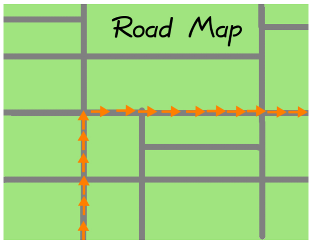
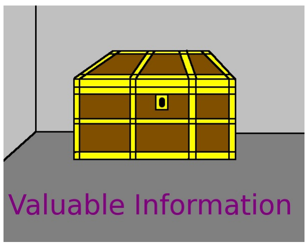
Module 7.5: A Formal Introduction to Probability



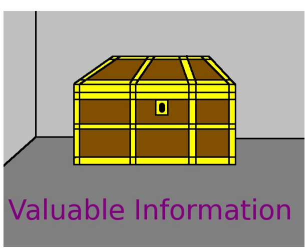
In this module I will define the building blocks of the theory of probability. It is a unique subject, because it is simultaneously very simple and very useful, yet the vast majority of the population has no knowledge of it. Modeling uncertain situations can enable one to assign fractions to outcomes to measure their relative likelihood. That is what we endeavor to learn here, and what we will continue to do for the remainder of this chapter.



What is a probability? A probability p is a number calculated for an event, to tell us about that event's likelihood. A larger probability represents a likelier event, and a smaller one a less likely event. It must always be the case that $0 \leq p \leq 1$.

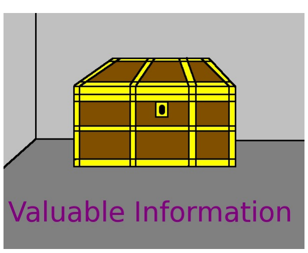
By the way, if you take several advanced statistics courses, you might eventually learn that there is a difference between "probability zero" and "absolutely impossible," as well as a difference between "probability one" and "absolutely certain."

However, that is extremely theoretical and does not reflect any remotely practical situations. We will not mention this distinction again.



For now, you should think of 0 as an event that is absolutely impossible, and 1 as an event that is absolutely certain.

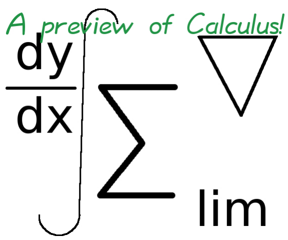
Yet, there is another important piece to the puzzle, which gives meaning to these fractions. That piece of the puzzle is the "Law of Large Numbers."



In an informal sense, the *Law of Large Numbers* says that if you repeat an experiment an enormous number of times, then the fraction of times that you get the event E (very nearly) equals the probability of E.

Another informal way of saying that is "if the event E has a probability of p, and you make n attempts, then for extremely large values of n, the number of times that event E actually occurs is very close to np."

This will make more sense after I show you some examples. We will have four examples momentarily.



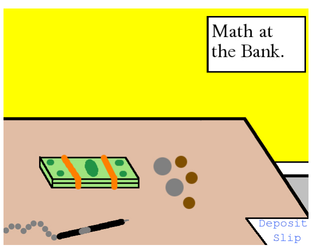
I'd really love to make the concept in the previous box more precise. However, it requires calculus to make the previous definition mathematically precise. If you've not had calculus, *do not read* the remainder of this box—instead skip to the next box.

Remember, you should only be reading this box if you've had calculus. If you haven't had calculus, then please skip to the next box. Suppose an event E has probability p, and you engage in n independent trials of the event E. Let x be the number of times that E actually occurs. Then, in the limit as n goes to infinity, $x/n = p$. We can even write

$$\lim_{n\to\infty} \frac{x}{n} = p$$

Last but not least, note that some textbooks write "$x = np$ (in the limit as n goes to infinity)," which clearly means the same thing.

Because the majority of my readers have not had calculus, I'd like to give a more "everyday" description of The Law of Large Numbers, through four examples, in the next two boxes.



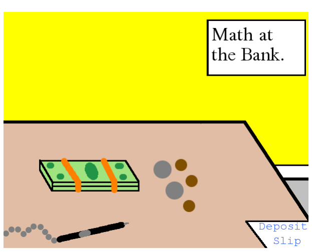
Homeowner's Insurance is an excellent way of understanding probability. It turns out there is a reliable rule of thumb, that a typical American home will be destroyed with probability 1 in 500 in any particular year, including fires, earthquakes, floods, mudslides, tornados, hurricanes, and so forth.

A particular insurance company might easily insure 100,000 customers, or many more. Let's work with 100,000 customers. Therefore, they can trust that roughly

$$100,000 \times \frac{1}{500} = 200$$

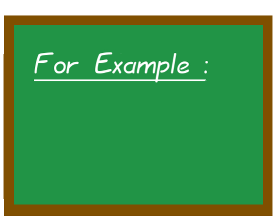
of their customers will have their homes destroyed each year.

Next, they can figure out how much to charge everyone in order to accommodate those 200 customers whose houses must be rebuilt. The insurance company must charge enough to be able to rebuild all those homes, plus make a profit, in the vast majority of years.



The theory of probability enables insurance companies to perform computations like the one in the previous box. Moreover, one can "factor in" special considerations, like proximity to a river or a fault line, or if the house is made of brick and therefore unlikely to burn completely.

What probability does not do, is given a list of 100,000 names, allow you to predict which 200 or so people will lose their houses. If you can understand this distinction, then you are already on your way to understanding probability as a subject.

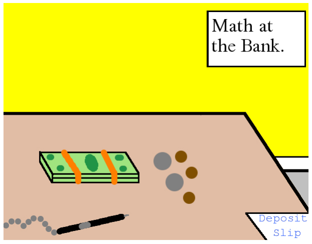


7-5-1

Imagine that a very large insurance company has insured 10 million homes. As we mentioned earlier, there is roughly a 1 in 500 chance that an ordinary American house will be destroyed in a given year. How many homes should they plan on rebuilding?

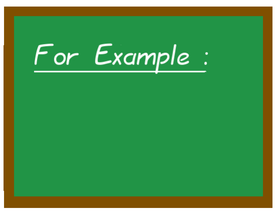
The Law of Large Numbers means that they can expect $(10,000,000)(1/500) = 20,000$ of their insured homes to be destroyed each year.

As it turns out, using a technique which we will learn later, called the Bernoulli-de Moivre-Laplace formula, we can say it is "very likely to be between 19,717 and 20,283, and extremely likely to be between 19,576 and 20,424."



That narrow interval of uncertainty is an important piece of the puzzle. It is one of the reasons why businesses care about probability. Even if the expected number of destroyed homes is 20,000, it seems very unlikely that it will be 20,000 “on the dot.” It could be 19,900 or 20,100 very easily. Society would not be able to have functioning insurance companies unless the uncertainty could be pre-computed.

What this company will do, in all likelihood, is budget for the destruction of 20,424 homes. They will do this by charging all 10,000,000 customers a relatively small dollar value, that is the *insurance premium*. Then, they'll buy a contract with a larger insurance company for the unlikely situation that it exceeds 20,424. This is called *underwriting*. It is very likely that the number of homes destroyed will be less than 20,424. Therefore, the insurance company will make a profit.

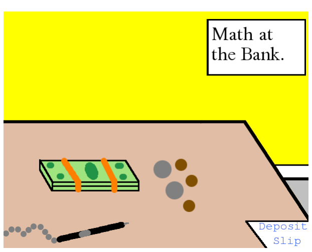


7-5-2

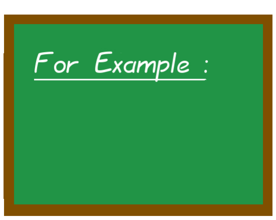
Imagine that a large restaurant in a major airport is having connection issues with their WiFi. It is often the case that there is a very large number of WiFi users in a small space in a busy modern airport. The manager is worried that customers who are having trouble connecting will simply move to a different restaurant, and that she'll lose their business. Her restaurant typically has 8000 guests per month. If analyzing the WiFi logs reveals that 5% of customers were unable to connect, how many customers (per month) should she expect will be unable to connect?

She should expect $(8000)(0.05) = 400$ customers per month will be unable to connect, though it might be slightly fewer or slightly more.

Using the Bernoulli-de Moivre-Laplace formula (which we will learn on Page 958), we can say it is very likely to be between 361 and 439 frustrated customers, and extremely likely to be between 341 and 459 frustrated customers.



Now the manager has a business decision to make. If upgrading the WiFi (perhaps by getting a more powerful base station, getting more bandwidth from the service provider, or perhaps by getting a more efficient router) costs more than the loss of 439 customers, then she should just accept the loss. However, if the upgrade costs less than the loss of 361 customers, then she should definitely carry out the upgrade. The space in between represents where judgement will come into play.

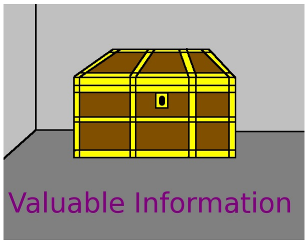


7-5-3

Suppose that a person with an exceptionally large amount of patience and free time were to flip a fair coin 10,000 times. The probability of getting heads or tails turns out to be 1/2. How many tails should he expect?

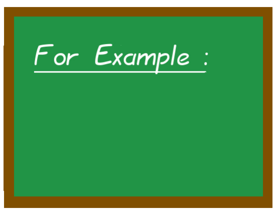
That person should expect $(10,000)(1/2) = 5000$ tails. However, it is unlikely to be exactly 5000. It could easily be 4951, 5027, or many possible numbers near 5000.

Using the Bernoulli-de Moivre-Laplace formula (which we will learn on Page 958), we can say it is very likely to be between 4900 and 5100, and extremely likely to be between 4850 and 5150.



You might be wondering if the words "very likely" and "extremely likely" have a definition. They do—here "very likely" means 95.45% probability in this context, and "extremely likely" means 99.73%. Mathematicians love to define everything. This values 95.45% and 99.73% come from statistics and calculus. In particular, they come from a very important integral.

We do not have the time to explain why it is 95.45% and 99.73% right now. Nonetheless, for those of my students who have taken statistics, you might be interested to know it comes from "the 68–95–99.7 rule."



7-5-4

A rule of thumb in Aerospace Engineering is that any given rocket launch of a satellite has a 2% probability of a catastrophic failure (exploding on the launch pad, falling into the ocean, and so forth). An ambitious satellite launching program might plan to eventually have 1000 launches, after many years. How many disasters should they expect?

They should expect $(1000)(0.02) = 20$ disasters, though it might be fewer or more.

Using the Bernoulli-de Moivre-Laplace formula (which we will learn on Page 958), we can say it is very likely to be between 11 and 29 destroyed satellites, and extremely likely to be between 6 and 34 destroyed satellites.

Did you notice how the interval was a lot larger, as a percentage, for the satellites than for the coins? This is why we call it "the law of large numbers" rather than "the law of sort-of biggish numbers." Let's look at the width of the interval of uncertainty of the last four examples, as a percentage of the expected value. To keep things simple, let's look at the "very likely" values.

Example Topic	n	width	$\frac{\text{width}}{n} \times 100\%$
Satellites	1000	$29 - 11 = 18$	1.80000%
WiFi Customers	8000	$439 - 361 = 78$	0.975000%
Coin Flips	10,000	$5100 - 4900 = 200$	2.00000%
Destroyed Houses	10,000,000	$20,283 - 19,717 = 566$	0.00566000%

By the way, if you looked at the "extremely likely" values, you'd see the same phenomenon. The n needs to be fairly large for "the law of large numbers" to work.

Here's another way of driving the point of the previous box into your head.

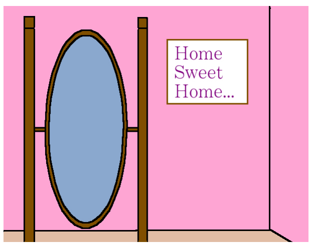
- If you flipped a coin one billion times, then we expected 500,000,000 tails and it is very likely to be between 499,968,377 tails and 500,031,623 tails. That's a width of 63,244, which is 0.0126488% of 500,000,000.
- Contrastingly, if you flipped a coin one hundred times, then we expect 50 tails and it is very likely to be between 40 tails and 60 tails. That's a width of 20, which is 40% of 50.

This happens because one billion is a large number, and one hundred is not a large number.

I'd like to address an important issue now. It is extremely likely that you've seen the topic of probability before, in other university classes, in high school, or even in middle school. Many of my students will say, "Probability is easy! What's the big deal? Why do we need these large readings for something that simple?!"

In the end, I am compelled to agree that probability is easy, as a topic, overall. However, there are some pitfalls that can trap you. These pitfalls can even trap very experienced professional scientists if they are not paying attention.

As far as students in my own courses are concerned, I have gone back in painstaking detail over their final exams. I analyzed the errors to try to see if there were some common mistakes, usually in mid-summer. For algebra, that analysis lead to Diagnostic One and Diagnostic Two, which you saw in Chapter Zero of this textbook. For probability, I have found five core errors that are remarkably common. They will be listed at the end of this module, on Page 952.



A Pause for Reflection. . .

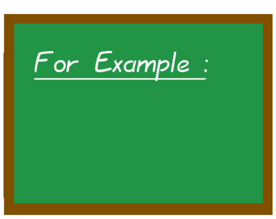
The remark which denigrates probability that I hear most often is "Isn't probability just all about problems where you divide two numbers? What's the big deal? It is just division!"

Technically, this is true. Most probability problems come down to a few very simple arithmetic operations. However, you have to pick the correct numbers for the division or other arithmetic step. As we are about to see, that isn't always easy.

We are about to consider an example with a very simple data set. From that simple data set, a whole host of questions can be asked. This will reveal that while probability might sometimes seem trivial, it is actually a very rich and deep topic.

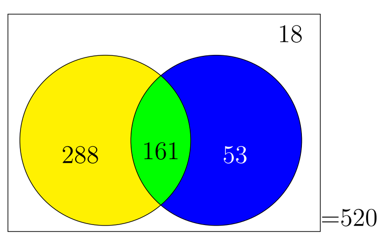
Between January 31st and February 2nd, 2015, the magazine *The Economist* surveyed 998 Americans with the following question: "Do you think the government should or should not require parents to have their children vaccinated against infectious diseases (e.g. measles, mumps, whooping cough)?" I found the results on the magazine's webpage. The survey included three answers: "Should Require," "Not Sure," and "Should Not Require." The survey also asked the respondent their political party, with the answers "Democrat," "Independent," and "Republican."

For our first attempt at this problem, to keep it simple and to keep the problem from becoming huge, we will exclude the "Not Sure" responses and any respondents who remarked "Independent." Otherwise, the problem would be much longer than it already is.



7-5-5

Here is a Venn Diagram representing the data from the survey in the previous box. There are 520 responses in the reduced data set. The diagram is below, and the left circle represents respondents who say that vaccines should be required, and republicans are the circle on the right.



From this one very simple Venn Diagram, we can ask a whole host of questions. In fact, over the next few boxes, we will explore 16 different questions from this one data set. Some of my readers will be able to answer these questions without difficulty. Other readers will be unable to answer them. If you find yourself unable to answer these 16 questions correctly, then do not be discouraged, but complete the module, and return to retry these 16 questions after you have finished with the rest of the module.

Play Around With it...

7-5-6

Referring back to the previous box, answer the following small questions.

- How many republicans are in the data set? [Answer: 214.]
- How many democrats are in the data set? [Answer: 306.]
- How many people in the data set wish to require vaccines? [Answer: 449.]
- How many people in the data set wish to not require vaccines? [Answer: 71.]

We will continue in the next box.

Play Around With it...

7-5-7

Continuing with the previous box,

- What is the probability that a random republican is in favor of requiring vaccines? [Answer: $161/214 = 75.2336\cdots\%$.]
- What is the probability that a random republican is not in favor of requiring vaccines? [Answer: $53/214 = 24.7663\cdots\%$.]
- What is the probability that a random democrat is in favor of requiring vaccines? [Answer: $288/306 = 94.1176\cdots\%$.]

We will continue in the next box.

Play Around With it...

7-5-8

Continuing with the previous box,

- What is the probability that a random democrat is not in favor of requiring vaccines? [Answer: $18/306 = 5.88235\cdots\%$.]
- What is the probability that a random person in the data set is in favor of requiring vaccines? [Answer: $449/520 = 86.3461\cdots\%$.]
- What is the probability that a random person in the data set is not in favor of requiring vaccines? [Answer: $71/520 = 13.6538\cdots\%$.]
- What is the probability that a person in favor of requiring vaccines is a democrat? [Answer: $288/449 = 64.1425\cdots\%$.]

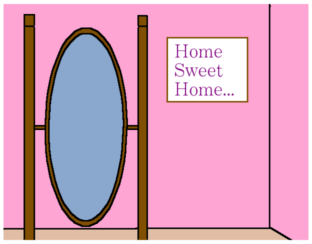
We will continue in the next box.

Play Around With it...

7-5-9

Continuing with the previous box,

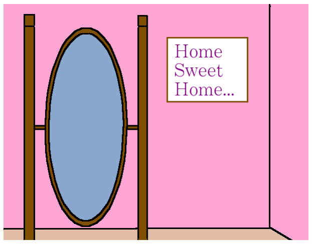
- What is the probability that a person in favor of requiring vaccines is a republican? [Answer: $161/449 = 35.8574\cdots\%$.]
- What is the probability that a person not in favor of requiring vaccines is a democrat? [Answer: $18/71 = 25.3521\cdots\%$.]
- What is the probability that a person not in favor of requiring vaccines is a republican? [Answer: $53/71 = 74.6478\cdots\%$.]
- What is the probability that a random person in the data set is a democrat? [Answer: $306/520 = 58.8461\cdots\%$.]
- What is the probability that a random person in the data set is a republican? [Answer: $214/520 = 41.1538\cdots\%$.]



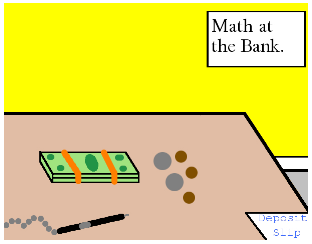
A Pause for Reflection...
In conclusion, yes, most probability questions just are a matter of dividing two numbers. However, as you can see from the previous box, there are a lot of possible questions that can be asked about even an easily imaginable situation: a Venn Diagram with two circles.

It is very common for someone to have a question in front of them, but to imagine that a slightly different question had been asked. Therefore, it is necessary to dwell on the most subtle distinctions in the wording. The tiniest change in phrasing can change the answer completely.

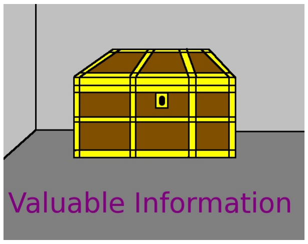
Remember, if you found yourself unable to correctly answer those 16 questions (about the vaccines), then do not be discouraged. Just complete the module, and return to retry those 16 questions after you have finished with the rest of the module.



A Pause for Reflection...
Here's another interesting perspective on those 16 questions that we just finished. Imagine that you happen to be a political writer, trying to write an article or a blog post about the survey that we just discussed. Then you have a large collection of twelve probabilities that you can choose from, to paint whatever picture you might want to paint.



By the way, now would be a good time to mention that probabilities can be written as fractions, decimals, or percentages. For example, it is common to see 75%, 3/4, or 0.75. Each is entirely acceptable (unless an examination question were to specifically note otherwise), and all three formats can be found in higher-level textbooks on economics, finance, or business.



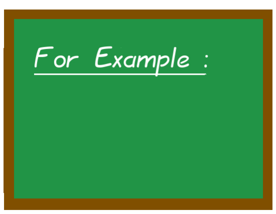
Now that we've established that probability is not a trivial topic, I'd like to introduce the first building block of probability theory.

A *sample space* is a set of outcomes connected to a probability problem or experiment. The outcomes must be *mutually exclusive* as well as *collectively exhaustive*.

- The "mutually exclusive" part means that it can never be the case that two outcomes in the space happen simultaneously. In other words, during one trial, if you know that one outcome has occurred, then you know that none of the other outcomes have occurred. (A mathematician would say "at most one outcome occurs.")
- Next, "collectively exhaustive" means that at least one of the outcomes will occur during one trial. No possibility is left out. It cannot be the case that none of the outcomes occur. (A mathematician would say "at least one outcome occurs.")
- Together, these two criterion guarantee that "exactly one outcome occurs" for any trial of the experiment.

Don't be alarmed if that's a strange definition at first glance. Just take a set of outcomes, and check to see:

- Are the outcomes mutually exclusive?
- Are the outcomes collectively exhaustive?
- If both criteria are met, then the set of outcomes is a sample space.



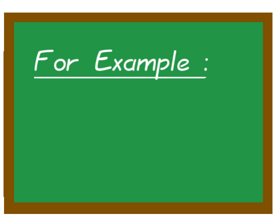
7-5-10

Suppose I'm going to begin a probability problem that involves a coin flip. The outcomes are "heads" and "tails." Is this a sample space?

I must ask myself two questions. Are the outcomes mutually exclusive? Yes, you cannot possibly get "heads" and "tails" simultaneously. Are the outcomes collectively exhaustive? Yes, there's no other side to the coin other than the "heads" side and the "tails" side.

Therefore, to be maximally formal, I can write

- Let "H" be the outcome "heads."
- Let "T" be the outcome "tails."
- The sample space is $\mathcal{S} = \{H, T\}$.



7-5-11

Now I'm going to study a sequence of three consecutive coin tosses, done in order. What is my sample space?

I'm going to write

$$\mathcal{S} = \{\{HHH\}, \{HHT\}, \{HTH\}, \{HTT\}, \{THH\}, \{THT\}, \{TTH\}, \{TTT\}\}$$

and as you can see, the set $\mathcal{S}$ covers every possible outcome, thus $\mathcal{S}$ is collectively exhaustive. Similarly, $\mathcal{S}$ is mutually exclusive, because no two of these outcomes could occur simultaneously. Therefore, $\mathcal{S}$ is a sample space.

Another valid sample space would be $\mathcal{N} = \{0, 1, 2, 3\}$, representing the number of heads obtained. Of course, I'm not as likely to get "all heads" or "no heads" as the other two outcomes. However, if I only care about how many heads I got, then this notation would be more compact.

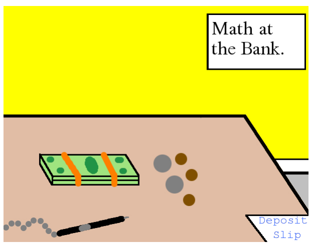
The choice of using $\mathcal{N}$ or $\mathcal{S}$ would really depend on what questions I was hoping to answer about the coin tosses.

In the previous box, you might be able to reason out the eight possible outcomes. However, it might be challenging for you to list all the outcomes if I were studying a sequence of 6 coin flips. (There are 64 outcomes in that case, by the way.)

That's okay, because later in the chapter we will learn tools that will allow us to solve such problems without listing out every possible outcome. For 20 coin tosses, there are over one million cases in the sample space—so we definitely do not want to list them! In fact there are

$$2^{20} = 1,048,576$$

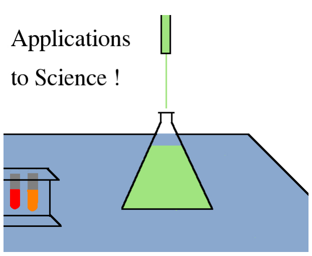
possibilities for a sequence of 20 coin flips. Surely, we do not want to list them.



The last few boxes about coin flips might seem childish and unrelated to business or science. However, suppose I am managing a smartphone factory, and that I am worried about defective parts. No manufacturing process is free of defects, but good industrial practices can manage them, making them either very rare or easily detectable. To mathematically work with this, I want to model a series of outcomes, namely "defective" or "functional."

During this chapter, you will learn how that is done by modeling the problem as a series of tosses of a weighted coin. For example, if 1% of the phones are defective, and 99% are good, then I will model each phone as a highly weighted coin that is 99% likely to fall on one side, and 1% likely to fall on the other. A batch of 500 phones would be represented by 500 coin tosses.

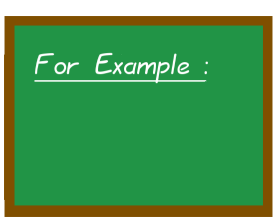
If the defect rate is too high, the company's reputation will be damaged and profits will be lost. However, it might be very expensive to get the defect rate extremely low. Probability theory allows management to navigate this balancing act.



Likewise, you might be working for a non-profit. Perhaps you are monitoring how many aid workers, among a set of 500 aid workers exposed to a rare virus like Ebola, are actually going to catch Ebola. Most of the time, either someone gets infected or they do not get infected. This too will be modeled as a series of coin flips of a weighted (i.e. unfair) coin. As you can see, many diverse problems can be modeled as a series of coin flips.

Furthermore, in diseases where there are three outcomes ("infected," "carrier," and "not infected,") probability theory can also work very well—it is only slightly harder.

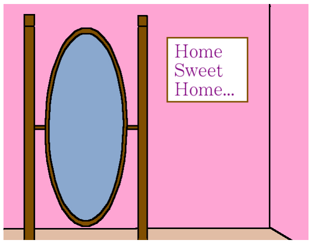
In any case, aid workers have to balance the risks with the "payoff" of saving lives. Volunteers and aid professionals often speak of "a calculated risk." Naturally the calculation is performed with probability theory.



7-5-12

In the USA, college admissions can be complicated. However, one simple issue is to forecast whether or not an admitted student will accept admission and attend the university, or decline. For example, if a university has 1000 freshman spots to fill, they must not offer admission to only 1000 students. Instead, they have to offer admission to more, because some will decline. Perhaps a student will decline in order to attend a different university, or perhaps the student will decline to attend any university at all. The reason for the decline is irrelevant. Probability theory helps to advise the admissions office on how many offers of admission they should make. For now, let us ask ourselves, "What is the sample space for this problem?"

We can write A for attend, and D for decline. For this problem, the sample space $\mathcal{S} = \{A, D\}$ is mutually exclusive and collectively exhaustive.

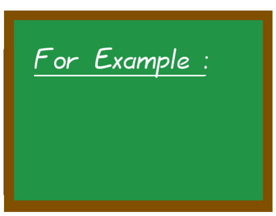


A Pause for Reflection...
Despite the utility of probability to science and business, the heritage of probability as a subject lies solidly within the analysis of gaming in general, and gambling in particular. Moreover, gambling provides us with a rich array of relatively simple problems to practice on. For this reason, perhaps your exposure to probability in high school or middle school was mostly a sequence of problems about coin flips, dice, and cards.

Unsurprisingly, many students do not pay much attention to such topics. That's completely understandable. Yet, probability is about much more than coin flips, dice, cards, and gambling! Indeed, there are many applications that have nothing to do with games. The applications include reliability engineering, statistics, risk management, insurance, and anything related to surveys—whether they are marketing surveys, political surveys or any other kind of surveys.

Moreover, I do not gamble; in my experience, students who successfully learn the content of this chapter are very unlikely to gamble with medium or large amounts of money. I completely understand that some students have a distaste for gambling. A few students even believe gambling to be gravely unethical, (e.g. to wager money on the throw of dice when that money can be instead used to aid the poor.)

For students who actually do want to learn about dice games and gambling, such as students in the UW Stout Game Design & Development program, I have a separate module which I wrote just for them. That module is called "Probability, Dice Games, and Odds," and begins on Page 1069. However, since almost all other students do not wish to learn about dice and gambling, I have avoided those topics throughout the vast majority of this chapter.



7-5-13

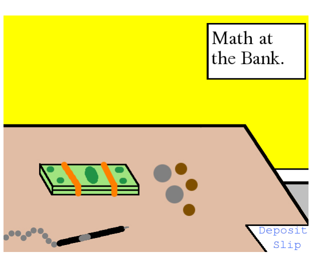
Let's start with a simple situation. There is a customer service team working at the phones, to answer questions. The members of the team are Alice, Bob, Charlie, Diane, Edward, Frank, Greg, and Harriet. A computer randomly assigns each incoming call to a customer service representative, and each representative is equally likely. Who will the next incoming caller get? What is the sample space?

$$\mathcal{S} = \{\text{Alice}, \text{Bob}, \text{Charlie}, \text{Diane}, \text{Edward}, \text{Frank}, \text{Greg}, \text{Harriet}\}$$

However, it might be nice to abbreviate the names, writing instead

$$\mathcal{S} = \{A, B, C, D, E, F, G, H\}$$

Suppose we are furthermore asked why we believe that $\mathcal{S}$ is mutually exclusive and collectively exhaustive. We would say that S is mutually exclusive, because it is not possible for a caller to be assigned multiple team members. We would say that S is collectively exhaustive because it is not possible for an incoming caller to fail to be assigned to someone.



It is interesting to note that at times, a business might cause random assignments on purpose.

This might be done in a customer service situation to make sure that some team members are not overly fatigued while other team members sit idle—the randomness divides the workload evenly among members of the team. This is particularly easy to accomplish via computer.

We've seen several examples now of sample spaces. Let's pause briefly, and summarize some of what we've seen. We'll do that over the next three boxes.

In the case of flipping a fair coin, either outcome is equally likely—H or T. Likewise, in the customer-service team problem, the eight outcomes $\{A, B, C, D, E, F, G, H\}$ were equally likely. Similarly, if you were to roll a fair die, from the six-sided dice that come with many boardgames, the six outcomes $\{1, 2, 3, 4, 5, 6\}$, are equally likely.

In flipping a weighted coin, detecting if smart phones are defective, estimating how many aid workers will get Ebola, or predicting if accepted students will accept or decline the offer of admission, the two outcomes within each sample space are not equally likely. How can we compute the probabilities in those situations? By analyzing data!

- For the smart phone factory, we can look into the logs over the past month or quarter, and see how many phones came out defective versus how many phones came out functional.
- For college admissions, we must log into the university database and perform some queries to discover what fraction of admitted students are estimated to attend, and what fraction are estimated to decline.
- For Ebola, the situation is more complicated, because sometimes one outbreak has different characteristics from the previous outbreaks. We would need to hire a microbiologist to help us.
- For flipping a weighted coin, we regrettably must take the weighted coin and flip it a very large number of times, carefully making note of the number of heads and the number of tails. This would be exceptionally boring, so let's not do that.

For flipping multiple fair coins, the outcomes in the sample space

$$\mathcal{S} = \{\{HHH\}, \{HHT\}, \{HTH\}, \{HTT\}, \{THH\}, \{THT\}, \{TTH\}, \{TTT\}\}$$

are equally likely but the outcomes in the sample space $\mathcal{N} = \{0, 1, 2, 3\}$, representing the number of heads, are not equally likely. For example, in $\mathcal{S}$ there are three outcomes which have exactly 2 heads, but only one outcome with exactly 3 heads. Later, we'll calculate that the probability of exactly getting 2 heads is 3/8, but getting exactly 3 heads has probability 1/8.

The act of assuming that the outcomes in a sample space are "equally likely" happens often, but it is also often incorrect. While "equally likely" outcomes do occur in certain probability problems (often involving dice, lottos, cards, roulette, or drawing names from a hat), it certainly is not true all the time.

I consider one of the five principle errors in basic probability theory to be "assuming the outcomes in the sample space are equally likely, when in fact they are not."

See Page 952 for the complete list of the five very common errors.

Play Around With it...

7-5-14

Suppose a colleague of mine at a not-so-great university is trying to model the attendance pattern of his class. There are 20 students in it. Each day for one semester he's going to record the number of people present. The data will then be used to generate a model. He thinks the sample space for any particular day is

$$\text{WRONG!} \rightarrow \quad \mathcal{S} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\} \quad \leftarrow \text{WRONG!}$$

but he's not exactly correct. Can you see what's wrong? (This is a hard one!)

The answer will be given on Page 964.

Play Around With it...

7-5-15

Suppose you are hired by an insurance company in California to examine the aftermath of some earthquake in order to get better data about damages, for future use. There are numerous houses covered by the insurance company that hired you, and some houses fared better than others. You find the quantity of damage as a percentage of the total value of the house. Then you decide to round to the nearest multiple of 10%. What is the sample space for any particular house? A colleague suggests (incorrectly)

$$\text{WRONG!} \rightarrow \mathcal{S} = \{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 100\%\} \leftarrow \text{WRONG!}$$

but why is he wrong?

The answer will be given on Page 965.

Play Around With it...

7-5-16

Imagine that Ned is tracking the performance of some software-company stocks, and measuring their annual performance over the last few years. Perhaps Ned will round to the nearest 5%. For a sample space, Ned suggests the following

$$\text{WRONG!} \rightarrow \mathcal{S} = \{-15\%, -10\%, -5\%, 0\%, 5\%, 10\%, 15\%\} \leftarrow \text{WRONG!}$$

and this is wrong. Why is it wrong?

The answer will be given on Page 965.

Play Around With it...

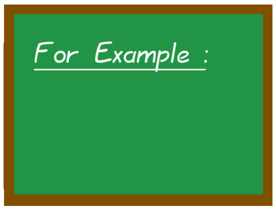
7-5-17

A particular airline requires that the check-in clerk type in the passenger's full name (first name, last name, and middle initial) during the check-in process, and it is to be typed from the passenger's photo identification, such as their driver's license or passport. Once in a while, a clerk mistypes the name, and the passenger has to be told that their reservation does not exist, or that there is no record of their reservation. As you can imagine, this will make the customer extremely irritated, if not furious. With that in mind, the IT office tries to investigate the situation, by looking at the data records of filed complaints. They discover that the middle initial is usually the problem.

Before proceeding further, an IT intern wants to use probability to analyze the middle initial. Since there are 26 letters of the alphabet, there are 26 possible middle initials. We want to determine if this is a good sample set.

- Is this set mutually exclusive?
- Is this set collectively exhaustive?
- Do you think that the "equally likely assumption" applies here? or not?

The answer is given on Page 965. Please go have a look, because there is an interesting anecdote embedded in the answer.



7-5-18

Suppose you are a manager at a warehouse, and you are analyzing the interaction of two teams, and how absenteeism impacts their productivity. This often can happen in jobs that are physically laborious, not entertaining, and paying the minimum wage—employees are often late, quit suddenly, or simply do not show up for work. Each team of six workers can have 0–6 absences on any given day, giving a sample space of

$$\mathcal{S} = \{0, 1, 2, 3, 4, 5, 6\}$$

if we want to analyze each team individually.

Suppose we want to compute the probabilities of various outcomes, operating the teams as a pair. One option is to calculate the number of ordered pairs. The first member of the ordered pair is the number absent from team one, and the second is the number absent from team two. Then we'd have 49 outcomes, in a much bigger sample space:

$$\mathcal{B} = \left\{ \begin{array}{ccccccc} (0,0); & (0,1); & (0,2); & (0,3); & (0,4); & (0,5); & (0,6); \\ (1,0); & (1,1); & (1,2); & (1,3); & (1,4); & (1,5); & (1,6); \\ (2,0); & (2,1); & (2,2); & (2,3); & (2,4); & (2,5); & (2,6); \\ (3,0); & (3,1); & (3,2); & (3,3); & (3,4); & (3,5); & (3,6); \\ (4,0); & (4,1); & (4,2); & (4,3); & (4,4); & (4,5); & (4,6); \\ (5,0); & (5,1); & (5,2); & (5,3); & (5,4); & (5,5); & (5,6); \\ (6,0); & (6,1); & (6,2); & (6,3); & (6,4); & (6,5); & (6,6) \end{array} \right\}$$

Alternatively, we could just add the number absent together. In that case, the outcomes would be the sums of the possible ordered pairs. Then we would instead have

$$\mathcal{T} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

As you can see, each of $\mathcal{S}$, $\mathcal{B}$ and $\mathcal{T}$ are mutually exclusive and collectively exhaustive. They are all valid sample spaces. We will continue the discussion in the next box.

Continuing with the previous box, depending on what you need to analyze, it is not clear if you would want $\mathcal{S}$, $\mathcal{B}$, or $\mathcal{T}$. For example,

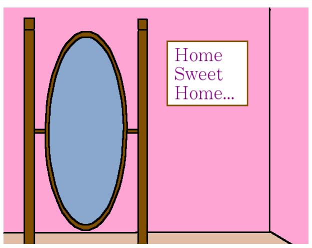
- If the two teams are often combined, and work together as a bigger team of twelve, then you would want to use $\mathcal{T}$.
- However, perhaps it is the case that the two teams do not have the same skill set. Perhaps you have a team unloading trucks and putting the boxes on carts, and a team carting the unloaded boxes to the right places inside the warehouse. The former team would be selected for strength, and the latter team selected for attention to details. Moreover, if one team had high absenteeism on a particular day, then it might well be the case that the other team cannot work, even if everyone is there. In this case, you'd want to use $\mathcal{B}$, so that you can talk about both teams simultaneously. It turns out this is somewhat complicated, and therefore, we won't do that in this book. However, it is studied in more advanced courses on probability and statistics, and is called "the joint distribution of $\mathcal{S}$ with $\mathcal{S}$."
- On the other hand, if the teams are in opposite parts of a large warehouse, and do not interact with each other at all, then you should analyze them separately. You would use $\mathcal{S}$ to do this, because it is simpler.
- In summary, the choice of which sample set to use would depend on what questions you desire to answer.

Still continuing with the previous box, even if we chose the simplest sample space,

$$\mathcal{S} = \{0, 1, 2, 3, 4, 5, 6\}$$

we might still fall victim to a common stumbling block. One of the five common errors in basic probability problems is to assume that all the outcomes are equally likely, when perhaps they are not.

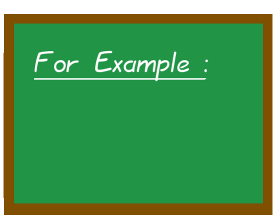
If the team has anywhere near typical rates of attendance, then either 0 or 1 absence would be the most common outcome. Even with catastrophic attendance, 2 might be the most common. In stark contrast, 5 or 6 are probably rather rare. Later, we can use the concept of "independence" in probability, as well as Bernoulli's Distribution formula (not the Bernoulli-de Moivre-Laplace formula), to determine a good model for this probability distribution.



A Pause for Reflection. . .

In many branches of mathematics, there is more than one road to an answer. In fact, in probability, this will be truly amazing. There may be several roads to the same answer, and each road might look very different from the others. In particular, the same problem can be described by different sample spaces that are shockingly distinct, yet in the end, the same answer should be obtained.

If you find that your answers *do match* those of a classmate, but your routes *do not match* those of your classmate, then do not ignore the variation in your approaches, nor panic over the variation. Instead, be sure to *recognize both ways* of solving the problem, to increase your depth of understanding of the material, as well as increase your speed on examinations.



7-5-19

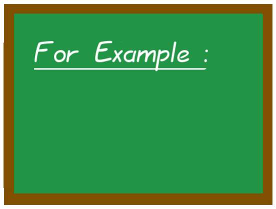
Let's return briefly to our earlier example of a smartphone factory. For simplicity, let's consider phones as either defective ("D") or functional ("F.") Once during each shift, the quality-control engineer is going to sample five phones coming off the production line. Naturally, each can be defective or functional. What is the sample space?

A particularly bad choice would be

$$\begin{aligned}\mathcal{S} = \{&FFFFF, FFFFD, FFFDF, FFFDD, FFDFF, FFDFD, FFDDF, FFDDD,\\ &FDFFF, FDFFD, FDFDF, FDFDD, FDDFF, FDDFD, FDDDF, FDDDD,\\ &DFFFF, DFFFD, DFFDF, DFFDD, DFDFF, DFDFD, DFDDF, DFDDD,\\ &DDFFF, DDFFD, DDFDF, DDFDD, DDDFF, DDDFD, DDDDF, DDDDD\}\end{aligned}$$

There are several reasons that $\mathcal{S}$ is an unwise choice. First of all, with 32 entries, it is inconvenient to write all that out. Second, it contains information we simply don't need! If two smartphones are defective out of five, then it doesn't matter if the order is $DFFDF$ or $FFFDD$. We do not want to keep track of that, because that extra work would serve no useful purpose.

Instead, a nicer sample space would be $\mathcal{N} = \{0, 1, 2, 3, 4, 5\}$ representing the total number of defective smart phones. Last but not least, in neither case (neither $\mathcal{N}$ nor $\mathcal{S}$) are the members of the sample space equally likely.



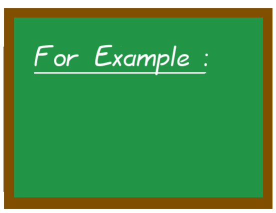
7-5-20

Suppose there are five workers on a team, and they are unhappy to learn that one of them must go to attend mandatory "fire safety training." Unfortunately, it is a busy week and there's lots of work to be done in the office, so no one wants to go. They place their names into a hat, and one name will be pulled out. That person will be the one to sit through the boring and not-very-useful training.

The people in the office are named Andrea, Bill, Chuck, Doris, and Edgar, so we can abbreviate their names by the first letter. The sample space is then

$$\mathcal{S} = \{A, B, C, D, E\}$$

As you can see, the set $\mathcal{S}$ is mutually exclusive because we only pull one name, and collectively exhaustive because we did remember to list all five workers on the team.



7-5-21

The above example is a bit over-simplified. Suppose the team has a dispute with the boss, Fred, and they're going to send a delegation of three workers to their boss's boss (Fred's boss). They are going to pull three names out of a hat containing each of their names. What does the sample space look like?

Well, we have to consider all possible triplets drawn from the 5 names. The set of all possible triplets happens to be

$$\mathcal{S} = \{ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE\}$$

which might require a moment for you to verify.

A few modules later, you will learn how to know ahead of time how many there must be, which is an immense aid in making sure that you've gotten all the possible combinations. That is how you will take care of collective exhaustion, by counting what you have written, and by ensuring that the count matches what the formulas predict. As far as mutual exclusivity, note that ABC and ABD both have A (or Andrea) in common. Is this a problem? No, because ABC and ABD are distinct outcomes. We cannot have both ABC and ABD at the same time, because that would involve four names (Andrea, Bill, Chuck, and Doris) and not three names.

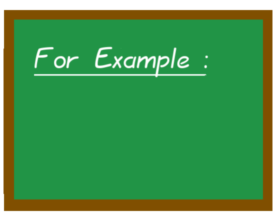
One of the things that can be maddening about probability is that small changes to the problem can have drastic consequences. In the previous problem, if Bill were drawn first, Andrea second, and Chuck third, that would be the same delegation as Chuck first, Bill second, and Andrea third. That's because a delegation is a set of people, and it doesn't matter in what order the names were drawn.

Suppose instead of choosing a 3-person delegation, the office is instead going to unionize and elect a president, vice-president, and secretary—with the president drawn first, vice-president second, and secretary third. Then it is not the same thing if Bill is first or Chuck is first. If the order is Bill-Andrea-Chuck, then Bill is president, but if Chuck-Bill-Andrea, then Chuck is president. Therefore, we would have to list all the triplets, including triplets that are identical except for ordering.

If we try to analyze the problem mentioned in the previous box, where the names of Andrea, Bill, Chuck, Doris, and Edgar are going to be drawn to select a president, a vice-president, and a secretary, then we get a mess. Do not attempt to write such a list down.

Using formulas that you will learn a few modules from now, one can calculate that there are 60 such possibilities. Each of the 10 “orderless” outcomes becomes six separate “ordered” outcomes, and $(10)(6) = 60$. It would be tedious, even pointless, to write them all down. I’ve generated the list for you, below, but this is not how we want to solve such problems. No one has the time to generate such long lists.

ABC	becomes	ABC, ACB, BAC, BCA, CAB, CBA
ABD	becomes	ABD, ADB, BAD, BDA, DAB, DBA
ABE	becomes	ABE, AEB, BAE, BEA, EAB, EBA
ACD	becomes	ACD, ADC, CAD, CDA, DAC, DCA
ACE	becomes	ACE, AEC, CAE, CEA, EAC, ECA
ADE	becomes	ADE, AED, DAE, DEA, EAD, EDA
BCD	becomes	BCD, BDC, CBD, CDB, DBC, DCB
BCE	becomes	BCE, BEC, CBE, CEB, EBC, ECB
BDE	becomes	BDE, BED, DBE, DEB, EBD, EDB
CDE	becomes	CDE, CED, DCE, DEC, ECD, EDC



7-5-22

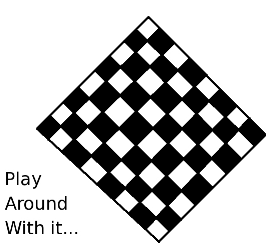
Suppose there is a survey (done before the use of the internet was common) to determine how people get the news. The choices are TV, radio, newspapers, or none. One might imagine that the sample space for one person’s responses then has four choices (outcomes), but this would be an error—it actually has eight. The reason is that someone might use both the TV and the radio, or the radio and newspapers, or all three, or perhaps only newspapers and TV. The sample space is thus

$$\mathcal{S} = \{ \quad \{\}\,;\{T\}\,;\{R\}\,;\{N\}\,;\{T,R\}\,;\{R,N\}\,;\{N,T\}\,;\{N,T,R\} \quad \}$$

which you can see has all of the 8 possible subsets of $\{T, R, N\}$.

Recall that the symbol “$\{\}$” represents “the empty set.” Here, that means someone who uses neither newspapers, TV, nor radio.

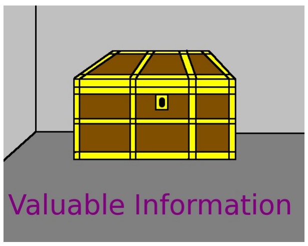
Remember, a set with 3 members, like $\{T, R, N\}$, has $2^3 = 8$ subsets. In fact, a set with n members has 2^n possible subsets. We discussed this back on Page 827 in the module “Introduction to Sets.”



7-5-23

Of course, nowadays many people take their news from the internet, so we have to adjust the sample space accordingly.

- How many outcomes will be in the sample space? [Answer: $2^4 = 16$.]
- What are they? [Answer: $\{$ $\{\}$; $\{T\}$; $\{R\}$; $\{N\}$; $\{T,R\}$; $\{R,N\}$; $\{N,T\}$; $\{N,T,R\}$; $\{I\}$; $\{T,I\}$; $\{R,I\}$; $\{N,I\}$; $\{T,R,I\}$; $\{R,N,I\}$; $\{N,T,I\}$; $\{N,T,R,I\}$ $\}$.]



Now we are extremely well-practiced with the idea of a sample space and we are ready to learn what to do with them. Probabilities can be given to the outcomes in the sample space according to four approaches:

1. One can assume that the outcomes are equally likely. This is called "the equally likely assumption" or "uniform distribution assumption", and sometimes simply "uniformity." Problems of this sort often, but not always, come from gaming.

2. One can perform a suitably large survey or large experiment and assume that the percentages observed are very close to the actual probabilities. This is a favorite method in many subjects, from economics to criminology. However, this must be done very carefully. We'll see a lot of that, later in this module.

3. One can calculate the probabilities from some other sample space and probability distribution. That's done by merging events together to form bigger events. We will learn how to do just that later, in the next module, on Page 980.

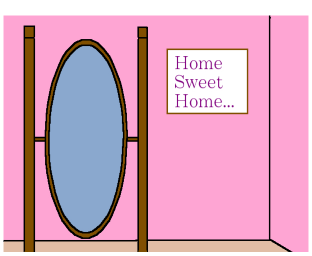
4. One can simply assume particular values. This is usually done only as a very rough approximation, or to make examination problems shorter. The assumed values must all be positive or zero, and furthermore, they must add up to 100% (or the number 1 if using decimals or fractions).

- Just to be clear, an example for # 4 might be helpful. When modeling employee tardiness in a troublesome team, a manager might estimate that employees are punctual 85% of the time, late 10% of the time, and absent 5% of the time. If these numbers are given from the manager's experience and perspective, then they have the mathematical status of assumptions.

Once all the outcomes in a sample space are assigned probabilities, the sample space is considered a *probability distribution*, regardless of which of the four mechanisms were used.

You can see that in each case, there is an assumption:

- In the first case (the equally likely assumption) we explicitly assume that the outcomes are equally likely.
- In the second case, we have a survey and we must assume that the people surveyed are a representative sample of the general population. If you want to study how Americans dress in December, it is unwise to only interview people in Minnesota. It is also unwise to only interview people in Hawaii.
- In the third case, the assumption is inherited from the previous sample space.
- The fourth case is nothing but assumptions.

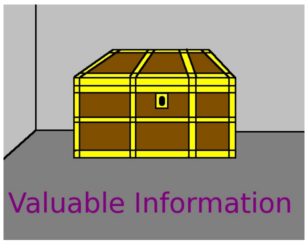


A Pause for Reflection. . .
We can summarize the last few boxes by saying that in probability, you never get something for nothing. You have to have a starting point, a foundation for your analysis.

The root of the analysis is either some sort of survey or experiment, as is common in business or science problems, or alternatively from an assumption of equally likely outcomes, as is common in gaming and gambling.

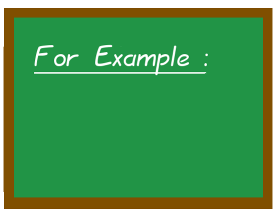
Until recently, the last three cases could be referred to by Latin names. Those names are "a priori", "a posteriori", and "per fiat." To do so today would be considered extremely pompous, except in a math paper. In fact, not too many math papers use that terminology either.

One does occasionally hear "a priori" uttered by senior mathematicians, physicists, chemists and engineers. The usage of the other terms is extremely uncommon; I mention them only for completeness. Please do not attempt to remember these Latin names.



At this point, we're going to explore problems that are modeled well by the "equally likely assumption."

A sample space of n elements, governed by the "uniform distribution assumption", also known as the "equally likely assumption," has each outcome in the sample space at probability $1/n$.



7-5-24

Three simple examples to clarify the above would be:

- A fair coin, as we mentioned earlier, has a sample space of $\mathcal{S} = \{H, T\}$, and so both H and T have probability $1/2$.
- The sequence of three flips of a coin has a sample space of

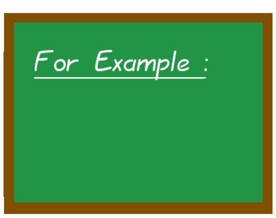
$$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

therefore, each outcome has a probability of $1/8$.

- From the customer service team on Page 929, where callers were assigned to team members randomly as they called in, we had the sample space

$$\mathcal{S} = \{\text{Alice}, \text{Bob}, \text{Charlie}, \text{Diane}, \text{Edward}, \text{Frank}, \text{Greg}, \text{Harriet}\}$$

and each outcome (each person) has probability $1/8$.



7-5-25

Sometimes it is nice to make a list or chart of the outcomes in a sample space, along with their probabilities. That list or chart is called a *probability distribution*. Let's write down a probability distribution for the customer service team in the previous box.

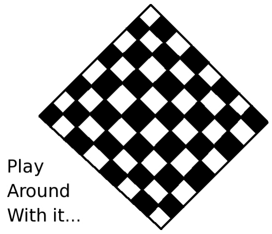
Since the customer service representative is assigned by a computer so that each representative is equally likely, we have

Team Member:	A	B	C	D	E	F	G	H
Probability:	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Similarly, for the flipping of a fair coin, we would have

- Heads: 1/2
- Tails: 1/2

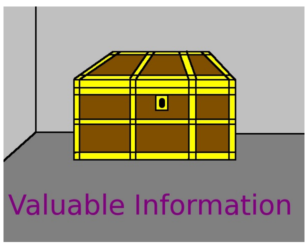
That's all a probability distribution is. It is a set of outcomes that form a sample space, along with the probabilities of those outcomes.



7-5-26

Suppose that Xavier DuBois is checking in at the airport. He notices that his 6-letter "record locator" happens to begin with an X. This cheers him up, because he often thinks about how rare the letter X is. He wonders what the probability is that a record locator will begin with an X. Note, a record locator is just a sequence of six random letters, taken from the alphabet.

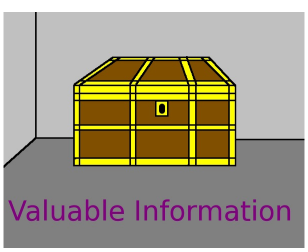
What is the probability that a random record locator will have an X as its first letter? [Answer: 1/26.]



The concept of an event is very useful in solving probability problems. For any probability distribution, any subset of the set of outcomes is called an *event*.

- A simple event has 1 outcome. A simple event is a "singleton set" consisting of one outcome. The probability of the simple event is just the probability of that outcome.
- A compound event has more than one outcome; it is a set of two or more outcomes. The probability of the compound event is the sum of the probabilities of the outcomes that it contains.

However, there are two technicalities which I must mention now. They are stated in the next box.

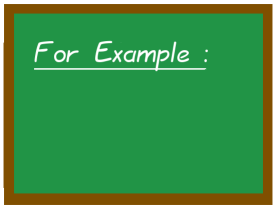


There are two technicalities that are important to understand about events.

1. Technically, the empty set is considered an event. That's because the empty set is a subset of any set whatsoever—including the set of outcomes in a probability problem. For technical reasons, this "trivial subset" is given the probability of 0.
2. The set of all the outcomes together (i.e. the sample space), is also considered an event. That's because $\mathcal{S} \subseteq \mathcal{S}$ for all sets $\mathcal{S}$ in all of set theory—including the set of outcomes in a probability problem. For technical reasons, this "improper subset" is given the probability of 1.

The last technicality has a useful consequence. The sum of all of the probabilities of all the outcomes must be exactly 1, for any probability distribution. This fact will be useful on many occasions while solving problems.

Note, we first talked about "singleton subsets," "proper subsets," and "non-trivial subsets," on Page 812, Page 825, and Page 826. Feel free to review those definitions, if you feel that you need to.



7-5-27

When trains arrive at a busy train station, an important task for the computer running the station is to dispatch them to empty tracks. Suppose a train is pulling into a train station and the computer (or alternatively the signaling system connecting the train and the computer) has failed. The train's driver must choose a track at random, because it is very difficult for trains to stop quickly, or even moderately quickly. There are 20 tracks, and 3 of them are occupied. What is the probability that the train chooses an occupied track, and crashes? What is the probability that a train chooses an unoccupied track, and arrives safely?

Since the driver is choosing randomly, we can assume that all the tracks are "equally likely." There are 3 of them that are occupied out of 20, so the probability of a collision caused by two trains sharing the same track is given by the following:

$$3/20 = 0.15 = 15\%$$

which is far too high of a risk to bear.

We will continue in the next box.

In the previous box, we computed the probability that the incoming train will pick a track that is already occupied. Now we should compute the probability that the incoming train picks a track that is not already occupied, therefore avoiding a crash.

We know that $20 - 3 = 17$ tracks are unoccupied. Therefore, the chance of picking an unoccupied track and not crashing is given by the following

$$17/20 = 0.85 = 85\%$$

which is far too low.

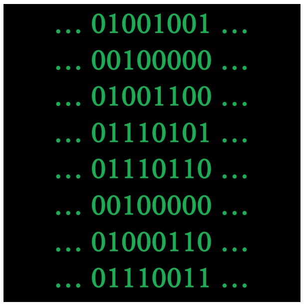
The way to analyze the problem in the previous box is to realize that we're talking about an interesting set—the set of train tracks that could be chosen, a set of size 20. That will be the denominator of our fraction. Then we are asking about a subset of the interesting set, namely the subset with some property.

In the first calculation, we are asking about the subset of tracks that happen to already have a train on them, a subset of size 3. That was our first numerator, resulting in 3/20.

In the second calculation, we are asking about the subset of tracks that happen to not already have a train on them, a subset of size $20 - 3 = 17$. That was our first denominator, resulting in 17/20.

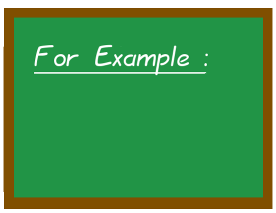
By the way, the "interesting set" must always meet the requirements of being a sample space. As we saw many times earlier in this module, it is possible for a probability problem to have many different sample spaces—all of which are valid. You have to look to the heart of the problem and its context to be able to determine what the interesting set is. That insight does not come from some set of rules in probability theory, but rather by trying to visualize what is happening inside of the problem.

Last but not least, please understand that "interesting set" is not a technical or vocabulary term. It is merely a set that is interesting in the context of a given problem.



By the way, the previous example illustrates why the computers which run important transportation systems have protections to make sure they are always operating, including being attached to a UPS (*Uninterruptible Power Supply*). Many business people are unfamiliar with the existence of such power supplies, which usually guarantee anywhere from 30 minutes to 2 hours of power in the event of a power failure. This can help your business survive a minor natural disaster, like a hurricane or an earthquake, which is why I am bringing it to your attention.

Many of the techniques of this chapter and the next are useful in the design of high-reliability systems. In fact, the mathematics department in which I teach (the University of Wisconsin—Stout) lost a faculty member to a startup company that uses mathematical software to help analyze and design ultra-reliable digital equipment for critical systems like mass-transit controls.

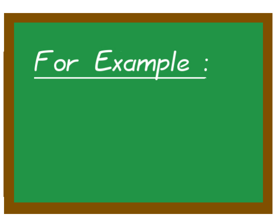


7-5-28

Earlier, on Page 934, we had the names Andrea, Bill, Chuck, Doris, and Edgar. They were putting their names in a hat, so that one name drawn out of the hat would be their representative to attend the fire-safety training. What is the probability that the representative will be male? What is the probability that the representative will be female? What is the probability that the representative will have the letter "e" in their name?

These three questions are compound events. We will abbreviate each name by its first letter. The sample space is $\{A, B, C, D, E\}$, and since the names are all equally likely, each outcome has probability $1/5$.

- The males are $\{B, C, E\}$, so the probability is $3/5$ that a male will be chosen.
- The set of females is $\{A, D\}$, and so the probability is $2/5$ that a female will be chosen.
- The set of persons with "e" in their name is $\{A, E\}$, so the probability is $2/5$.



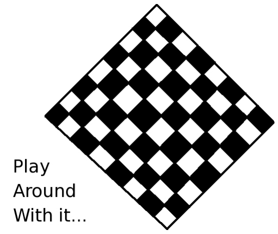
7-5-29

Marlann, Laura and Alan are taking an economics class, and they have a small test today. The test is going to consist of one homework question from each of three chapters, chosen randomly from the set of assigned homework questions. The three students are really only worried about Chapter 5, as they've studied the other chapters very well. In Ch 5, section 5.1 had 20 questions, section 5.2 had 10 questions, section 5.3 had 15 questions, and section 5.4 had 5 questions. As it comes to pass, Alan studied 5.1 and 5.2, but not 5.3 and 5.4; Laura studied all the sections except 5.4; Marlann studied all the sections, without exception. We now wish to compute the probabilities that Marlann, Laura, and Alan get a question for Ch 5 drawn from a section that they studied. Of course, since they studied different amounts, the probabilities will not be the same.

Using the method of the previous box, we can identify that our interesting set is the set of questions from Chapter 5, and there are $20 + 10 + 15 + 5 = 50$ of them.

The numerator for Alan will be a subset of those 50 questions, namely those questions that come from sections which Alan studied. What is the size of that subset? The subset has $20 + 10 = 30$ questions in it. Therefore, the probability that Alan gets a question for Ch 5, drawn from one of the sections which he studied, is $30/50 = 0.6$.

In the next box, you will compute the probabilities for Laura and for Marlann.



7-5-30

Looking at the previous box, compute the probabilities that Marlann or Laura gets a question for Chapter 5, from a section that he or she has studied.

- What is the probability for Laura? [Answer: $45/50 = 0.9$.]
- What is the probability for Marlann? [Answer: $50/50 = 1.0$.]

As you can see, it is wise to make sure that you can answer all of the questions in a chapter, before showing up to a test.

In the last few problems, we made use of the "equally likely assumption."

- For the example about the train pulling into a train station with 20 tracks, the train tracks were equally likely. Each had probability $1/20$. We essentially added 3 copies of $1/20$ to get the probability of $3/20$ of the train driver unluckily picking an occupied track, and essentially added 17 copies of $17/20$ to get the probability of the train driver luckily picking an unoccupied track.

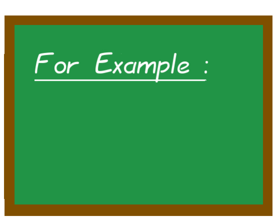
$$\underbrace{\frac{1}{20} + \frac{1}{20} + \frac{1}{20} + \cdots + \frac{1}{20}}_{\text{17 copies}} = \frac{17}{20}$$

- For the example about the automatically generated test, there were 50 questions in chapter 5. Since the questions are chosen at random, each had probability $1/50$. We essentially added 30 copies of $1/50$ to get the probability of $30/50 = 0.6$ for Alan.

$$\underbrace{\frac{1}{50} + \frac{1}{50} + \frac{1}{50} + \cdots + \frac{1}{50}}_{\text{30 copies}} = \frac{30}{50} = 0.6$$

- Likewise, we added 45 copies and 50 copies of $1/50$ to get $45/50 = 0.9$ and $50/50 = 1.0$ for Laura and Marlann.

In the next box, we'll see how to take a similar approach in problems where the outcomes are not equally likely.



7-5-31

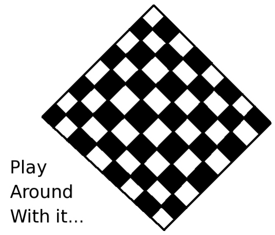
Some problems involve assuming various probabilities for the outcomes in a sample space. These are often educated guesses, but they might be based on extensive experience. For example, a manager might be frustrated with a particular cluster of employees (perhaps stock-room workers). He might simply state that there's a 5% chance one of them will be absent on a given day; a 10% chance that they'll be late; and an 85% chance they'll be on time. We could ask what is the probability that such an employee is "present," i.e. either late or on time, but not absent.

First, we ask ourselves if the set of outcomes "absent," "late," and "on time" represent a sample space. They appear to be mutually exclusive and collectively exhaustive, therefore, the set is a sample space. Second, we should verify that the probabilities do add to 100%.

$$5\% + 10\% + 85\% = 100\%$$

Since they add to 100% and none are negative, this is a valid probability distribution.

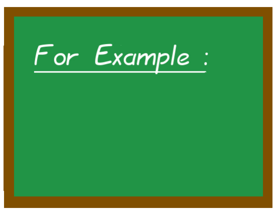
Third, we must realize that the event we are asked about is a compound event, consisting of the outcomes "late" and "on time." The probability of this event "present" is easily calculated by $10\% + 85\% = 95\%$.



\# 7-5-32

Later in this chapter, we will be able to deal with much more complicated and therefore useful calculations. Unfortunately, we don't have the tools for that just yet. For now, I present you with an easy question.

Returning to the stock-room employees of the previous box, suppose there is a form that the supervisor must fill out when a stock-room employee is either absent or late. What is the probability that the form must be used, for any particular employee on any particular day? [Answer: 15%.]



\# 7-5-33

This problem was suggested by Prof. Mark Fenton. Suppose a father and young son attend a sporting event together, and there is a raffle for a football helmet worn by a famous athlete. The young son really wants the helmet, and naturally the father will be buying raffle tickets. It turns out that the drawing is coming up soon. The raffle-ticket salesman tells the dad that he can purchase as many tickets as he likes, but then those will be the last tickets sold—he's going to the stage right away, so that the winning ticket can be drawn. The raffle-ticket salesman also mentions that only 35 tickets have been sold. How many raffle tickets would the father have to buy in order to have a 60% chance of winning? 70%? 80%? 90%?

The key to this problem is to realize that if the father buys x tickets, then $35 + x$ tickets have been bought. The set of tickets bought comprise the interesting set, and we're concerned with the subset of tickets that were bought by the dad—there are x of those. Therefore, for any x, the probability that the father wins will be given by

$$f(x) = \frac{x}{35+x}$$

which we can use to find the sample probabilities. We'll continue in the next box.

Let's consider how many tickets the father would have to buy, in order to cause a 60% probability of success.

$$\begin{aligned} f(x) &\geq 0.6 \\ \frac{x}{35+x} &\geq 0.6 \\ x &\geq (0.6)(35+x) \\ x &\geq 21 + 0.6x \\ 0.4x &\geq 21 \\ x &\geq 21/0.4 \\ x &\geq 52.5 \end{aligned}$$

Of course, since he cannot buy half a ticket, he would have to buy 53 tickets to obtain a 60% chance of success.

Before we continue, I should mention a technicality. We only have the right to multiply both sides by $35 + x$ in an inequality, if we can be certain that $35 + x$ is always positive.

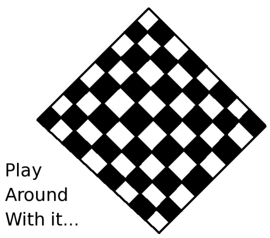
Note, it is impossible for the father to buy a negative number of raffle tickets. Since $x \geq 0$, adding 35 to both sides tells us that $35 + x \geq 35$. In plain English, if the father buys zero tickets, then $35 + x$ is positive; if he buys some non-zero number of tickets, $35 + x$ gets bigger, so it certainly stays positive. Therefore, $35 + x$ is always positive for this problem, and we were justified in what we did.

Admittedly, most students would not notice this, but would solve the problem correctly anyhow.

Let's check our work from the previous example. If we've solved the problem correctly, then 52 tickets purchased should result in a probability just under 60%, and 53 tickets purchased should result in a probability just over 60%. Let's see if that's true.

- $f(52) = \frac{52}{35+52} = \frac{52}{87} = 0.597701\cdots$
- $f(53) = \frac{53}{35+53} = \frac{53}{88} = 0.6022\overline{27}$

Alright, now we are certain that we've successfully solved for 60%. I will let you solve for 70%, 80%, and 90% yourself, in the next box.

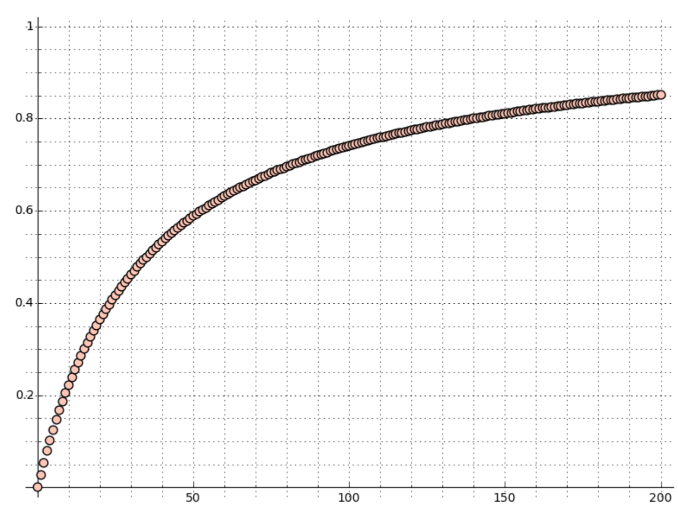


\# 7-5-34

Continuing with the previous example, how many tickets would the father have to buy in order to win...

- ...with probability 70%? [Answer: 82 tickets.]
- ...with probability 80%? [Answer: 140 tickets.]
- ...with probability 90%? [Answer: 315 tickets.]
- ...with probability 99%? [Answer: 3465 tickets.]

It is shocking how many tickets must be purchased in order to guarantee 99% success. Luckily, it is far easier to guarantee 90% or 80% success. Let's explore this in more detail, in the next box.



The graph at the left shows our function from the previous example:

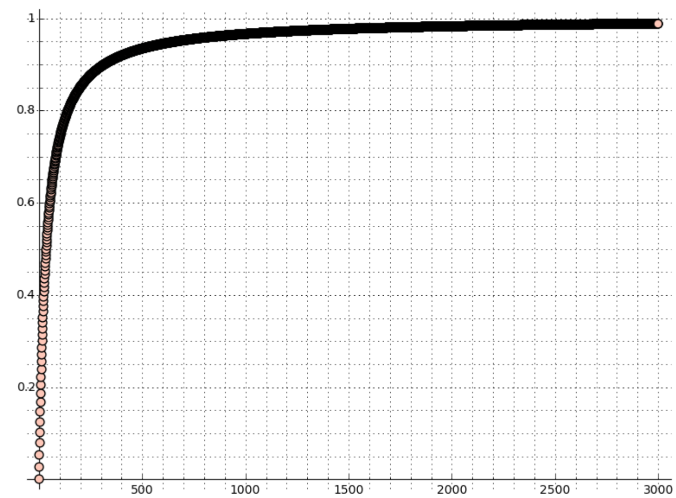
$$f(x) = \frac{x}{35+x}$$

As you can see, the function grows rapidly for the first bunch of tickets, but it eventually levels out. Since only integer values of x make sense, the function is plotted as a collection of dots representing

$$x \in \{0, 1, 2, 3, \ldots\}$$

instead of a smooth curve, representing all real values of x for some interval. In this case, our plot represents all integers between 0 and 200, inclusive.

It can be informative to see where the function crosses the $y = 0.7$ and $y = 0.8$ lines, and see that they appear to be at approximately $x = 81$ or $x = 140$, as you calculated in the previous box. Moreover, you can see that the function becomes very flat after about 100 tickets. Let's explore that more in the next box.



As you saw in the previous box, the graph of the function levels out after about 100 or 200 tickets. That's why we could reach 80% probability with only 140 tickets, but required 315 tickets to reach 90% probability, and a whopping 3465 tickets to reach 99% probability.

On the left, I've redrawn the plot to include all integers between 0 and 3000, inclusive. As you can see, the function is clearly leveling out. The probability will approach 1, getting closer and closer, but it will never reach 1. This behavior is an example of an *asymptote*.

One neat way to see that the probability will never reach one is to compute what happens if one million tickets are bought. In that case, we have

$$f(1,000,000) = \frac{1,000,000}{1,000,000 + 35} = \frac{1,000,000}{1,000,035} = 0.999965001\cdots$$

which is really close to 1, but not equal to 1.

Play
Around
With it...

7-5-35

Depending on how you count, there are 238 countries in the world. The reason that the number isn't so clear has to do with places that have an intermediate-degree of self-government, like Hong Kong, Puerto Rico, or Svalbard. In any case, the CIA World Factbook estimated (for July of 2015) that the population of the world was 7,256,490,011. The top ten countries in population were as follows.

1:	China	1,367,485,388	6:	Pakistan	199,085,847
2:	India	1,251,695,584	7:	Nigeria	181,562,056
3:	USA	321,368,864	8:	Bangladesh	168,957,745
4:	Indonesia	255,993,674	9:	Russia	142,423,773
5:	Brazil	204,259,812	10:	Japan	126,919,659

If you picked a person at random from the world's population, then what is the probability that they come from one of these ten countries? [Answer: $58.1514\cdots\%$.]

For the sake of completeness, it might be good to mention that the sample space (the interesting set) was the set of people on the planet. The subset we want are those people who happen to be in the population of one of those ten listed countries.

For the previous box, please note that you definitely should not respond

$$\frac{4,219,752,402}{7,256,490,011}$$

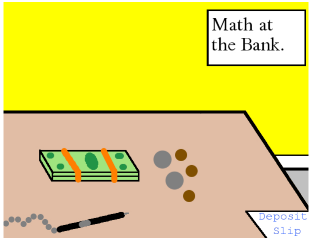
as this is not a human-readable number. It is hard to imagine that fraction in a concrete or tangible way, except perhaps we can see that it is slightly more than 1/2.

Play
Around
With it...

7-5-36

Looking at the previous checkerboard box and its data set, if you picked a random person in the world, what is the probability that they are...

- ...from China? [Answer: $18.8449\cdots\%$.]
- ...from the USA? [Answer: $4.42870\cdots\%$.]
- ...from Japan? [Answer: $1.74905\cdots\%$.]



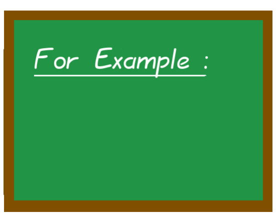
This would be a good moment to mention that when you talk about probabilities in business or industry, usually you report them to the nearest basis point. A basis point is 1% of 1%. Therefore, we would normally write 18.84% or 18.85%, 4.42% or 4.43%, and 1.74% or 1.75%, for the three answers in the previous box.

I reported six digits there, because we are using six significant figures throughout this textbook.

When working with probability problems related to surveys or experiments, there is one simple event (outcome) in the sample space for each choice on the survey, or outcome of the experiment. However, the survey's choices (or the experimental outcomes) must be constructed so that the simple events remain mutually exclusive and collectively exhaustive.

If these criteria of mutual exclusivity and collective exhaustion are met, then probability of an outcome is simply the ratio of people who chose that survey item, to the total number of people who answered the survey. In plainer words, the probability that a random member of the population sampled chose outcome A is equal to the percentage of the sample population that chose outcome A.

The compound events are treated identically as before—you add up the probabilities of the simple events (the outcomes) in the compound event.



7-5-37

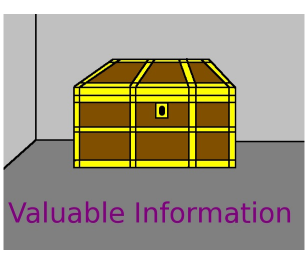
Suppose there is a survey with 487 respondents, to see if they read the local newspaper. The choices were "always read it," "sometimes read it," or "never read it." There were 68 people who said "always," 206 who said "sometimes," and 213 who said "never." What are the probabilities, as a percentage, that a random member of the set of respondents would give each of those answers, when asked whether they read the local newspaper?

The sample space is

$$\mathcal{S} = \{Always, Sometimes, Never\}$$

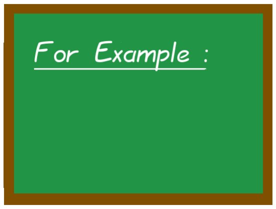
First, we can calculate the probabilities of the simple events (the outcomes).

- $Pr\{Always\} = 68/487 = 0.139630\cdots$
- $Pr\{Sometimes\} = 206/487 = 0.422997\cdots$
- $Pr\{Never\} = 213/487 = 0.437371\cdots$



This is the first time we've used the notation $Pr\{E\}$, which indicates the probability of an event E. It can be used for simple or compound events. This notation is extremely common.

Moreover, our list of three outcomes in the previous box, "always," "sometimes," and "never" form a sample space, by being mutually exclusive and collectively exhaustive. When we assign probabilities to those events, we have a probability distribution. Recall that a probability distribution is a list of outcomes in a sample space along with their probabilities.



7-5-38

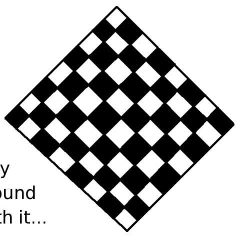
Now, using the data of the previous example, we can calculate some compound events.

- What is the probability, given the data in the previous problem, of someone reading the newspaper "at least sometimes"?

$$\frac{68}{487} + \frac{206}{487} = \frac{274}{487} = 0.562628\cdots$$

- What is the probability, given the data in the previous problem, of someone reading the newspaper "at most sometimes"?

$$\frac{206}{487} + \frac{213}{487} = \frac{419}{487} = 0.860369\cdots$$



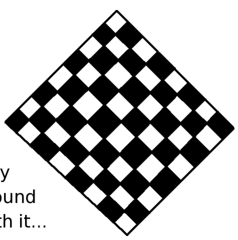
Play Around With it...

7-5-39

An ice-cream survey is being performed in a school cafeteria to determine what kind of flavors will sell the most. The choices are

Chocolate	213 votes	Strawberry	51 votes
Vanilla	88 votes	Mint Chocolate Chip	148 votes
Pistachio	23 votes	Salt-Water Taffy	2 votes
Banana Nut	18 votes		

- What is the probability that some random person from the survey likes strawberry? [Answer: $51/543 = 0.0939226\cdots$.]
- What is the probability a random person from the survey likes an ice-cream that contains chocolate in its name? [Answer: $361/543 = 0.664825\cdots$.]



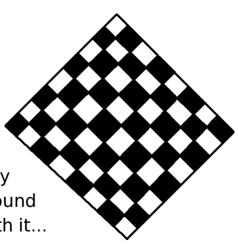
Play Around With it...

7-5-40

Let's continue with further questions from the previous box. Perhaps both the pistachio and the mint chocolate chip contain green food coloring, and a parent is concerned that a certain type of green food coloring was recalled by the manufacturer last year. No other flavors contain green food coloring, as it turns out.

What is the probability that a random student in the survey has chosen one of these two tainted ice creams as their choice?

[Answer: $171/543 = 0.314917\cdots$.]



Play Around With it...

7-5-41

Still continuing with the ice cream survey of the previous two boxes, let's consider the following questions.

- Both pistachio and banana nut contain nuts, but none of the other flavors do. What is the probability that an ice-cream flavor containing nuts is chosen by a random student in the survey? [Answer: $41/543 = 0.0755064\cdots$.]
- Similar to the previous bullet, what is the probability that an ice-cream without nuts is chosen? [Answer: $502/543 = 0.924493\cdots$.]

There are two different ways to arrive at the answer $502/543 = 0.924493\cdots$, in the previous box. On the one hand, we could say that the non-nut flavors are Chocolate, Strawberry, Vanilla, Mint Chocolate Chip, and Salt-Water Taffy. We can then add these simple events to get a complex event.

$$\frac{213}{543} + \frac{51}{543} + \frac{88}{543} + \frac{148}{543} + \frac{2}{543} = \frac{502}{543} = 92.4493\%$$

On the other hand, we know that percentages add to 100% when considering the parts of a whole, and therefore if 7.55064% of surveyed students choose something with nuts, then surely

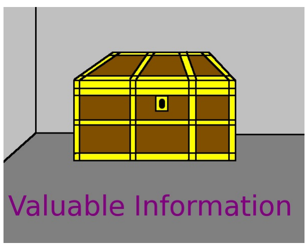
$$100\% - 7.55064\% \approx 92.4493\%$$

choose something without nuts.

Let's look again at the work of the previous box.

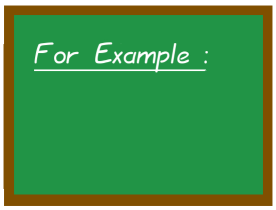
As you can see, if we had already known the value 7.55064%, then the second option is a much faster way of arriving at 92.4493%. We had to do fewer mathematical operations to reach that point.

This technique is extremely common in probability. If we work with percentages, we will subtract a probability from 100%. If we work with decimals or fractions, we will subtract a probability from 1. In fact, this technique is so common, that it has a name. We'll present that in the next box.



If p is the probability of an event happening, then $1 - p$ is the probability of an event not happening.

This is called *the complement principle* of probability.



7-5-42

Returning to the ice cream survey (from Page 946), what is the probability of a random student in the survey *not* choosing strawberry?

The data says that 51 students out of 543 students have chosen strawberry. Therefore, the probability of a student not choosing strawberry is computed by

$$1 - \frac{51}{543} = \frac{492}{543} = 0.906077\cdots$$

Isn't that much easier than calculating the following?

$$\frac{213}{543} + \frac{88}{543} + \frac{23}{543} + \frac{18}{543} + \frac{148}{543} + \frac{2}{543} = \frac{492}{543}$$

That is why the use of the complement principle is so popular. It's just a huge timesaver.

For Example :

7-5-43

I'd like to pose one last follow up question about the ice-cream survey of the last few boxes. If someone asked you to take the data, and form a probability distribution, you have to know what that means. It means you're going to list all the outcomes in the sample space along with their assigned probabilities. In this case, it comes out as

Chocolate	$0.392265\cdots$	Strawberry	$0.0939226\cdots$
Vanilla	$0.162062\cdots$	Mint Chocolate Chip	$0.272559\cdots$
Pistachio	$0.0423572\cdots$	Salt-Water Taffy	$0.00368324\cdots$
Banana Nut	$0.0331491\cdots$		

We don't really need six-significant figures here, but that's the standard used in this textbook. By the way, these numbers were obtained by taking the original entries from Page 946 and dividing them by the total number of survey respondents, 543.

For Example :

7-5-44

Let's look back at the checkerboard problem about the world's population and the ten most populous countries (see Page 944). Let's say that you are helping a younger sibling with their homework, and that they are writing a report about the world's population. Your younger sibling wants to know what percentage of people *do not* live in one of the top ten most populous countries. In other words, what percentage of the world's population live in the other $238 - 10 = 228$ countries? Must you now add up 228 numbers? That would be very tedious, because most of the populations would be in the millions!

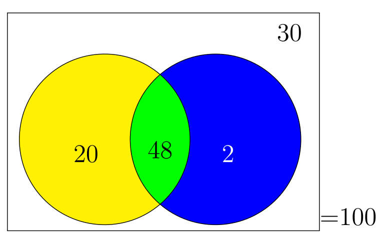
No, of course, we do not have to add up all those 228 population estimates. We can use the complement principle of probability, and simply calculate

$$100\% - 58.1514\% = 41.8486\%$$

For Example :

7-5-45

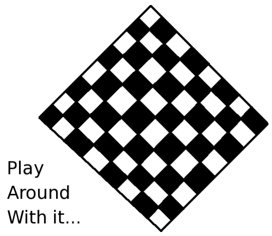
On Page 835, we studied a problem about a survey of graduating seniors and their employment. It was summarized by the following Venn Diagram. The left circle represent students who've had internships, and the right circle represent students with a job lined up for after graduation.



Based on the above, we can ask questions like "What is the probability that a student who has done an internship has a job lined up for after graduation?" and "What is the probability that a student who has not done an internship has a job lined up for after graduation?"

There are 68 students with internships, and 48 of them have a job lined up after graduation. Therefore, the probability that a student who had an internship has a job lined up is $48/68 = 70.5882\cdots\%$.

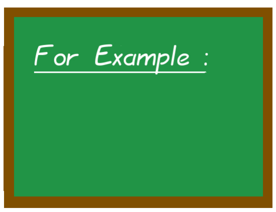
We will leave the others for you to find, in the next box.



7-5-46

Continuing with the previous box, what is the probability that. . .

- . . . a student who did not have an internship has a job lined up?
  [Answer: $2/32 = 6.25\%$.]
- . . . a student who did not have an internship has no job lined up?
  [Answer: $30/32 = 93.75\%$.]
- . . . a student who did have an internship has no job lined up?
  [Answer: $20/68 = 29.4117\cdots\%$.]



7-5-47

According to "Table 163: Osteopathic Physicians" of *The Statistical Abstract of the United States*, 131st Edition (2012–2013), in the year 2001 there were 46,962 osteopaths in the United States. Furthermore, they have the following age structure:

- 4838 were age 65 and over.
- 9544 were age 55 and over.
- 22,298 were age 45 and over.
- 37,096 were age 35 and over.

What is desired is a probability distribution of the ages of osteopaths. The outcomes are "65 and over," "age 55–64," "age 45–54," "age 35–44," and "below age 35." We certainly can't just divide the above numbers by 46,962, because that won't add up to 100%. Moreover, an osteopath who is age 66 will be counted in all four of those categories, violating mutual exclusivity.

Instead, we need only perform some subtractions. We'll do those in the next box.

Continuing with the previous box, first we are going to find out how many osteopaths are in each age bracket.

- There are $46,962 - 37,096 = 9866$ osteopaths who are under age 35.
- There are $37,096 - 22,298 = 14,798$ osteopaths who are 35–44.
- There are $22,298 - 9544 = 12,754$ osteopaths who are 45–54.
- There are $9544 - 4838 = 4706$ osteopaths who are 55–64.
- We already know that there are 4838 osteopaths who are 65 or older.

Then it is an easy matter to covert these to percentages, using division.

- Under age 35 $= 9866/46,962 = 21.0084\cdots\%$.
- Age 35–44 $= 14,798/46,962 = 31.5105\cdots\%$.
- Age 45–54 $= 12,754/46,962 = 27.1581\cdots\%$.
- Age 55–64 $= 4706/46,962 = 10.0208\cdots\%$.
- Age 65 and over $= 4838/46,962 = 10.3019\cdots\%$.

Those five bullets above assign probabilities to five events. Those five outcomes are mutually exclusive and collectively exhaustive, and thus form a sample space. The set of outcomes, taken with their assigned probabilities, form a probability distribution.

Now we can check the work of the previous two boxes with some sums.

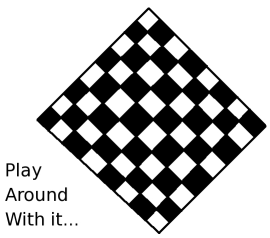
- $4838 + 4706 + 12,754 + 14,798 + 9866 = 46,962$ osteopaths are in the data set. ✓
- $4838 + 4706 + 12,754 + 14,798 = 37,096$ osteopaths are age 35 and over. ✓
- $4838 + 4706 + 12,754 = 22,298$ osteopaths are age 45 and over. ✓
- $4838 + 4706 = 9544$ osteopaths are age 55 and over. ✓

It is probably a good idea to check that the percentages get very close to 100%.

$$21.0084\% + 31.5105\% + 27.1581\% + 10.0208\% + 10.3019\% = 99.9997\% \checkmark$$

and surely it is clear that the last bit there is due to rounding error.

It looks like we got this one correct!

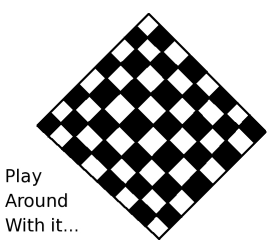


\# 7-5-48

Also according to "Table 163: Osteopathic Physicians" of *The Statistical Abstract of the United States*, 131st Edition (2012–2013), in the year 2010 there were 70,068 osteopaths in the United States. That's an impressive increase in only nine years, wouldn't you say? Furthermore, they have the following age structure:

- 6528 were age 65 and over.
- 17,723 were age 55 and over.
- 33,673 were age 45 and over.
- 53,791 were age 35 and over.

Figure out how many osteopaths are in each age band, and then construct a probability distribution for the age band of a randomly selected osteopath. The answers will be given on Page 965.



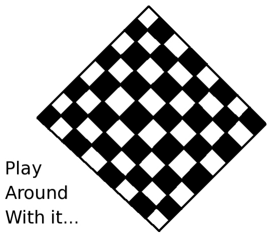
\# 7-5-49

The Pew Research Center did a study, published in the summer of 2015, on how Americans get their news. A portion of the research was studying how Americans use Twitter for news. A sample of 176 Twitter users were voluntarily monitored. For each user, four randomly selected weeks between August 2014 and February 2015 were analyzed. The tweets were read, categorized, and tweets that were not about one's friends and family, but about other events and issues, were tabulated. Here is a table that summarizes their data.

Never	28%	At least 50 times	18%
At least once	72%	At least 100 times	12%
At least 10 times	39%	At least 200 times	5%
At least 25 times	26%		

- Hint: It is probably easier to do this problem by first constructing a probability distribution.
- What percentage of people sent at least one tabulated tweet?
- What percentage of people sent between 10 and 24 tabulated tweets?
- What percentage of people sent between 50 and 199 tabulated tweets?

The answers will be given on Page 966. You can read an article about this data, by Michael Barthel and Elisa Shearer, called "How do Americans use Twitter for News?", posted to the blog "`KnightBlog`" on August 19th, 2015.



7-5-50

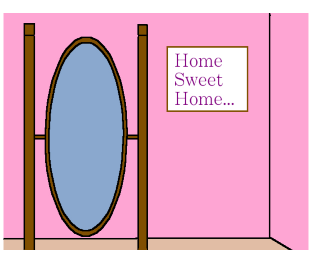
There was a Gallup Poll, conducted by phone during November 4–8, 2015, which asked the 824 people called the following question: "What do you think is the most important problem facing this country today?" Here is a summary of the responses.

- The economy in general: 17%
- Dissatisfaction with Government, Congress, Politicians; Poor leadership, Corruption, Abuse of power: 15%
- Immigration/Illegal Aliens: 9%
- Unemployment/Jobs: 7%
- Poor healthcare/hospitals; High cost of healthcare: 6%
- Federal Budget Deficit/Federal Debt: 5%
- Ethics/moral/religious/family decline; Dishonesty: 5%
- Gap between Rich and Poor: 4%
- There were 34 other categories that each received 3% or less, getting 29% of the population.
- "No opinion" was the response of 3% of the population.

Imagine that Prof. Witherspoon is writing a newsletter that deals with macroeconomics. He believes it will be of interest to those who responded about "the economy in general," "unemployment/jobs," and "federal budget deficit/federal debt," and roughly half of those who said "gap between rich and poor."

If the sample population of the survey is representative of the American population, then what is the probability that a randomly selected American will be interested in Prof. Witherspoon's newsletter, given his belief about who is interested? [Answer: 31%.]

Of course, simply because Prof. Witherspoon believes that the interest will work out along those lines, doesn't mean that it actually will.

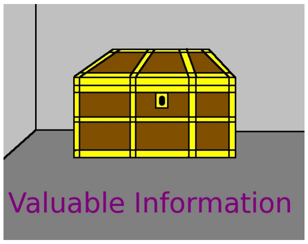


A Pause for Reflection...
Do you see any flaws in Prof. Witherspoon's reasoning? I see two big flaws, but perhaps you might see other flaws.

The first flaw is that we did not investigate the 34 other categories that each got less than 4% of the responses. Maybe some of those categories were very related to Prof. Witherspoon's bulletin. You can see the entire data set by clicking on this link: `http://www.gallup.com/file/poll/187667/MIP_151211.pdf`

The second flaw is that the Gallup poll only asked about "the most important problem." Perhaps many of the people who listed a non-economic reason as their #1 problem might list a macroeconomic topic as their #2. Just as an example, imagine a medical doctor who thought that the #1 problem was "Poor healthcare/hospitals; High cost of healthcare" but that the #2 problem was "unemployment/jobs," and the #3 problem was "federal budget deficit/federal debt." Perhaps that medical doctor might want to read Prof. Witherspoon's newsletter.

I've seen student research in the social sciences take this approach from time to time. It is better to ask something like "Do you think the issue of xyz is a major problem in the country right now?" Then your answers are simply "yes," "no," or perhaps "don't know/no opinion." It is much easier to calculate probabilities with that sort of questioning.

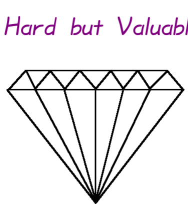


As promised, I now present you with the top five probability mistakes that I have found, looking at the exams and projects of students.

- Adding probabilities of events when that is not possible. The probabilities of outcomes (simple events) can be added. The probabilities of compound events cannot be added, in general. Recall, a compound event is an event that consists of more than one outcome, whereas a simple event has only one outcome in it. This pitfall is so dangerous that the entire following module is dedicated to discussing this matter.
- Making the "equally likely assumption" when it does not apply.
- Using a sample population that is not representative of the population that matters. This is not so much from exams, but from particularly badly designed student projects.
- Attempting to use a set that is not mutually exclusive, or that is not collectively exhaustive, as if it were a sample space.
- The fifth error is hard to explain, but it deals with assuming events will be independent when they are not. This will be explained in the module "Independence, Repetition, and Bernoulli's Formula."

I hope at this point that you've understood most of the module. If you did not get the 16 questions about the vaccine survey correct, then please retry those at this time. You'll find them on Page 924. As you do them, try to keep in mind the questions:

- What is the interesting set? (This will be your denominator.)
- What subset am I looking for? (This will be your numerator.)



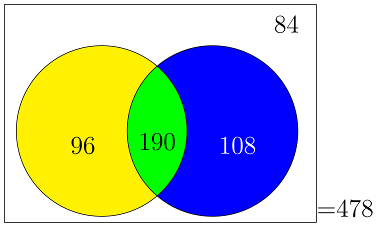
At this point, we have finished the mainstream topics of this module. What follows is a collection of advanced thoughts.

- A detailed look at the right and wrong uses of some hypothetical data about MBA admissions.
- A discussion of sample bias, including a very famous story from World War II. (In some ways, these are the least important boxes in the module, because we cannot shoot up a bunch of aircraft and have you analyze their hulls during a mathematics examination. Yet, in some ways, these are the most important boxes in the module, because the mistakes that get made in probability in the business world are often conceptual, not computational.)
- An exploration of some consequences of the fact that the sum of the probabilities of the outcomes is always 100% or 1 in a probability distribution.
- Further information about the Bernoulli-de Moivre-Laplace formula.

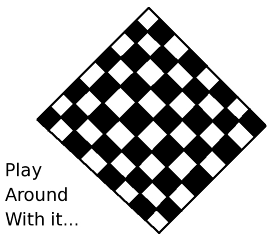
Some instructors will want to cover this material, and some will not. I hope that you will keep reading, either way.

We will now commence a long discussion on what I hope you will find is an interesting problem.

On Page 835, we studied a problem about the students enrolled in an international commerce MBA program, summarized by the following Venn Diagram.



All of the students can speak two languages, but those in the left circle can speak three or more languages. Also, those in the right circle have done a study abroad. To ensure that you understand the data in this box before we study this problem further, I'm going to ask you a few questions in the next box.



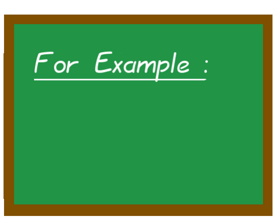
7-5-51

Based on the data in the previous box, tell me. . .

- As a percentage, what is the probability that a random international commerce MBA student neither speaks three or more languages nor has done a study abroad? [Answer: $17.5732\cdots\%$.]
- As a percentage, what is the probability that a random international commerce MBA student has done a study abroad? [Answer: $62.3430\cdots\%$.]
- As a percentage, what is the probability that a random international commerce MBA student speaks three languages? [Answer: $59.8326\cdots\%$.]

It is important to reflect on the right and wrong uses of the data in the previous box. If the MBA program is a two-year program, and if you're a corporate recruiter trying to hire people for internships/co-ops during that middle summer, then this survey is relevant to you. It will tell you what percentages of the students there have the language skills and the international experience that your business needs.

However, if you want to use the survey to decide questions about admissions, then this is very unwise. For example, let's suppose that you are an undergraduate senior, and you find the survey data of the previous box on the internet. You might be eager to use the data to answer questions about the probability of your own admission to that MBA program. Let's explore this further in the next box.



7-5-52

The fatal flaw is that the survey data, given in the international commerce MBA program, does not include any students who were rejected. Therefore, we cannot use the data to analyze the probability of admissions. To demonstrate this, we will consider two hypothetical cases, focusing on how having three or more languages affects the probability of admissions.

The consideration of these two cases constitutes a fairly advanced problem. Some students might find this very hard to understand. Don't be alarmed if you find the following difficult. Nonetheless, I think it is healthy to try to force yourself to follow the reasoning.

We already know that out of 478 accepted students, there were 286 with three or more languages, and 192 who spoke only two languages. In both cases, we will imagine that 1522 applications were rejected, giving a total of 2000 applicants.

Let's start by examining a hypothetical situation, relating to the problem of the previous box.

Case 1: Imagine that among the 1522 rejected students, there were 1400 who spoke only two languages. In this case, $1400 + 192 = 1592$ applicants spoke only two languages, and 192 got admitted, so the probability of a two-language applicant being admitted is a dismal $192/1592 = 12.0603\%$. On the other hand, $1522 - 1400 = 122$ rejected students have three or more languages, and 286 admitted students have three or more languages, for a total of $122 + 286 = 408$ applicants with three or more languages. The probability of a three-plus-language applicant being admitted is $286/408 = 70.0980\%$, which is much higher.

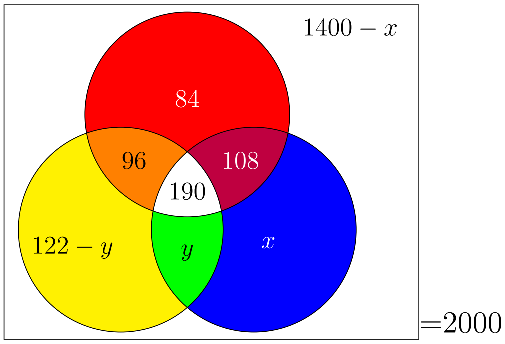
We'll examine Case 2 in the next box.

Now we'll continue with the previous example, examining a different hypothetical outcome.

Case 2: Contrastingly, imagine that among the 1522 rejected students, there were 600 who spoke only two languages. In this case, $600+192 = 792$ applicants spoke only two languages, and 192 got admitted, so the probability of a two-language applicant being admitted comes to $192/792 = 24.2424\%$. Meanwhile, $1522-600 = 922$ rejected students have three or more languages, and 286 admitted students have three or more languages, for a total of $922 + 286 = 1208$ applicants with three or more languages. The probability of a three-plus-language applicant being admitted is $286/1208 = 23.6754\%$. As you can see, the probability of admission is almost the same in both cases, with a tiny disadvantage, surprisingly.

In the next two boxes, we will graphically display Case 1 and Case 2.

This is a graphical representation of Case 1 from the previous example. The top circle represents admitted students, the left circle represents those students who speak three or more languages, and the right circle represents students who have completed a study abroad.



As you can see, some data is missing, signified by the x and y in the Venn Diagram. Luckily, that will not impede our progress. The probability for admission of a student who speaks only two languages is given by

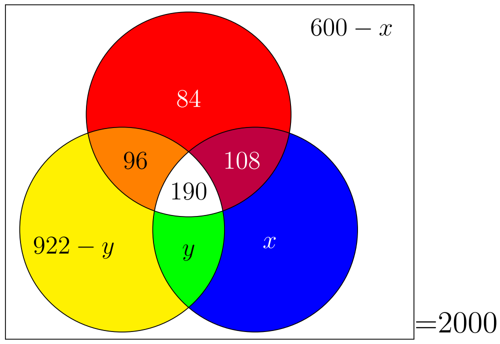
$$\frac{84 + 108}{1400 - x + x + 84 + 108} = \frac{192}{1592} = 12.0603\cdots\%$$

whereas the probability for admission of a student who speaks three or more languages is given by

$$\frac{96 + 190}{122 - y + y + 96 + 190} = \frac{286}{408} = 70.0980\cdots\%$$

In the next box, we will see a graphical representation of Case 2.

This is a graphical representation of Case 2 from the previous example. As before, the top circle represents admitted students, the left circle represents those students who speak three or more languages, and the right circle represents students who have completed a study abroad.



As you can see, only two spots changed. Again, some data is missing, signified by the x and y in the Venn Diagram. As before, that will not impede our progress. The probability for admission of a student who speaks only two languages is given by

$$\frac{84+108}{600-x+x+84+108} = \frac{192}{792} = 24.2424\cdots\%$$

whereas the probability for admission of a student who speaks three or more languages is given by

$$\frac{96+190}{922-y+y+96+190} = \frac{286}{1208} = 23.6754\cdots\%$$

What have we learned from considering two cases, in the previous several boxes, that are so different from each other? In Case 1, speaking three or more languages vastly increased a student's probability of admission. In Case 2, it had almost no effect, and was actually a slight disadvantage.

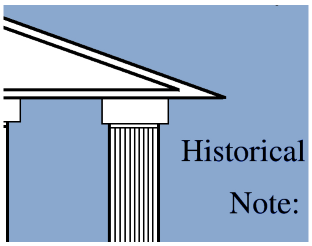
The bottom line is that we simply don't know the breakdown among the rejected applicants as far as how many spoke only two languages, and how many had three or more languages. Furthermore, we can see that two particular values of that unknown quantity (represented in Case 1 and Case 2) produce drastically different answers.

With that in mind, we must conclude that it is not possible *to say anything at all* about the impact of the number of languages spoken on the probability of admission.

When using survey data or experimental data, it is important that the sample population be a very good representation of the actual population that matters. Probability is about predicting the uncertain, and the prediction will be useless if this assumption is violated.

We looked at the MBA admissions question in great detail. I'm now going to show you three additional very brief examples, in a more casual way.

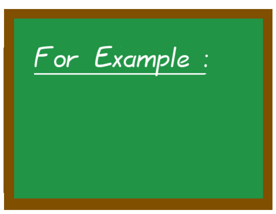
The name for this pitfall or error is *sample bias*. We witnessed sample bias in the international commerce MBA program because our data set only included accepted students, and did not contain any rejected students. In general, sample bias occurs when the sample population differs, in some important way, from the actual population that matters.



There is a famous story about the application of statistics to the aerial war in Europe during World War II. British and American bombers had been returning from their flights over Nazi Germany, with bullet holes from anti-aircraft guns. Of course, not all bombers made it back—some were destroyed over their targets or crashed into the North Sea. Particular regions of the bomber's surface were marked out, and it was noted that some regions had many bullet holes, some had a few bullet holes, and some had no bullet holes.

Archibald Wald (1902–1950), from the SRG (Statistical Research Group) was assigned to figure out where armor plating should be added to the bomber's surface. Of course, the armor plating was heavy, so one cannot armor the entire bomber—it would be too heavy to fly. Legend has it that those working with the bombers every day thought it pointless to bring on a statistician. After all, isn't the solution obvious? One should armor the parts of the plane with the most bullet holes. Paradoxically, Wald suggested the exact opposite!

Wald suggested that the parts of the aircraft where there were no holes to be found should be armored. Why is this case? We will explain in the next box.



7-5-53

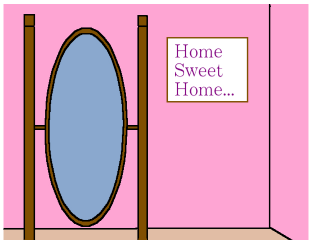
Regarding the question of the previous box, let us consider a simplified model where the surface of the bomber has been divided into regions A, B, C, D, E, F, G, and H, each having roughly equally exposed area. The bullets would land uniformly across those eight regions.

However, it might be that some particularly fragile components (e.g. gas tanks, fuel lines, and so forth) might be located behind those panels. If it comes to pass that, perhaps, no bomber can survive a hit to regions D or G, then what will the ground crews observe? Surely any bombers that are struck in regions D or G will not survive the return flight, and therefore they will not appear in the data set. Accordingly, every bomber that lands and is examined by the ground crew after its mission will have precisely zero bullet holes in region D and precisely zero bullet holes in region G. Every bomber in the data set will have 0 holes in both regions. It is for this reason that Wald suggested armoring the regions where no bullet holes were found.

In other words, the sample population of the survey was the set of bombers that survived the return trip. The ground crews in Britain cannot possibly include the locations of the bullet holes on the surface of the bombers that exploded over Nazi Germany or that crashed into the North Sea. The population that matters was the set of bombers that were destroyed, not the set that made it back.

Whenever the sample population is significantly different from (or entirely excludes) the population that matters, extreme caution must be taken when interpreting the results.

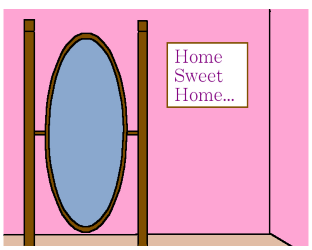
The previous two boxes represent the story as it is frequently told in books about statistics and operations research. However, in reality, Archibald Wald did a great deal of quantitative analysis to complement the above qualitative analysis. A concise and very readable 9-page article has been written, summarizing Wald's work. The title is "Abraham Wald's Work on Aircraft Survivability" by Marc Mangel and Francisco Samaniego, published by the *Journal of the American Statistical Association* in 1984. The article is freely available on the internet.



A Pause for Reflection...
Surveys about sexual habits, preferences, activities, and experience level are notorious for sample bias. Imagine a group of graduate students from psychology or sociology armed with clipboards and pencils, standing near a busy building on a college campus. They can stop students randomly to ask them if they'd be willing to take a 10-minute survey about their sexual habits. Surely, some students would agree, but also, some students would not agree, and a few might turn away in disgust. One can imagine that a prudish person who is not adventurous is rather unlikely to take the survey, whereas a person who is very sexually active and adventurous might be perfectly willing to fill out the survey.

Clearly, the data set would be drawn entirely from people who had filled out the survey. The students who do not fill out the survey cannot possibly appear in the data set. Therefore, the prudish students are under-represented and the adventurous students are over-represented. This particular situation is exceptionally bad, because both the under-representation of the prudish and the over-representation of the adventurous will increase measures of average experience level and average frequency of sexual activity.

The human desire to impress others might also lead one to exaggerate certain "prestigious" experiences and downplay other "socially unacceptable" experiences.



A Pause for Reflection...
Another challenging area would be dietary and nutrition studies. Perhaps a new drug or a new diet plan is being tested to measure its effectiveness. Patients would be recruited into the study, possibly to be paid or possibly not, and they would be closely tracked over the period of the study. Let us imagine that the drug or diet plan is indeed effective, and gets approved, marketed, and mass-produced.

The typical consumer, living their life with their job and family obligations, is not closely monitored. They are not taking part in a study, and therefore, their compliance with the rules of the diet might be excellent or it might be imperfect. If it is a drug, they might forget to take it from time to time, or perhaps not.

What it comes down to is that the set of patients taking part in a nutrition study or a dietary study is not perfectly representative of the population in general. The monitoring during a study is likely to increase patient compliance to dietary rules or drug-taking schedules. Therefore it is very possible that the effectiveness of the diet plan or drug will be higher during the medical studies than in ordinary life.

Our discussion of sample bias—whether it comes from bombers in World War II, surveys about sexual habits, or studies about weight loss, is now concluded.

Over the next four boxes, we're going to explore a very cool property of probability distributions. The sum of the probabilities of the outcomes is always 100% or 1.

When we take the probabilities of the outcomes of a sample space by assumption (or the manager's experience), we require all the probabilities to add up to 1 (see Page 936). Therefore, if the compound event $\mathcal{A}$ consists of all the outcomes (the simple events), then the probability of $\mathcal{A}$ must be 1.

It turns out that this will be true for all probability distributions. First, let's consider the case of the "uniform distribution assumption," in the next box.

Let's consider a problem where the "uniform distribution assumption" applies.

If the sample space has n outcomes, then each event has a probability of $1/n$. If my compound event $\mathcal{A}$ has all of the outcomes (all the simple events), then it has n outcomes. Then the probability of $\mathcal{A}$ is

$$\underbrace{\frac{1}{n}+\frac{1}{n}+\frac{1}{n}+\cdots+\frac{1}{n}}_{n \text{ times}} = n\left(\frac{1}{n}\right) = \frac{n}{n} = 1$$

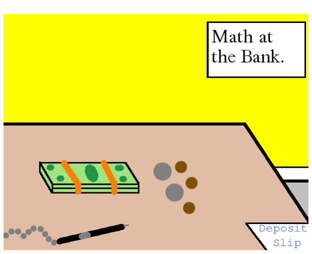
Similar in thought to the previous box, suppose we're working on a problem in the survey/experiment style. What is the probability of the compound event $\mathcal{A}$ that has all of the outcomes? Surely the denominator is the number of data points (or experimental trials), as always, in a survey problem. However, because the events are collectively exhaustive, every thing that ever happened in the data set can be found inside one of those events. Therefore, the numerator is also the number of data points.

Furthermore, there's no risk of counting the same data point twice, as a member of two distinct outcomes (simple events), because the outcomes are mutually exclusive.

Now we can be certain that the numerator and denominator in the probability of $\mathcal{A}$ will be equal. What is the value of a fraction that has numerator and denominator equal? It is equal to 1, of course!

Now we have shown that in all the "foundational" cases of building a probability distribution, the sum of the probabilities of the outcomes will be 100%, or 1.

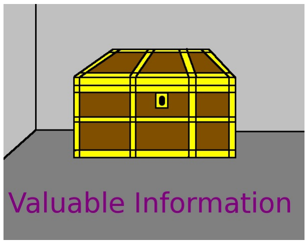
However, we did also state (on Page 936) that it is possible to make a new probability distribution from an old probability distribution, based on certain formulas. Even in this case, as it turns out, the sum of the probabilities of the outcomes will still be 1. Unfortunately, we are not in a position to prove that fact until we learn those formulas later in this chapter.



It was rather unfair of me to mention the Bernoulli-de Moivre-Laplace formula, and its important uses in the insurance industry, without actually showing you the formula. Actually, this formula comes up in many different financial situations.

In any case, some of my readers will have taken statistics already, and will know what a standard deviation is. Other readers will not have taken statistics yet, and do not know what a standard deviation is. With that in mind, I will state the Bernoulli-de Moivre-Laplace formula three times, in three different formats.

In the next box, is a statement of it that anyone can understand.



This is a version of the Bernoulli-de Moivre-Laplace formula which anyone can understand.

Given n independent trials where event E will occur with probability p, and thus not occur with probability $q = 1-p$, let x be the number of times that event E actually happens. For very large n, . . .

- . . . it is "probably" the case that

$$np - \sqrt{npq} < x < np + \sqrt{npq}$$

- . . . it is "very likely" that

$$np - 2\sqrt{npq} < x < np + 2\sqrt{npq}$$

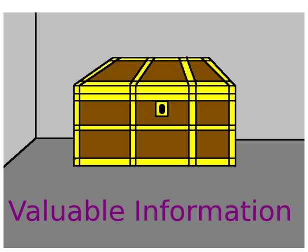
(Note: that's the inequality that I used to write the first few boxes of this module.)

- . . . it is "extremely likely" that

$$np - 3\sqrt{npq} < x < np + 3\sqrt{npq}$$

- However, this only works if $np > 5$ and $nq > 5$.

The only problem is that we haven't described what the word independent means, and we won't be able to do that until Page 1068.



This is a version of the Bernoulli-de Moivre-Laplace formula which will make more sense if you had exposure to probability, but not necessarily statistics.

Given n independent trials where event E will occur with probability p, and thus not occur with probability $q = 1-p$, let x be the number of times that event E actually happens. For very large n, . . .

- This inequality will hold with probability 68.27%, and be violated with probability 31.73%.

$$np - \sqrt{npq} < x < np + \sqrt{npq}$$

- This inequality will hold with probability 95.45%, and be violated with probability 4.55%.

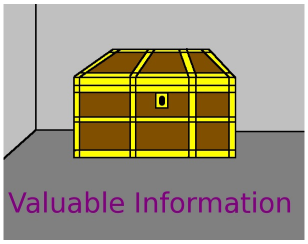
$$np - 2\sqrt{npq} < x < np + 2\sqrt{npq}$$

(This inequality, and the next one, were used to write the first few examples of this module.)

- This inequality will hold with probability 99.73%, and be violated with probability 0.27%.

$$np - 3\sqrt{npq} < x < np + 3\sqrt{npq}$$

- However, this only works if $np > 5$ and $nq > 5$.



By the way, almost no one ever calls it the Bernoulli-de Moivre-Laplace formula. Instead, it is almost always called the $\sqrt{npq}$ rule ("the square-root N P Q rule" in spoken mathematical slang). If you look at the previous box, you'll know why.

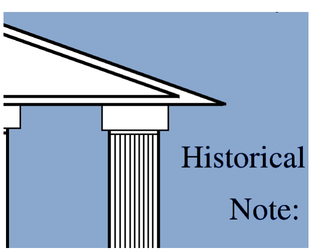
Computing the expected value, np, is easy. This can be remembered because "N P" is a classic slang abbreviation for "no problem." It is "no problem" to compute the expected value. Therefore, if you can remember the name of the $\sqrt{npq}$ rule, then all you have to remember are the coefficients—1, 2, and 3. It is really easy to memorize if you break it down this way.

Besides, it is much easier to say "the square-root N P Q rule" than "the Bernoulli-de Moivre-Laplace formula."

This is a version of the Bernoulli-de Moivre-Laplace formula which will make sense if you had statistics, because you will know what the normal distribution is. (It is sometimes called the Gaussian distribution.)

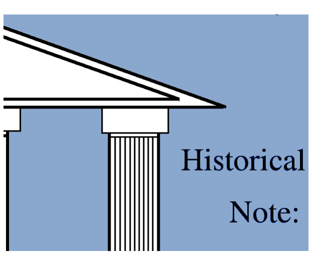
Given n independent trials where event E will occur with probability p, and thus not occur with probability $q = 1-p$, let x be the number of times that event E actually happens.

For very large n, the variable x is normally distributed, with mean equal to np and standard deviation equal to $\sqrt{npq}$. However, this only works if $np > 5$ and $nq > 5$.



Many textbooks omit Bernoulli and refer to this as the "de Moivre-Laplace Central Limit Theorem." However, that's not very fair, as Bernoulli was the first to develop this idea. In spoken English, it is common for both students and faculty to refer to this result as "the Central Limit Theorem" to avoid awkwardly mispronouncing the French, or as the "square root N P Q" rule.

If you'd like to pronounce the names correctly, then the pronunciation is kind of like "Bear-NOO-lee duh-MWAHV lah-PLAHS." Keep in mind that Bernoulli has also invented another important formula in probability, a formula that lies at the heart of reliability. We'll learn about that on Page 1116. Meanwhile, let's discuss Abraham de Moivre.



At times in history, we come across someone who made major contributions yet who lead a relatively miserable life. In this sense, Abraham de Moivre (1667–1754) is the Mozart of mathematics, but his life wasn't really as tragic as that of Wolfgang Amadeus Mozart (1756–1791). (You can watch the excellent movie *Amadeus* (1984) if you'd like a glimpse into Mozart's life.) What makes this story (told over the next few boxes) very exciting is that it touches upon the lives of some of the most famous people in all of history.

Historical Note:

The contributions of Abraham de Moivre are numerous. He's primarily known among mathematicians for a formula which allows for the roots and exponents of complex-imaginary numbers to be computed, such as

$$\sqrt[5]{\frac{\sqrt{2}}{2}+\frac{\sqrt{2}}{2}\sqrt{-1}} \qquad \text{or} \qquad \left(\frac{\sqrt{3}}{2}+\frac{1}{2}\sqrt{-1}\right)^7$$

but that has no place in a book like this one. He wrote a book about probability, called *The Doctrine of Chances*, and while this was not the first (or the second) book on probability, it was the first to be widely read by gamblers, so it must have been rather readable. He would have been comfortable with Chapter 4 of our textbook, because he also wrote a treatise about annuities—including annuities that take into account the estimating of the customer's date of death via probability theory. His work included the famous Fibonacci numbers, multinomials, and he got most of the way toward Sterling's formula for estimating factorials.

Professionally, Abraham de Moivre had major contributions. Yet, he was not able to earn a proper living, for reasons that we'll now explore in the next few boxes.

Historical Note:

Despite Abraham de Moivre's mathematical accomplishments, listed in the previous box, his actual life was extremely disrupted. It turns out that he was a Protestant, and in particular, a Huguenot.

By now, you've surely noticed a bunch of French names in the history of mathematics, but also in other scholarly subjects like science and philosophy. That's because France, in stark contrast to the rest of Europe, had religious freedom for a critical window of time. People from all over Europe would immigrate to France, making Paris an intellectual melting pot. This began in 1598, with *The Edict of Nantes*, signed by Henry IV (1553–1610) who was Protestant, and who at age 18 barely escaped the St. Bartholomew's Day massacre (1572), but who became Roman Catholic in order to accept the throne of France. Please understand that religious differences frequently lead to the death penalty, often accompanied by torture, throughout medieval and renaissance Europe, so this is a far from minor point. Actually, Henry IV himself was assassinated in 1610 by François Ravaillac (1578–1610). That story is extremely interesting as well, but there is no room to explore it here.

In any case, these 87 years of religious freedom made Paris a center of free thinking and scholarship, at a time when that freedom was otherwise unavailable. While it lasted for about three generations, this atmosphere of freedom was ruined by *The Edict of Fontainebleau* in 1685, signed by Louis XIV (1638–1715), which completely revoked religious freedom in France. Just as a warning—this is not the Louis that got his head cut off in the guillotine. You have two more Louis's to go, because it was Louis XVI (1754–1793) who lost his head along with his queen, Marie Antoinette (1755–1793).

The abrupt termination of religious freedom and the consequent trials and persecutions helped inspire both the American Revolution and the French Revolution, making this story a crucial turning point in the history of both Europe and the USA. In fact, New Rochelle, NY, which is a few miles from Fordham University where I once taught, was founded in 1688 by Huguenots fleeing *The Edict of Fontainebleau*. This is as much an American story as a European story. We learn that many of our early immigrants to the USA came here "because of religious freedom," four sterile and often-repeated words which do not seem to capture the full threat of torture and death.

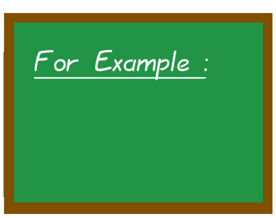
With this context, we return to Abraham de Moivre in the next box.

Keeping in mind the context of the previous box, Abraham de Moivre was only 18 years old when *The Edict of Fontainebleau* was signed. He was forced to renounce his faith and attend Roman Catholic services, while his father was imprisoned. When he was 20 years old, he fled to London with his brother and mother, and joined the Huguenot community-in-exile there. This was a high risk maneuver, because Huguenots were forbidden to emigrate, and those who were caught were sentenced to be oarsmen, rowing on royal galleys.

He was poor for his whole life, despite his mathematical accomplishments and the fact that he was the son of surgeon. He never received a university teaching position, and had to tutor students one-on-one in their homes. This took a lot of time. The (unverifiable) story is that he did not have time to read Newton's calculus book *Principia Mathematica* so he had to rip the pages out, putting them in his pocket and reading them as he walked from one tutoring job to the next. He earned supplementary money by playing and teaching chess in coffee houses.

Abraham de Moivre was friends with Edmond Halley (1656–1742) for whom the comet was named, Brook Taylor (1685–1731) for whom Taylor polynomials in calculus are named, and James Stirling (1692–1770), who completed Abraham de Moivre's work on estimating factorials numerically. If you've taken calculus, then you've probably heard of the furious debate over who invented it: Isaac Newton (1642–1726) or Gottfried Leibnitz (1646–1716). Abraham de Moivre was appointed to the committee for deciding the Newton versus Leibnitz dispute, next to many of these famous mathematicians. Yet, he never held a university position.

Having learned about the problems of the past, we now return to the problems of the present day in the next box. However, if you'd like to learn more, see *Abraham De Moivre: Setting the Stage for Classical Probability and Its Applications* by David R. Bellhouse, published by CRC Press in 2011.



\# 7-5-54

Suppose the probability of a car being stolen in your town, for a typical year, is 1%. Further suppose that 20,000 cars are in your town. A big insurance company is thinking of opening a branch in your town. What inequality does the Bernoulli-de Moivre-Laplace formula provide, to tell us the "very likely" values for the number of cars that will be stolen in a typical year? Hint: use the version of the inequality that has a "2" in front of the square root signs.

First we recall the formula,

$$np - 2\sqrt{npq} < x < np + 2\sqrt{npq}$$

and then we must figure out what the letters are. There are $n = 20,000$ cars, which might or might not get stolen. We have $p = 0.01$, the probability of car being stolen, and $q = 0.99$, the probability of a car not being stolen. We plug those in and get

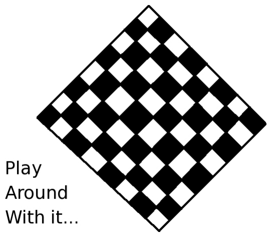
$$(20,000)(0.01) - 2\sqrt{(20,000)(0.01)(0.99)} < x < (20,000)(0.01) + 2\sqrt{(20,000)(0.01)(0.99)}$$

which simplifies to

$$200 - 2\sqrt{198} < x < 200 + 2\sqrt{198}$$

and further simplifies to

$$171.857\cdots < x < 228.142\cdots$$



7-5-55

Suppose in a nearby city, 1% is also the probability that a car will be stolen during a given year, and 5% is the probability that a car will be vandalized. However, the city has 250,000 cars.

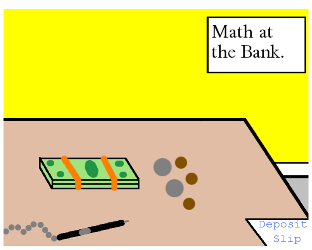
- What inequality does the Bernoulli-de Moivre-Laplace formula provide, to tell us the "very likely" values for the number of cars that will be vandalized in a typical year? Hint: use the version of the inequality that has a "2" in front of the square root signs.

  [Answer: $12,282.0\cdots < x < 12,717.9\cdots$.]

- What inequality does the Bernoulli-de Moivre-Laplace formula provide, to tell us the "very likely" values for the number of cars that will be stolen in a typical year? Hint: use the version of the inequality that has a "2" in front of the square root signs.

  [Answer: $2400.50\cdots < x < 2599.49\cdots$.]

It is worth noting that both this box and the previous box refer to rather high crime rates, as one might find in a dangerous city.

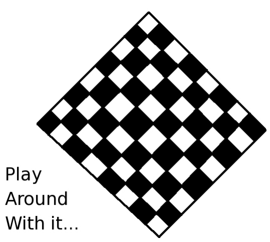


It is a bit funny to have the decimals go out to six significant figures in the above examples, because that represents less than one car. However, I've done that because we're using six significant figures as our standard in this textbook.

In industry, the standard practice is to round "outward" in this case. That is to say, you round to make the interval of values just slightly larger. The lower number is rounded down, and the upper number is rounded up. I did this in the first few examples of this module.

For example, in the previous box, we would say that between 12,282 and 12,718 cars will be vandalized. Meanwhile, between 2400 and 2600 cars will be stolen.

Last but not least, it is worth noting that a car can be both vandalized and stolen in the same year. We do not have enough data to be able to answer questions about the probability of being "either vandalized or stolen" nor questions about the probability of being "both vandalized and stolen."



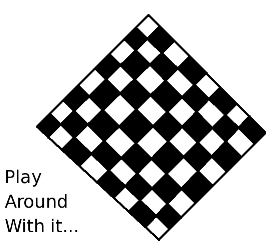
7-5-56

Let's repeat the previous checkerboard box, but with the "extremely likely" values.

Hint: use the version of the inequality that has a "3" in front of the square root signs.

- With the standard of "extremely likely," how many cars do we expect to be stolen? [Answer: 2350 to 2650.]

- With the standard of "extremely likely," how many cars do we expect to be vandalized? [Answer: 12,173 to 12,827.]



7-5-57

One last problem will wrap up the module. Suppose I flip a fair coin a trillion times. (Of course, that's not possible within a human lifetime, but this is a thought experiment.) I'm going to count how many heads I get, and call it x. Into what interval is x very likely to fall, in the sense of the Bernoulli-de Moivre-Laplace formula?

[Answer: $499,999,000,000 < x < 500,001,000,000$.]

Note that the values in the previous box reinforce what we expect from the law of large numbers. We expect

$$np = (1,000,000,000,000)(1/2) = 500,000,000,000.$$

and the Bernoulli-de Moivre-Laplace formula tell us that we will be very likely to get an answer within 1,000,000 of that.

By the way, there's a really cool way to check your work with the Bernoulli-de Moivre-Laplace formula. If you average the lower bound and the upper bound, before rounding them, then you should get the expected value. Even if you average them after rounding, it should be very close. This works regardless if you use the "very likely" or the "extremely likely" interval.

Let's check the work for the problem about stolen and vandalized cars. The expected values are

$$np = 250,000(0.01) = 2500 \qquad \text{and} \qquad np = 250,000(0.05) = 12,500$$

The averaging, even after the rounding, gives us the following numbers:

$$\begin{aligned} \frac{2350 + 2650}{2} &= 2500 \\ \frac{2400 + 2600}{2} &= 2500 \\ \frac{12,173 + 12,827}{2} &= 12,500 \\ \frac{12,282 + 12,718}{2} &= 12,500 \end{aligned}$$

It looks like I got it right!

The module is now complete. Here are a few answers to questions posed earlier in the reading.

Here is the answer to the question about the instructor who wants to model, on any given day, how many of his students will be absent from class. The question was on Page 931.

[Answer: He forgot to consider the possibility that no one shows up at all. He's missing 0, so the set he wrote is not collectively exhaustive, and thus is not a sample space.]

Even though the outcome "zero students present and all students absent" is really unlikely, the totals will be slightly off with this sample space. If we perform some mathematical analysis with a sample space that is not collectively exhaustive, then we are very likely to make an error.

Actually, I did once "fill in" for a colleague who had very liberal attendance policies while teaching Math-123: *Finite Mathematics*, at the University of Wisconsin—Stout. He was traveling to a conference and I covered his classes for him. His class had approximately 30 students enrolled, but that day only 1 student attended, while all of the other 29 students were absent. On the one hand, this is horrifying, but on the other hand, it was a Friday in April.

Naturally, I still delivered the topic that I had prepared, for the benefit of the one student who showed up.

This is the answer to the question about earthquake insurance on Page 931.

[Answer: Your colleague has overlooked the possibility that a house might not have been damaged at all by an earthquake (0% damage). Moreover, if the damage were between 0.01% and 4.99% of the value of the home, then the correct "category" after rounding would be 0%. This set is not collectively exhaustive, and is therefore not a sample space.]

This is the answer to the question about tracking annual stock performance on Page 931.

[Answer: It is possible for a software-company's stock to rise by 20% or more. Moreover, it is easily possible for a software company's stock to fall by 20% or more—especially if they go bankrupt. Therefore, we need to have something similar to

$$\mathcal{S} = \{\text{“}-20\% \text{ or worse,”} -15\%, -10\%, -5\%, 0\%, 5\%, 10\%, 15\%, \text{“}+20\% \text{ or better”}\}$$

or perhaps instead

$$\mathcal{S} = \{\text{“}-25\% \text{ or worse,”} -20\%, -15\%, -10\%, -5\%, 0\%, 5\%, 10\%, 15\%, 20\%, \text{“}+25\% \text{ or better”}\}$$

to correctly model this problem.]

These are answers for the question about middle initials and airline reservations from Page 931.

For the question of "collectively exhaustive" we can firmly say that the answer is no, because some people do not have a middle name at all. However, not all students might be aware of that, so it isn't a very good question.

For the question of "mutually exclusive" we can firmly say that the answer is no, because some people have multiple middle names. Consider the famous military historian, C. R. M. F. Cruttwell (1887–1941). His full name was "Charles Robert Mowbray Fraser Cruttwell." He is essentially famous for three things. First, having written *A History of the Great War, (*1914–1918*)*, which is particularly interesting, since Cruttwell fought in that war and was wounded. Second, for having offended the novelist Evelyn Waugh (1903–1966), who then retaliated by naming many unpleasant or silly characters "Cruttwell" in his numerous stories and novels. Third, he is famous for having a really long name.

For the question of the "equally likely assumption" the answer is clearly no. Surely more people have a middle name that begins with a common letter like "J" or "T" than a middle name which begins with "X" or "Q."

Here are the answers for converting the osteopath data from Page 950 into a probability distribution. First, the actual counts of the number of osteopaths by age band.

- There are 16,277 osteopaths that are under age 35.
- There are 20,118 osteopaths that are 35–44.
- There are 15,950 osteopaths that are 45–54.
- There are 11,195 osteopaths that are 55–64.
- There are 6528 osteopaths that are 65 or older.

We will continue in the next box.

Continuing with the previous box (about the osteopaths), as percentages we have the following:

- Under age 35: $23.2302\cdots\%$.
- Age 35–44: $28.7121\cdots\%$.
- Age 45–54: $22.7636\cdots\%$.
- Age 55–64: $15.9773\cdots\%$.
- Age 65 or over: $9.31666\cdots\%$.

I'd like to note something. It is particularly odd to report the last few decimal places. The sixth significant figure represents much less than 1 person in this case. However, I have kept six significant figures in this problem because that is our standard throughout the textbook.

These are the answers about Twitter usage, taken from the question on Page 950.

- What percentage of people sent at least one tabulated tweet? [Answer: 72%.]
- What percentage of people sent between 10 and 24 tabulated tweets? [Answer: 13%.]
- What percentage of people sent between 50 and 199 tabulated tweets? [Answer: 13%.]

Here is the probability distribution. I think it is a lot easier "to see" how to get the answers to the above questions after constructing the probability distribution. However, this is not (strictly speaking) necessary.

0 tabulated tweets	28%	25–49 tabulated tweets	8%
1–9 tabulated tweets	33%	50–99 tabulated tweets	6%
10–24 tabulated tweets	13%	100–199 tabulated tweets	7%
200 or more tabulated tweets	5%		