions one at a time. Because A and B are independent $P(A \cap B) = P(A)P(B)$.

(a) B is the disjoint union of $A \cap B$ and $A^c \cap B$. Thus,

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

Subtract $P(A \cap B)$ to obtain

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(A^c)P(B)$$

and A^c and B are independent.

(b) Just switch the roles of A and B in part (a) to see that A and B^c are independent.

(c) Use the complement rule and inclusion-exclusion

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B) = 1 - P(A) - P(B) - P(A \cap B) \\ &= 1 - P(A) - P(B) - P(A)P(B) = (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

and A^c and B^c are independent.

6.15. Let A_i be the event $\{i\text{-th coin turns up heads}\}$. Then the event can be written $A_1 \cap A_3^c \cap A_7 \cap A_9$. Thus,

$$\begin{aligned} P(A_1 \cap A_3^c \cap A_7 \cap A_9) &= P(A_1)P(A_3^c)P(A_7)P(A_9) \\ &= p_1(1 - p_3)p_7p_9. \end{aligned}$$

6.17. Multiply both of the expressions in (6.10) by the appropriate probability to see that they are equivalent to

$$P(A \cap B) > P(A)P(B).$$

6.18. By using Bayes formula we have

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \approx \frac{P(B)}{P(A)}.$$

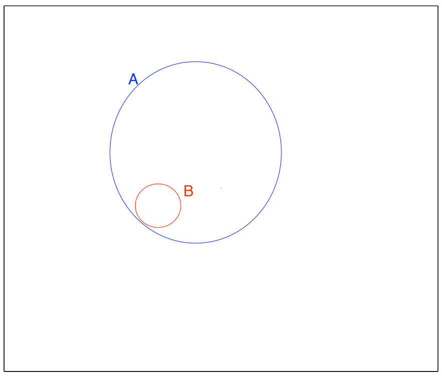

Figure 6.8: If $P(A|B) \approx 1$, then most of B is inside A and the probability of $P(B|A) \approx P(B)/P(A)$ as shown in the figure.

6.20 Because A is the disjoint union of $A \cap B$ and $A \cap B^c$, we have $P(A) = P(A \cap B) + P(A \cap B^c)$ or $P(A \cap B^c) = P(A) - P(A \cap B)$. Thus,

$$D_{A,B^c} = P(A)P(B^c) - P(A \cap B^c) = P(A)(1 - P(B)) - (P(A) - P(A \cap B)) = -P(A)P(B) + P(A \cap B) = -D_{A,B}.$$