4.1 Randomness

A random phenomenon has outcomes that we cannot predict but that nonetheless have a regular distribution in very many repetitions.

The probability of an event is the proportion of times the event occurs in many repeated trials of a random phenomenon.

Trials are independent if the outcome of one trial does not influence the outcome of any other trial.

4.2 Probability Models

SAMPLE SPACE

The sample space S of a random phenomenon is the set of all possible outcomes.

PROBABILITY RULES

Rule 1. The probability $P(A)$ of any event A satisfies $0 \le P(A) \le 1$.

Rule 2. If S is the sample space in a probability model, then $P(S) = 1$.

Rule 3. Two events A and B are **disjoint** if they have no outcomes in common and so can never occur together. If A and B are disjoint,

$$P\left(A \text{or} B\right) = P\left(A\right) + P\left(B\right)$$

This is the **addition rule for disjoint events.**

Rule 4. The **complement** of any event A is the event that A does not occur, written as A^c. The **complement rule** states that

$$P\left(A^c\right) = 1 - P\left(A\right)$$

EQUALLY LIKELY OUTCOMES

If a random phenomenon has k possible outcomes, all equally likely, then each individual outcome has probability $1/k$. The probability of any event A is

$$P\left(A\right) = \frac{\text{count of outcomes in } A}{\text{count of outcomes in } S} = \frac{\text{count of outcomes in } A}{k}$$

MULTIPLICATION RULE FOR INDEPENDENT EVENTS

Rule 5. Two events A and B are **independent** if knowing that one occurs does not change the probability that the other occurs. If A and B are independent,

$$P\left(A \text{ and } B\right) = P\left(A\right)\ P\left(B\right)$$

This is the **multiplication rule for independent events**.

4.3 Random Variables

DISCRETE RANDOM VARIABLE

A **discrete random variable** X has possible values that can be given in an ordered list. The **probability distribution** of X lists the values and their probabilities:

Value of X	x_1	x_2	x_3	...
Probability	p_1	p_2	p_3	...

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. $p_1 + p_2 + \cdots = 1$.

Find the probability of any event by adding the probabilities p_i of the particular values x_i that make up the event.

CONTINUOUS RANDOM VARIABLE

A **continuous random variable** X takes all values in an interval of numbers. The **probability distribution** of X is described by a density curve. The probability of any event is the area under the density curve and above the values of X that make up the event.

Normal distributions as probability distributions

EXAMPLE 4.26

Texting while driving. Texting while driving can be dangerous, but young people want to remain connected. Suppose that 26% of teen drivers text while driving. If we take a sample of 500 teen drivers, what percent would we expect to say that they text while driving?[11]

The proportion $p = 0.26$ is a number that describes the population of teen drivers. The proportion $\hat{p}$ of the sample who say that they text while driving is used to estimate p. The proportion $\hat{p}$ is a random variable because repeating the SRS would give a different sample of 500 teen drivers and a different value of $\hat{p}$.

We will see in the next chapter that in this setting, with teen drivers answering honestly, $\hat{p}$ has approximately the $N(0.26, 0.0196)$ distribution. The mean 0.26 of this distribution is the same as the population proportion because $\hat{p}$ is an unbiased estimate of p. The standard deviation is controlled mainly by the size of the sample.

What is the probability that the survey result differs from the truth about the population by no more than 3 percentage points? We can use what we learned about Normal distribution calculations to answer this question. Because p = 0.26, the survey misses by no more than 3 percentage points if the sample proportion is between 0.23 and 0.29.

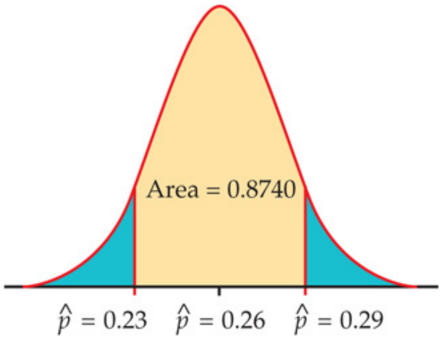

FIGURE 4.11 Probability as area under a Normal density curve, Example 4.26.

Figure 4.11 shows this probability as an area under a Normal density curve. You can find it by software or by standardizing and using Table A. From Table A,

$$\begin{aligned} P(0.23 \leq \hat{p} \leq 0.29) &= P\left(\frac{0.23-0.26}{0.0196} \leq \frac{\hat{p}-0.26}{0.0196} \leq \frac{0.29-0.26}{0.0196}\right) \\ &= P(-1.53 \leq Z \leq 1.53) \\ &= 0.9370 - 0.0630 = 0.8740 \end{aligned}$$

About 87% of the time, the sample $\hat{p}$ will be within 3 percentage points of the proportion p.

SECTION 4.3 SUMMARY

A random variable is a variable taking numerical values determined by the outcome of a random phenomenon. The probability distribution of a random variable X tells us what the possible values of X are and how probabilities are assigned to those values.

A random variable X and its distribution can be discrete or continuous.

A discrete random variable has possible values that can be given in an ordered list. The probability distribution assigns each of these values a probability between 0 and 1 such that the sum of all the probabilities is exactly 1. The probability of any event is the sum of the probabilities of all the values that make up the event.

A continuous random variable takes all values in some interval of numbers. A density curve describes the probability distribution of a continuous random variable. The probability of any event is the area under the curve and above the values that make up the event.

Uniform distributions are continuous probability distributions that are very similar to equally likely discrete distributions.

Normal distributions are one type of continuous probability distribution.

You can picture a probability distribution by drawing a probability histogram in the discrete case or by graphing the density curve in the continuous case.

4.4 Means and Variances of Random Variables

MEAN OF A DISCRETE RANDOM VARIABLE

Suppose that X is a **discrete random variable** whose distribution is

Value of X	x_1	x_2	x_3	...
Probability	p_1	p_2	p_3	...

To find the **mean** of X, multiply each possible value by its probability, then add all the products:

$$\begin{aligned} \mu_X &= x_1 p_1 + x_2 p_2 + \dots \\ &= \sum x_i p_i \end{aligned}$$

LAW OF LARGE NUMBERS

Draw independent observations at random from any population with finite mean μ. Decide how accurately you would like to estimate μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values eventually approaches the mean μ of the population as closely as you specified and then stays that close.

Rule 1. If X is a random variable and a and b are fixed numbers, then

$$\mu_{a+bX} = a + b\mu_X$$

Rule 2. If X and Y are random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

Rule 3. If X and Y are random variables, then

$$\mu_{X-Y} = \mu_X - \mu_Y$$

VARIANCE OF A DISCRETE RANDOM VARIABLE

Suppose that X is a **discrete random variable** whose distribution is

Value of X	x_1	x_2	x_3	...
Probability	p_1	p_2	p_3	...

and that μ_X is the mean of X. The **variance** of X is

$$\begin{aligned} \sigma_X^2 &= (x_1 + \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \dots \\ &= \sum (x_i - \mu_X)^2 p_i \end{aligned}$$

The **standard deviation** σ_X of X is the square root of the variance.

EXAMPLE 4.34

Find the mean and the variance. In Example 4.32 (pages 254–255), we saw that the distribution of the number X of fall courses taken by students at a small liberal arts college is

Courses in the fall	1	2	3	4	5	6
Probability	0.05	0.05	0.13	0.26	0.36	0.15

We can find the mean and variance of X by arranging the calculation in the form of a table. Both μ_X and σ_X^2 are sums of columns in this table.

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$
1	0.05	0.05	$(1 - 4.28)^2(0.05) = 0.53792$
2	0.05	0.10	$(2 - 4.28)^2(0.05) = 0.25992$
3	0.13	0.39	$(3 - 4.28)^2(0.13) = 0.21299$
4	0.26	1.04	$(4 - 4.28)^2(0.26) = 0.02038$
5	0.36	1.80	$(5 - 4.28)^2(0.36) = 0.18662$
6	0.15	0.90	$(6 - 4.28)^2(0.15) = 0.44376$
	$\mu_X = 4.28$		$\sigma_X^2 = 1.662$

We see that $\sigma_X^2 = 1.662$. The standard deviation of X is $\sigma_X = \sqrt{1.662} = 1.289$. The standard deviation is a measure of the variability of the number of fall courses taken by the students at the small liberal arts college. As in the case of distributions for data, the standard deviation of a probability distribution is easiest to understand for Normal distributions.

RULES FOR VARIANCES AND STANDARD DEVIATIONS OF LINEAR TRANSFORMATIONS, SUMS, AND DIFFERENCES

Rule 1. If X is a random variable and a and b are fixed numbers, then

$$\sigma_{a+bX}^2 = b^2\sigma_X^2$$

Rule 2. If X and Y are independent random variables, then

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$
$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

This is the **addition rule for variances of independent random variables**.

Rule 3. If X and Y have correlation ρ, then

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$$
$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$$

This is the **general addition rule for variances of random variables**.

To find the standard deviation, take the square root of the variance.

- The **standard deviation** σ_X is the square root of the variance. The standard deviation measures the variability of the distribution about the mean. It is easiest to interpret for Normal distributions.
- The **mean and variance of a continuous random variable** can be computed from the density curve, but to do so requires more advanced mathematics.
- The means and variances of random variables obey the following rules. If a and b are fixed numbers, then

$$\begin{aligned} \mu_{a+bX} &= a + b\mu_X \\ \sigma^2_{a+bX} &= b^2\sigma^2_X \end{aligned}$$

- If X and Y are any two random variables having correlation ρ, then

$$\begin{aligned} \mu_{X+Y} &= \mu_X + \mu_Y \\ \mu_{X-Y} &= \mu_X - \mu_Y \\ \sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y \\ \sigma^2_{X-Y} &= \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y \end{aligned}$$

- If X and Y are **independent**, then $\rho = 0$. In this case,

$$\begin{aligned} \sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y \\ \sigma^2_{X-Y} &= \sigma^2_X + \sigma^2_Y \end{aligned}$$

- To find the standard deviation, take the square root of the variance.

4.5 General Probability Rules

Rule 1. $0 \le P(A) \le 1$ for any event A

Rule 2. $P(S) = 1$

Rule 3. Addition rule: If A and B are **disjoint** events, then

$$P(A \text{or} B) = P(A) + P(B)$$

Rule 4. Complement rule: For any event A,

$$P(A^c) = 1 - P(A)$$

Rule 5. Multiplication rule: If A and B are **independent** events, then

$$P(A \text{and} B) = P(A)P(B)$$

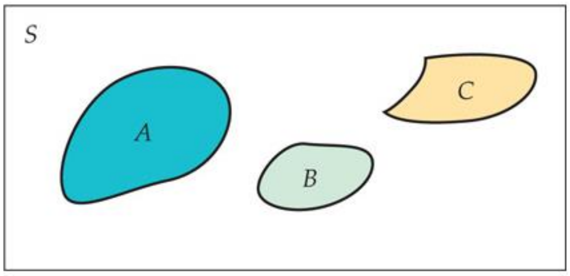

FIGURE 4.15 The addition rule for disjoint events: $P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$ when events A, B, and C are disjoint.

ADDITION RULE FOR DISJOINT EVENTS

If events A, B, and C are disjoint in the sense that no two have any outcomes in common, then

$$P(\text{one or more of } A, B, C) = P(A) + P(B) + P(C)$$

This rule extends to any number of disjoint events.

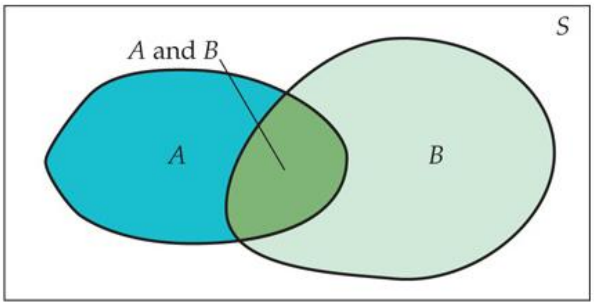

FIGURE 4.17 The union of two events that are not disjoint. The general addition rule says that $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

GENERAL ADDITION RULE FOR UNIONS OF TWO EVENTS

For any two events A and B,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

MULTIPLICATION RULE

The probability that both of two events A and B happen together can be found by

$$P(A \text{ and } B) = P(A)P(B|A)$$

Here $P(B \mid A)$ is the conditional probability that B occurs, given the information that A occurs.

Probability of a favorable draw. Doyle is still at the poker table. At the moment, he has two cards and they are both hearts. He has seen 24 cards and none of other players have any hearts. What is the chance that the next three cards he draws will be hearts? The full deck of 52 cards contains 13 hearts. Therefore, 11 of the unseen cards are hearts. There are 28 (52 − 24) unseen cards. To find Doyle's probability of drawing three hearts, we first calculate

$$\begin{aligned} P(\text{firstcardisaheart}) &= \tfrac{11}{28} \\ P(\text{secondcardisaheart}|\text{firstcardisaheart}) &= \tfrac{10}{27} \\ P(\text{thirdcardisaheart}|\text{firsttwocardsarehearts}) &= \tfrac{9}{26} \end{aligned}$$

Doyle finds both probabilities by counting cards. The probability that the first card drawn is a heart is 11/28 because 11 of the 28 unseen cards are hearts. If the first card is a heart, that leaves 10 hearts among the 27 remaining cards. So the *conditional* probability of another diamond is 10/27. The multiplication rule now says that

$$P(\text{nexttwocardsarehearts}) = \tfrac{11}{28} \times \tfrac{10}{27} = 0.146$$

We again apply the multiplication rule for the third card. The probability that the next three draws are hearts is equal to the probability that the first two draws are hearts times the probability that the third card is a heart given that the first two draws are hearts. This probability is

$$P(\text{nextthreecardsarehearts}) = \tfrac{11}{28} \times \tfrac{10}{27} \times \tfrac{9}{26} = 0.050$$

It is very unlikely that Doyle's next three cards will be hearts, even though his hearts are the only ones that he has seen.

DEFINITION OF CONDITIONAL PROBABILITY

When $P(A) > 0$, the **conditional probability** of B given A is

$$P(B|A) = \frac{P(A\text{and}B)}{P(A)}$$

INTERSECTION

The **intersection** of any collection of events is the event that *all* the events occur.

To extend the multiplication rule to the probability that all of several events occur, the key is to condition each event on the occurrence of *all* the preceding events. For example, the intersection of three events A, B, and C has probability

$$P(A \text{ and } B \text{ and } C) = P(A)P(B|A)P(C|A\text{and}B)$$

EXAMPLE 4.47

Online chat rooms. Online chat rooms are dominated by the young. Teens are the biggest users. If we look only at adult Internet users (aged 18 and over), 47% of the 18 to 29 age group chat, as do 21% of the 30 to 49 age group and just 7% of those 50 and over. To learn what percent of all Internet users participate in chat, we also need the age breakdown of users. Here it is: 29% of adult Internet users are 18 to 29 years old (event A_1), another 47% are 30 to 49 (event A_2), and the remaining 24% are 50 and over (event A_3).

tree diagram

What is the probability that a randomly chosen adult user of the Internet participates in chat rooms (event C)? To find out, use the **tree diagram** in Figure 4.19 to organize your thinking. Each segment in the tree is one stage of the problem. Each complete branch shows a path through the two stages. The probability written on each segment is the conditional probability of an Internet user following that segment, given that he or she has reached the node from which it branches.

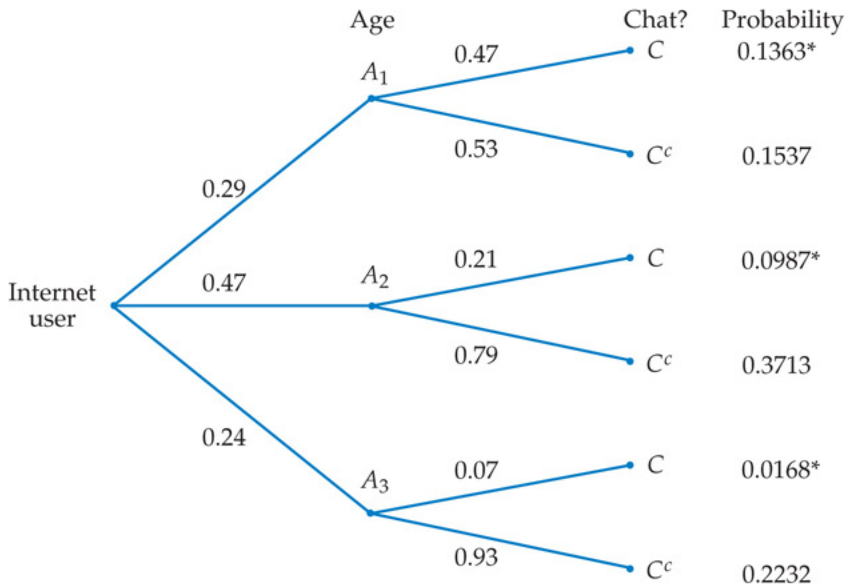

FIGURE 4.19 Tree diagram, Example 4.47. The probability $P(C)$ is the sum of the probabilities of the three branches marked with asterisks (*).

Bayes's rule

There is another kind of probability question that we might ask in the context of thinking about online chat. What percent of adult chat room participants are aged 18 to 29?

EXAMPLE 4.48

Conditional versus unconditional probabilities. In the notation of Example 4.47, this is the conditional probability $P(A_1 \mid C)$. Start from the definition of conditional probability and then apply the results of Example 4.47:

$$P(A_1|C) = \frac{P(A_1 \text{ and} C)}{P(C)} = \frac{0.1363}{0.2518} = 0.5413$$

More than half of adult chat room participants are between 18 and 29 years old. Compare this conditional probability with the original information (unconditional) that 29% of adult Internet users are between 18 and 29 years old. Knowing that a person chats increases the probability that he or she is young.

We know the probabilities $P(A_1)$, $P(A_2)$, and $P(A_3)$ that give the age distribution of adult Internet users. We also know the conditional probabilities $P(C \mid A_1)$, $P(C \mid A_2)$, and $P(C \mid A_3)$ that a person from each age group chats. Example 4.47 shows how to use this information to calculate $P(C)$. The method can be summarized in a single expression that adds the probabilities of the three paths to C in the tree diagram:

$$P(C) = P(A_1)P(C|A_1) + P(A_2)P(C|A_2) + P(A_3)P(C|A_3)$$

In Example 4.48, we calculated the "reverse" conditional probability $P(A_1 \mid C)$. The denominator 0.2518 in that example came from the previous expression. Put in this general notation, we have another probability law.

BAYES'S RULE

Suppose that $A_1, A_2, \ldots, A_k$ are disjoint events whose probabilities are not 0 and add to exactly 1. That is, any outcome is in exactly one of these events. Then if C is any other event whose probability is not 0 or 1,

$$P(A_i|C) = \frac{P(C|A_i)P(A_i)}{P(C|A_1)P(A_1)+P(C|A_2)P(A_2)+\cdots+P(A_k)P(C|A_k)}$$

The numerator in Bayes's rule is always one of the terms in the sum that makes up the denominator. The rule is named after Thomas Bayes, who wrestled with arguing from outcomes like C back to the A_i in a book published in 1763. It is far better to think your way through problems like Examples 4.47 and 4.48 than to memorize these formal expressions.

Independence again

The conditional probability $P(B \mid A)$ is generally not equal to the unconditional probability $P(B)$. That is because the occurrence of event A generally gives us some additional information about whether or not event B occurs. If knowing that A occurs gives no additional information about B, then A and B are independent events. The formal definition of independence is expressed in terms of conditional probability.

INDEPENDENT EVENTS

Two events A and B that both have positive probability are **independent** if

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

This definition makes precise the informal description of independence given in Section 4.2 (page 229). We now see that the multiplication rule for independent events, $P(A \text{ and } B) = P(A)P(B)$, is a special case of the general multiplication rule, $P(A \text{ and } B) = P(A)P(B \mid A)$, just as the addition rule for disjoint events is a special case of the general addition rule.