

# Topic 6: Conditional Probability and Independence\*

September 15-20, 2011

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? This question leads us to the concept of **conditional probability**.

## 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 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

$$\frac{\#(\text{outcome that result in a win and also have a heads on the first coin toss})}{\#(\text{outcomes with heads on the first coin toss})}$$

$$= \frac{\#\{\text{HHH, HHT, HTH}\}}{\#\{\text{HHH, HHT, HTH, HTT}\}} = \frac{3}{4}.$$

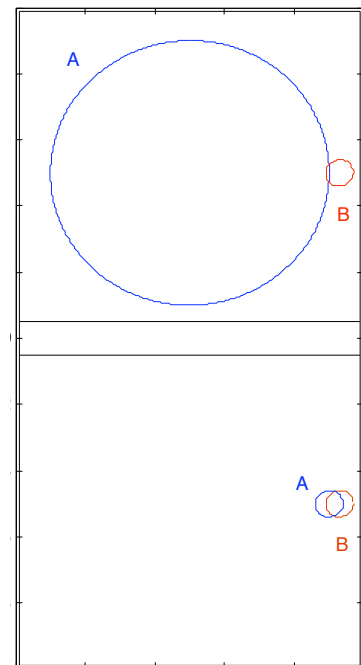
We can take this idea to create a formula in the case of equally likelihood outcomes for the statement the **conditional probability of A given B**.

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)}$$

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  $\#(\Omega)$ .

$$P(A|B) = \frac{\#(A \cap B)/\#(\Omega)}{\#(B)/\#(\Omega)} = \frac{P(A \cap B)}{P(B)} \quad (1)$$

We thus take this version (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 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.

\*© 2011 Joseph C. Watkins

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

## 2 The Multiplication Principle

The defining formula (1) for conditional probability can be rewritten to obtain **the multiplication principle**,

$$P(A \cap B) = P(A|B)P(B). \quad (2)$$

Now, we can complete an earlier problem:

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

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 3.** 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

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

**Exercise 4.** Show that

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

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

### 3 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 5.** A **partition** of the sample space  $\Omega$  is a finite collection of pairwise mutually exclusive events

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

whose union is  $\Omega$ .

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  $i \neq j$ )

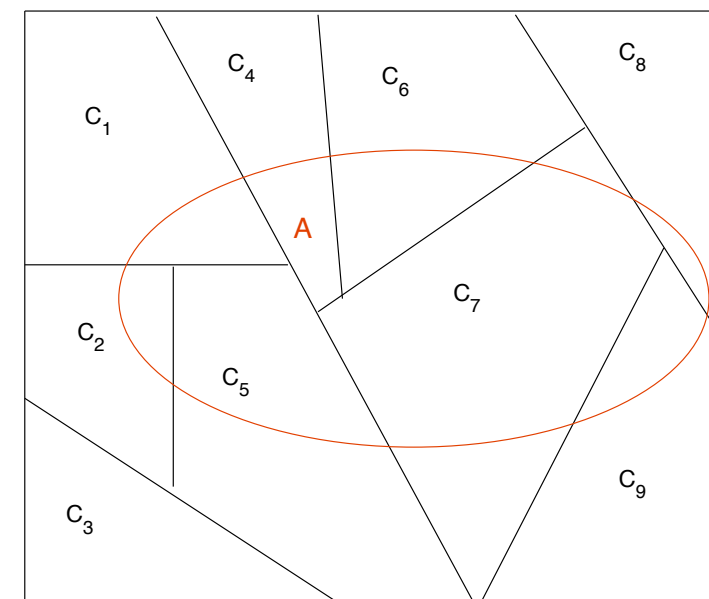
If we know the fraction of the population from 18 to 25 that has been infected by the H1N1 influenza A virus in each of the 50 states, then we cannot just average these 50 values to obtain the fraction of this population infected in the whole country. This method fails because it give equal weight to California and Wyoming.

**Theorem 6** (Law of total probability). Let  $P$  be a probability on  $\Omega$  and let  $\{C_1, C_2, \dots, C_n\}$  be a partition of  $\Omega$  chosen so that  $P(C_i) > 0$  for all  $i$ . Then, for any event  $A \subset \Omega$ ,

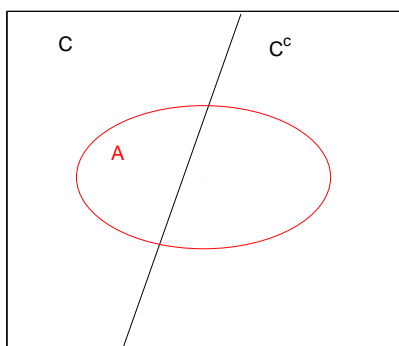
$$P(A) = \sum_{i=1}^n P(A|C_i)P(C_i). \quad (3)$$

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



**Figure 4:** A partition  $\{C_1, \dots, C_9\}$  of the sample space  $\Omega$ . The event  $A$  can be written as the union  $(A \cap C_1) \cup \dots \cup (A \cap C_9)$  of mutually exclusive events.



**Figure 5:** A partition into two events  $C$  and  $C^c$ .

Thus,

$$P(A) = \sum_{i=1}^n P(A \cap C_i). \quad (4)$$

Finish by using the multiplication identity (2),

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

and substituting into (4) to obtain the identity in (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

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c). \quad (5)$$

## 4 Bayes formula

Let  $A$  be the event that an individual test 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,

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

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

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

**Example 7.** *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 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,

900 test positive and 100 test negative

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

$999,000 \times 0.01 = 9990$  test positive and remaining 990,010 test negative.

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

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

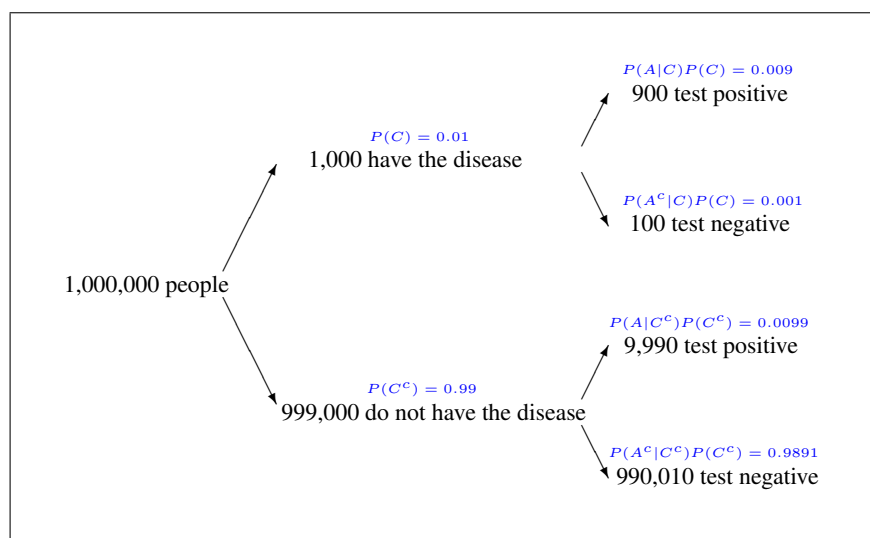
900:9990

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 (2) **twice**

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



**Figure 6: Tree diagram.** We can use a tree diagram to indicate the number of individuals, on average, in each group (in black) or the probability (in blue). Notice that in each column the number of individuals adds to give 1,000,000 and the probability adds to give 1. In addition, each pair of arrows divides an event 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.

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**

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

**Example 8.** 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 (5) and write Bayes formula as

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)}. \quad (6)$$

This shows us that we can determine  $P(A)$  if 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 (5) as the first step in Bayes formula (6),

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) = 0.90 \cdot 0.001 + 0.01 \cdot 0.999 = 0.0009 + 0.00999 = 0.01089.$$

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 common sense predictions 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 have the disease  $C$  and does not have the disease  $C^c$  as seen in (5), but rather to any partition of the sample space. Indeed, Bayes formula can be generalized to the case of a partition  $\{C_1, C_2, \dots, C_n\}$  of  $\Omega$  chosen so that  $P(C_i) > 0$  for all  $i$ . Then, for any event  $A \subset \Omega$  and any  $j$

$$P(C_j|A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^n P(A|C_i)P(C_i)}. \quad (7)$$

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 (7) and use one more time the definition of conditional probability

**Example 9.** 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

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

and

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

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 recessive 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 (heterozygous 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

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

Then, from the law of total probability

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

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

## 5 Independence

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

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

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

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

This formula

$$P(A)P(B) = P(A \cap B) \tag{8}$$

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 (8) is satisfied, we say that  $A$  and  $B$  are **independent**.

**Example 10.** 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.

**Exercise 11.** 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 12.** The events  $A_1, \dots, A_n$  are called **independent** if for any choice  $A_{i_1}, A_{i_2}, \dots, A_{i_k}$  taken from this collection of  $n$  events, then

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}).$$

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

**Exercise 13.** Flip 10 biased coins. Their outcomes are independent with the  $i$ -th coin turning up heads with probability  $p_i$ . Find

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

**Example 14.** 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 gene was present in the population with probability  $p$ , then the recessive gene is expressed in an individual when it receive this gene from both of its parents. If we assume that the presence of the gene is independent for the two parents, then

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

In Mendel's experimental design,  $p$  was set to be  $1/2$ . Consequently,  $P\{\text{recessive gene expressed}\} = (1 - 1/2)^2 = 1/4$ .

Using the complement rule,

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

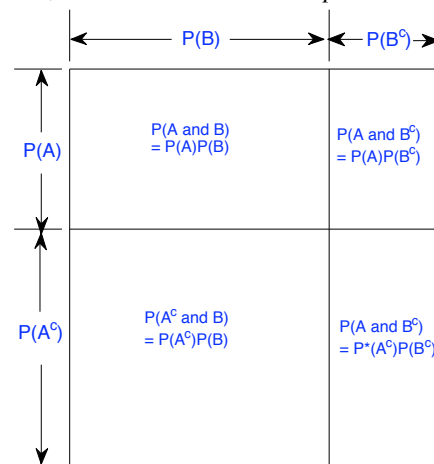
This number can also be computed by added the three alternatives shown in the Punnett square in Table 1 -  $p^2 + 2p(1 - p) = p^2 + 2p - 2p^2 = 2p - p^2$ .

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

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

**Exercise 15.** Show that if two traits are genetically linked, then the appearance of one increases the probability of the other. Thus,

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



**Figure 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 10 are indicated in this Figure.



implies

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

More generally, for events  $A$  and  $B$ ,

$$P(A|B) > P(A) \text{ implies } P(B|A) > P(B) \tag{9}$$

then we say that  $A$  and  $B$  are **positively associated**.

**Exercise 16.** A genetic marker  $B$  for a disease  $A$  is one in which  $P(A|B) \approx 1$ . In this case, approximate  $P(B|A)$ .

## 6 Answers to Selected Exercises

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  $\Omega$ ,

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

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

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

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

4. We modify both sides of the equation.

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

11. We take them 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.

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

**Table I:** 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 14. The probabilities of a given genotype are given in the lower right hand corner of the box.

(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.

13. 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,

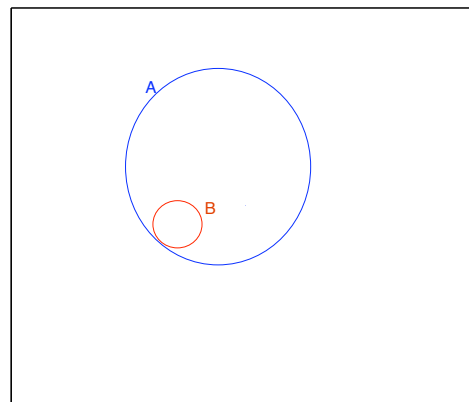
$$P(A_1 \cap A_3^c \cap A_7 \cap A_9) = P(A_1)P(A_3^c)P(A_7)P(A_9) = p_1(1 - p_3)p_7p_9.$$

15. Both of the expressions in (9) are equivalent to

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

16. By using Bayes formula we have

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



**Figure 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.