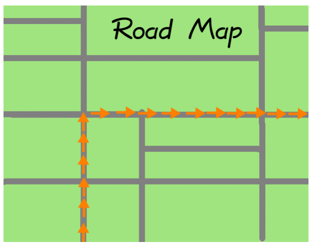
# Module 3.1: A Formal Introduction to Probability Theory



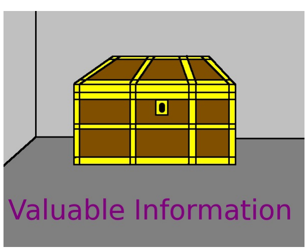

In this module, we're going to explore some, but not all, of the formal underpinnings of probability. Getting the basics down correctly is important for launching your exploration of the theory of probability. The flavor of this module might appear theoretical, and in many ways it is. However, much of the rest of the theory of probability has a problem-solving feel to it. Moreover, probability is extremely practical—as we will see throughout this chapter.

I know that almost all of my readers will have taken calculus before this course. Probability is not unlike calculus, where you had to learn the formal definition of a derivative and take derivatives "the long way" at first, but never did that again—except on exams. Instead, for the rest of calculus, you solved interesting problems using shortcut formulas for derivatives and integrals. If you've not had calculus so far, then do not worry—that's not a barrier to learning probability in discrete mathematics.

The crucial items we need to define are outcomes (also called simple events), compound events, and sample spaces. These, in turn, depend on the ideas called "mutually exclusive" and "collectively exhaustive." Before we turn to those formal details, we will try to define what a probability is, and in doing so, we will visit with the idea of the "Law of Large Numbers."

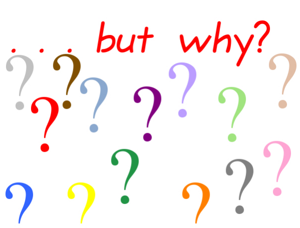

By the way, if the concepts discussed here seem strange, or arbitrary, then do not be alarmed. A lot of what happens in the formal definitions might appear to be unnecessary, awkward, or even incomprehensible. However, everything will be a lot clearer after several modules go by. It is by solving problems, not by analyzing definitions, that we will come to master the theory of probability.

On the other hand, without formal definitions, we would have no idea what we are doing and we would have no idea how to begin.




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.

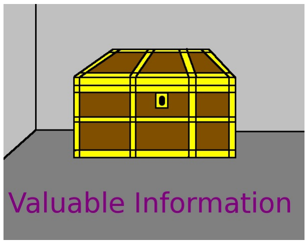
Furthermore, it must always be the case that $0 \leq p \leq 1$.




Many of us are taught in high school that a probability of $p = 0$ means "never" and a probability of $p = 1$ means always. This is actually wrong, but only in the most theoretical and unrealistic of questions.

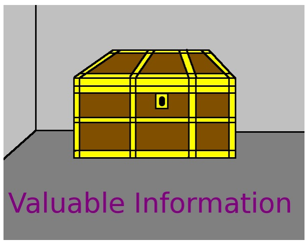
It turns out that there is a difference between "probability zero" and "absolutely impossible," as well as a difference between "probability one" and "absolutely certain." We will touch on this concept very late in this module, on Page 290.

However, those extremely theoretical circumstances will never come up in a real-life problem, and do not reflect any remotely practical situations. This is mentioned only for completeness.




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

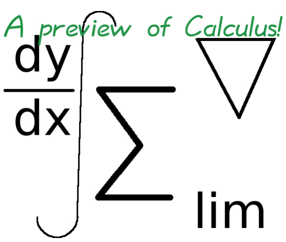
Yet, there is another important piece to the puzzle, which gives meaning to these numbers that we assign to events. That piece of the puzzle is the "Law of Large Numbers."




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.

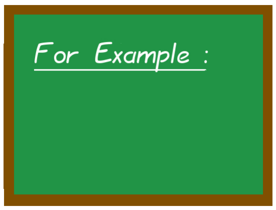

There are many variations of the "Law of Large Numbers." For example, advanced probability textbooks will distinguish between the "*Strong* Law of Large Numbers" and the "*Weak* Law of Large Numbers." There is also "*Borel*'s Law of Large Numbers," and the "Uniform Law of Large Numbers."

That's of no interest to us in discrete mathematics. It comes down to the following: There are some random situations which are so unpredictable that their variance and standard deviation are infinity. (If you don't know what that means, then don't worry—just skip to the next box.) In those situations, the easy-to-prove version of the Law of Large Numbers does not apply. That easy-to-prove version is called the "*Strong* Law of Large Numbers." However, a harder-to-prove version does work for many of those situations, and that's called the "*Weak* Law of Large Numbers."

To make matters more confusing, the conclusions of the strong law and the weak law are not exactly the same. The distinction is very subtle and there's no way for me to explain it at this time. To do that, we would need some very advanced concepts—which you might or might not learn in later courses.

Nonetheless, permit me to guarantee for you that this distinction will never matter when solving a real world problem, because the Law of Large Numbers is absolutely true. It is merely a matter of whether or not an easy-to-read proof, or a hard-to-read proof, is necessary for proving the law.

I think it would be very helpful for us to step away from theory now, and move toward some practical questions. I'd like to give a more "everyday" description of the "Law of Large Numbers," through four examples, in the next two boxes.

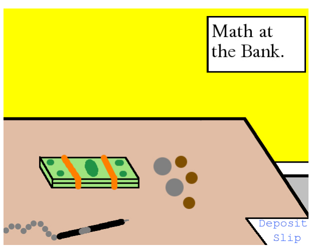



# 3-1-1

Imagine that a somewhat large insurance company has insured 10 million homes. 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. How many homes should they plan on rebuilding?

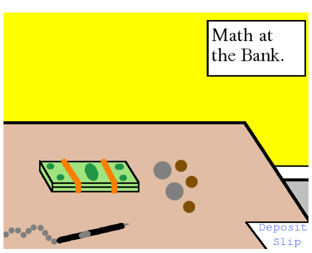
The "Law of Large Numbers" means that they can expect $(10,000,000)(1/500) = 20,000$ of their customers' homes to be destroyed each year.

As it turns out, using a technique which we will learn later (called the Bernoulli-de Moivre-Laplace inequalities) 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. The economy 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 rather unlikely situation that the number of destroyed homes 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.

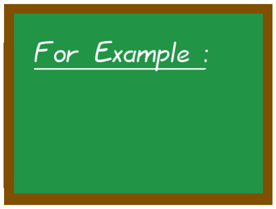



The theory of probability enables insurance companies to perform computations like the ones in the previous two boxes.

What probability *does not do*, is given a directory of 10,000,000 customer files, allow you to predict *which* 20,000 or so customers will lose their houses. If you can understand this distinction, then you are already on your way to understanding probability as a subject.

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. In any case, this is why the insurance industry employs enormous numbers of statisticians.

To see a county-by-county assessment, across the USA, of the risk levels for a home being destroyed, you might want to look at the article "Which natural disaster will likely destroy your home?" by Les Christie, published on `@CNNMoney` on June 19th, 2014. It has a cool interactive map.
`http://money.cnn.com/2014/06/19/pf/insurance/natural-disaster-risk/index.html`

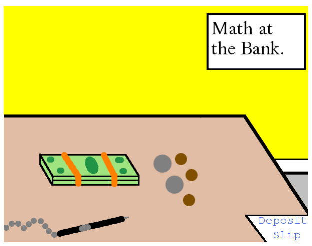



# 3-1-2

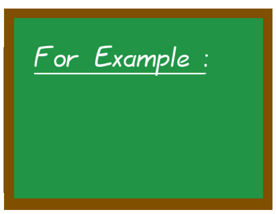
Imagine that a large restaurant in a major airport is having connection issues with their WiFi. It is often the case that there is a very large number of WiFi users in a small space in a busy modern airport. The manager is worried that customers who are having trouble connecting will simply move to a 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 inequalities (which we will learn on Page 348 of the module "The Square Root of NPQ Rule"), 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.

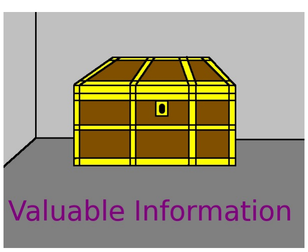



# 3-1-3

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 heads should he expect?

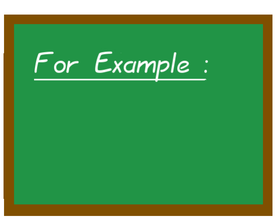
That person should expect $(10,000)(1/2) = 5000$ heads. However, it is unlikely to be exactly 5000. It could easily be 4951, 5027, or many possible numbers near 5000.

Using the Bernoulli-de Moivre-Laplace inequalities (which we will learn on Page 348 of the module "The Square Root of NPQ Rule"), 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. These 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, but I'm going to show you that integral on Page 357. 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," which is sometimes called "the empirical rule."




# 3-1-4

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 inequalities (which we will learn on Page 348 of the module "The Square Root of NPQ Rule"), we can say it is very likely (95.54%) to be between 11 and 29 destroyed satellites, and extremely likely (99.73%) to be between 6 and 34 destroyed satellites.


Did you notice how the interval of uncertainty was a lot smaller for the 10,000,000 homes, as a percentage, than for the other three examples?

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 $np$. To keep things simple, let's look at the "very likely (95.45%)" values.

| Example Topic | $n$ | width | (width/$n$)(100%) |
|---|---|---|---|
| Coin Flips | 10,000 | $5100 - 4900 = 200$ | 2.00000% |
| Satellites | 1000 | $29 - 11 = 18$ | 1.80000% |
| WiFi Customers | 8000 | $439 - 361 = 78$ | 0.975000% |
| Destroyed Houses | 10,000,000 | $20,283 - 19,717 = 566$ | 0.00566000% |

By the way, if you looked at the "extremely likely (99.73%)" values, you'd see the same phenomenon. The $n$ needs to be fairly large for the "Law of Large Numbers" to work. This point is important, so let's discuss it further.


Here's another way of explaining the point of the previous box.

- If you flipped a coin one billion times, then we expected 500,000,000 heads and it is very likely (95.45%) to be between 499,968,377 heads and 500,031,623 heads, and extremely likely (99.73%) to be between 499,952,565 heads and 500,047,435 heads. The 95.45% interval has 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 heads and it is very likely (95.45%) to be between 40 heads and 60 heads, and extremely likely (99.73%) to be between 35 heads and 65 heads. The 95.45% interval has a width of 20, which is 40% of 50.
- We'll have Sage compute these two examples, and other examples, in the next box.

The reason that the uncertainty is tiny for a billion, but the uncertainty is large for a hundred is because one billion is a large number, and one hundred is not a large number. The next box will make this much clearer by providing more examples.

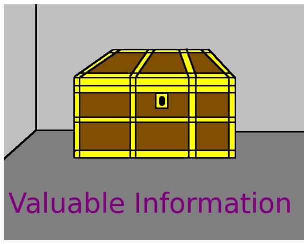
Using Sage, I can consult the Bernouli-DeMoivre-Laplace inequalities for several values of $n$. This time, I used the "extremely likely (99.73%)" range, so the following predictions will be true with probability 99.73%.

| | | | | | | |
|---|---|---|---|---|---|---|
| For n | = | 10 | it will be between | 0 | and | 10 |
| For n | = | 100 | it will be between | 35 | and | 65 |
| For n | = | 1000 | it will be between | 452 | and | 548 |
| For n | = | 10,000 | it will be between | 4,850 | and | 5150 |
| For n | = | 100,000 | it will be between | 49,525 | and | 50,475 |
| For n | = | 1,000,000 | it will be between | 498,500 | and | 501,500 |
| For n | = | 10,000,000 | it will be between | 4,995,256 | and | 5,004,744 |
| For n | = | 100,000,000 | it will be between | 49,985,000 | and | 50,015,000 |
| For n | = | 1,000,000,000 | it will be between | 499,952,565 | and | 500,047,435 |
| For n | = | 10,000,000,000 | it will be between | 4,999,850,000 | and | 5,000,150,000 |
| For n | = | 100,000,000,000 | it will be between | 49,999,525,658 | and | 50,000,474,342 |
| For n | = | 1,000,000,000,000 | it will be between | 499,998,500,000 | and | 500,001,500,000 |

As you can see, even for $n$ as small as ten million or one hundred million, the uncertainty is basically irrelevant. If you look at $n$ equal to ten billion, the uncertainty only shows up in the fifth decimal place! In stark contrast, if $n = 100$, there is a lot of uncertainty.

That's because numbers like ten million or ten billion are large numbers, whereas $n = 100$ is not a large number. This is why we call it the "Law of Large Numbers."

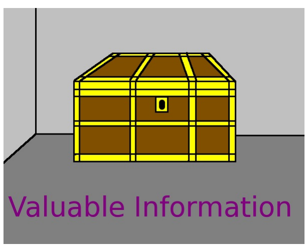
The Sage code for the above computations will be given on Page 358 of the module "The Square Root of NPQ Rule."




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

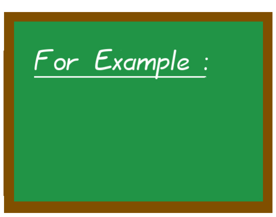
A *sample space* is a set of outcomes connected to a probability problem or experiment. The outcomes must be *mutually exclusive* as well as *collectively exhaustive*.

- 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 criteria guarantee that "exactly one outcome occurs" for any trial of the experiment.




Don't be alarmed if the previous definition is strange at first glance. Just take a set of outcomes that has been presented to you, 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.



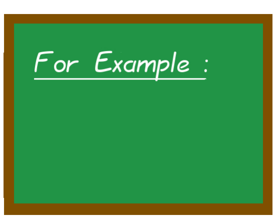

# 3-1-5

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 $\mathcal{S} = \{H, T\}$.



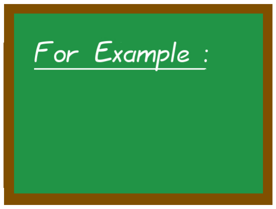

# 3-1-6

Imagine that 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.




# 3-1-7

Another valid sample space for the previous example 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.

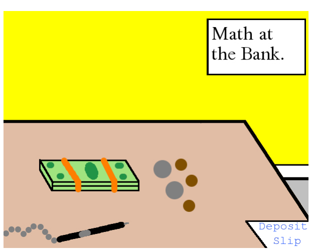
The choice of using $\mathcal{N}$ or $\mathcal{S}$ would really depend on what questions I was hoping to answer about the coin tosses.


In the previous box, you might be able to list the eight possible outcomes. However, it might be challenging for me to list all the outcomes if I were studying a sequence of 6 coin flips. (There are $2^6 = 64$ 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

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

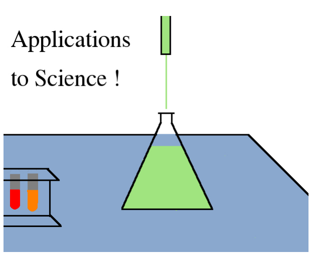
possibilities for a sequence of 20 coin flips. Surely, we do not want to list each of the 1,048,576 outcomes!




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.

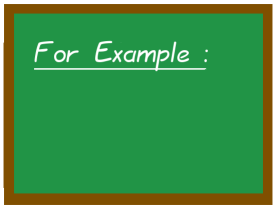
If the defect rate is too high, the company's reputation will be damaged and profits will be lost. However, it might be very expensive to get the defect rate extremely low. Probability theory allows management to navigate this balancing act.




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 a bit harder, so we'll examine problems like that later.

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.




# 3-1-8

In the USA, college admissions can be complicated. However, one vital 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.

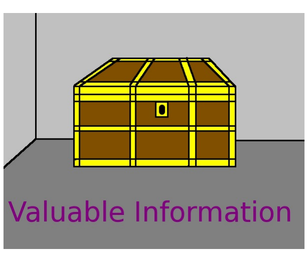
Perhaps a student will decline in order to attend a different university, or perhaps the student will decline to attend any university at all. The reason for the decline is irrelevant. Probability theory helps advise the admissions office on how many offers of admission they should make. For now, let us ask ourselves, "What is the sample space for this problem?"

For any particular student, 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. As you can see, this is similar to heads or tails on a coin. In stark contrast to that, if they offer admissions to 1500 students, then the sample space might be written as a set of sequences of 1500 $A$s and $D$s. As it turns out, there are $2^{1500} \approx 10^{451.544\cdots}$ possible sequences of 1500 $A$s and $D$s. You cannot possibly list all of them. The piece of paper would be the size of the solar system or larger.


To convince you that the issue in the previous box is a serious problem, consider a college with 1200 available beds in the dorms for new freshmen. They'd like to have a freshman class of 1000 students, and they know historically that they have had 2/3 of the admitted freshman accepting, and 1/3 declining. Therefore, they admit 1500 applicants.

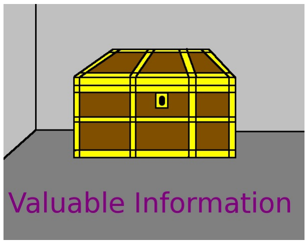
- The expected number of acceptances is $np = (1500)(2/3) = 1000$.
- It is fairly likely (68.27%) that the number will be between 981 and 1019.
- It is very likely (95.45%) that the number will be between 963 and 1037.
- It is extremely likely (99.73%) that the number will be between 945 and 1055.

These computations were done using the Bernoulli-de Moivre-Laplace inequalities, and you will learn how to do those calculations yourself on Page 348 of the module "The Square Root of NPQ Rule."




The previous few boxes gave us some examples of important situations modeled by weighted coin flips. Those examples are listed below.

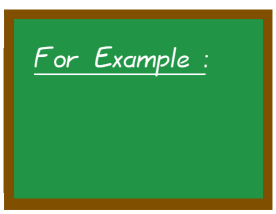
1. Manufacturing smartphones, which are either defective or functional.
2. Humanitarian aid workers, who will either get infected with Ebola or not.
3. College admissions, in trying to predict how many students will accept or decline.
4. Earlier, we analyzed customers at an airport restaurant, who either can or cannot connect to the WiFi.
5. Earlier, we also analyzed satellite launches, which are either successes or catastrophic failures.
6. We could also model the homeowner's insurance this way, as a home will either be destroyed, or not. However, intermediate levels of damage are fairly common, so that would not be a great model.




As it turns out, this technique of modeling binary situations (as repeated flips of a weighted coin) is so important, that it has a name. Each smartphone, aid worker, admitted student, restaurant customer, or satellite launch is one coin flip. The model is called a *Bernoulli Random Variable*.

In the module "The Binomial Distribution" we will learn some tricks that can solve problems of this type very quickly.


In the Summer of 2018, I hope to write a biography of Jacob Bernoulli (1655–1705) and insert it here.




# 3-1-9

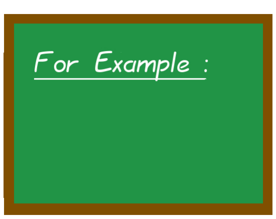
Let's return briefly to our earlier example of a smartphone factory. For simplicity, let's consider phones as either defective ("D") or functional ("F.") Once during each shift, the quality-control engineer is going to sample five phones coming off the production line. Naturally, each can be defective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



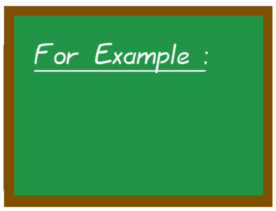

# 3-1-10

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.




# 3-1-11

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 treasurer—with the president drawn first, vice-president second, and treasurer 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.

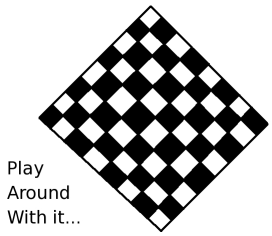

If we try to analyze the problem mentioned in the previous box, where the names of Andrea, Bill, Chuck, Doris, and Edgar are going to be drawn to select a president, a vice-president, and a secretary, then we get a mess. Do not attempt to write such a list down.

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 |

As you can see, listing things just is not satisfactory, even for small problems. It is too time-consuming, and for medium-sized problems it might take a very long time indeed. We need a better way, and that's one reason the subject of combinatorics was invented.

In the module "The Combination and Handshake Principles" we will learn a very nice tool that will resolve the situation in the previous few boxes, both quickly and easily.




# 3-1-12

Suppose 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. We are curious about who the next incoming caller will get?

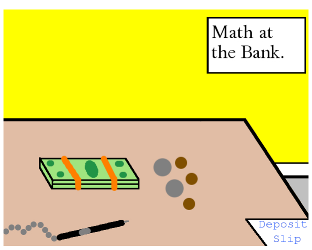
- What is the sample space?

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

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

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

- Is $\mathcal{S}$ mutually exclusive? [Answer: Yes, $\mathcal{S}$ is mutually exclusive, because it is not possible for a caller to be assigned multiple team members.]
- Is $\mathcal{S}$ collectively exhaustive? [Answer: Yes, $\mathcal{S}$ 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. Those sound unrelated to real life.

However, another example where we have the equally likely assumption is in a survey. Suppose you survey 500 people, and you ask them each "Do you approve of Senator John Q. Public?" or "Do you approve of Vaccinating Children?" Then you might want to ask, "What is the probability that a random survey respondent approves of Senator John Q. Public?" Of course, this is a proxy for "What is the probability that a random American approves of Senator John Q. Public?"


When 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

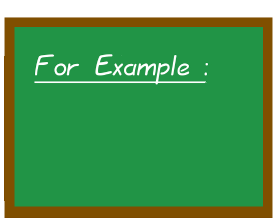
$$\mathcal{S} = \{\{HHH\}, \{HHT\}, \{HTH\}, \{HTT\}, \{THH\}, \{THT\}, \{TTH\}, \{TTT\}\}$$

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, on Page 288 of this module, we'll calculate that the probability of getting exactly 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, cards, roulette, drawing names from a hat, but also in surveys and opinion polls), it certainly is not true all the time.

I consider one of the six common 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 317 for the complete list of the six very common errors.




\# 3-1-13

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? While you're at it, do you think the "equally likely assumption" applies here?

Take a moment to think about it. The answer will be given on Page 291.

Play
Around
With it...

# 3-1-14

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 291. Please go have a look, because there is a neat story embedded in the answer.

Play
Around
With it...

# 3-1-15

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!)
- Also, do you think the "equally likely assumption" applies here?

The answers will be given on Page 291 of this module.

Play
Around
With it...

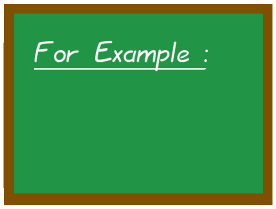
# 3-1-16

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 292.




# 3-1-17

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. What is the sample space?

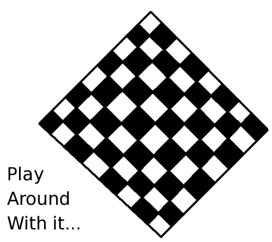
The sample space is

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

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

The previous box actually touches on three points from earlier modules.

- Recall that the symbol "{}" represents "the empty set." Here, that means someone who uses neither newspapers, TV, nor radio. (We talked about that on Page 35 of the module "Introduction to Sets.")
- Also recall, $\mathcal{S}$ is the power set of $\{T, R, N\}$, which is just a technical term for the set of all subsets of $\{T, R, N\}$. (See Page 159 of the module "Intermediate Set Theory and Irrationality.")
- Lastly recall, 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 51 in the module "Introduction to Sets.")

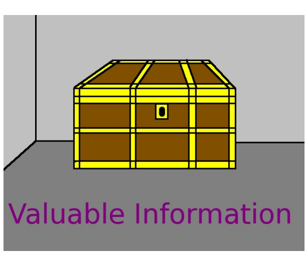



# 3-1-18

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

- How many outcomes will be in the sample space?
- What are they?
- Does the "equally likely assumption" apply?

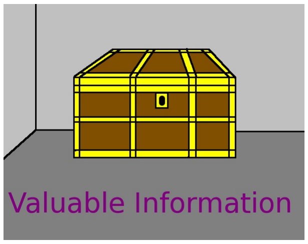
The answer will be given on Page 292 of this module.




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 or surveys.
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 in the next module, "Exploring Probability Through Problem Solving."

Our list will be continued in the next box.

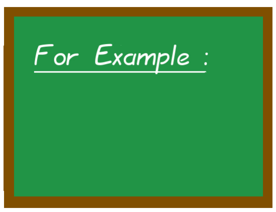



Continuing from the previous box,

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 also see a lot of that in the next module, "Exploring Probability Through Problem Solving," on Page 336.

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).

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 was used.



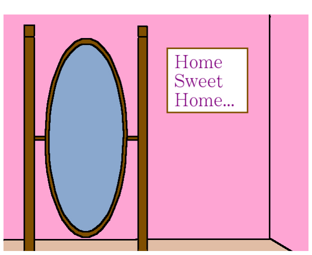

Looking at the previous box, an example for # 4 might be helpful. When modeling employee tardiness in a troublesome team, a manager might estimate that employees are punctual 85% of the time, late 10% of the time, and absent 5% of the time. If these numbers are given from the manager's experience and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

We'll reexamine this situation in the next module, "Exploring Probability Through Problem Solving," on Page 296.

# 3-1-19


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.

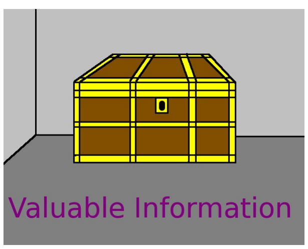



*A Pause for Reflection...*
We can summarize the last few boxes by saying that in probability, you never get something for nothing. You have to have a starting point, a foundation for your analysis.

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.

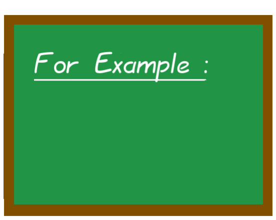

When a sample space has probabilities attached to each of the outcomes, it becomes a probability distribution. Now, we should describe where the probabilities actually come from, in each of our four cases.

- In the first case (the equally likely assumption) we will describe the event as some interesting set (possibly the entire sample space) and a very interesting subset. We will compute a division, with the size of the very interesting subset in the numerator, and the size of the interesting set in the denominator. This will be made clear by examples, throughout the next module, "Exploring Probability Through Problem Solving."
- In the second case, we have a survey and have collected data. The percentage of respondents who make a particular choice becomes the probability that a random survey respondent agrees with that choice. Then, we usually assume that the people surveyed are representatives of some larger population, which might be false, and conclude that this percentage also holds true for that larger population.
- In the third case, we can use formulas from probability—formulas that we will learn over many modules throughout this chapter and the two following it.
- In the fourth case, some experienced person is just stating estimated probabilities. Those often aren't very accurate.




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 $1/n$.



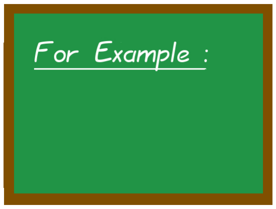

# 3-1-20

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 $1/2$.
- From the customer service team on Page 282, 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 $1/8$.
- If a "record locator" for a flight reservation is just six randomly generated letters of the English alphabet, then a sample space for the first letter of the record locator would be the set $\{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z\}$, and each outcome has probability $1/26$.




# 3-1-21

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.

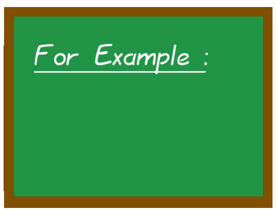
Since the customer service representative is assigned by a computer so that each representative is equally likely, we have

| 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.




# 3-1-22

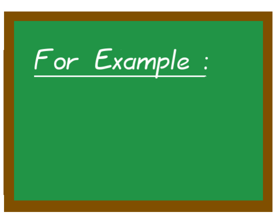
Instead of flipping one fair coin, as we did two boxes ago, we could flip three fair coins in succession. It might be interesting to see what the probability distribution looks like.

| 3-flip Sequence: | HHH | HHT | HTH | HTT | THH | THT | TTH | TTT |
|---|---|---|---|---|---|---|---|---|
| Probability: | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 |

Moreover, the probability of getting 0 heads is 1/8, from TTT; the probability of getting 1 head is 3/8, from HTT, THT, and TTH; the probability of getting 2 heads is 3/8, from HHT, HTH, and THH; the probability of getting 3 heads is 1/8, from HHH.

| How Many Heads?: | None | One | Two | Three |
|---|---|---|---|---|
| Probability: | 1/8 | 3/8 | 3/8 | 1/8 |

This provides us with a deep insight. We can see that for one single situation, you can have two sample spaces—one where the "equally likely assumption" applies, and one where it does not.

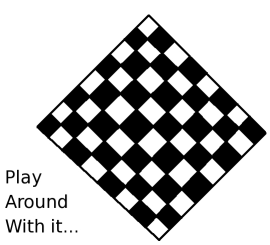



# 3-1-23

If someone asks you the following question: "If you know that a married couple has 3 children, but you don't know anything about the genders of the children, then what is the probability distribution for 0 girls, 1 girl, 2 girls, and 3 girls?"

Looking at the previous box, we can see that the probabilities are 1/8, 3/8, 3/8, and 1/8. After all, the gender of a child at birth is like the flip of a fair coin—it is a 50-50 chance of male or female.

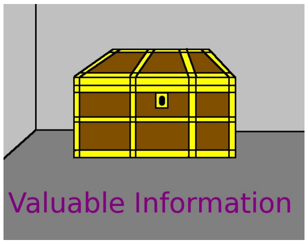
This is a very classic question in probability, and it has been a classic for many generations. I would not be surprised to see it as a job-interview question.




# 3-1-24

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.

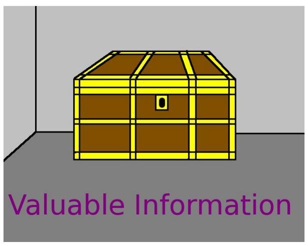
What is the probability that a random record locator will have an X as its first letter? [Answer: 1/26.]




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.

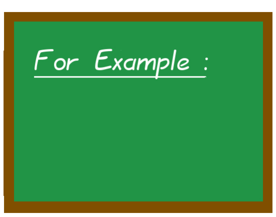



There are two technicalities that are important to understand about events.

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 36, Page 50, and Page 50. Feel free to review those definitions, if you feel that you need to.

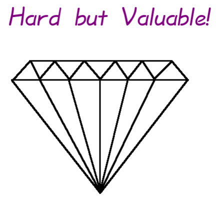



\# 3-1-25

Earlier, on Page 280, 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.




The bulk of the module is now complete. I have one last point to explore. Earlier, on Page 272 of this module, I claimed that there is a difference between "probability 0" and "absolutely never," and I also claimed that there is a difference between "probability 1" and "absolutely always."

This distinction will only matter if you take multiple additional courses in probability. However, it is mathematically interesting.


Suppose I am throwing darts at a square representing part of the coordinate plane, namely $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$. This is a perfect square of area 4.

There are all sorts of fun questions that I can ask. For example, if I throw darts at this dart board, what is the probability that I will strike a point where both coordinates are positive? Well, that's the area defined by $0 < x \leq 1$ and $0 < y \leq 1$, which is a perfect square of area 1. We would compute the probability by saying $1/4$, the ratio of the areas. Since there are four quadrants of equal size, this makes sense.

Now we'll see why "probability zero" and "never" are not exactly the same.


What is the probability that I would hit the $x$-axis? In other words, $x$ can be anything on the dartboard, but what is the probability that the point that is struck has $y = 0$? The $x$-axis has area equal to 0. Therefore, the probability is $0/4 = 0$. Of course, there are points on the $x$-axis, so I might hit it. But the probability of getting $y = 0$ on the nose, absolutely exactly, is 0—because the $x$-axis is a line and lines are infinitely thin. It is not impossible to hit the $x$-axis, but it is extraordinarily unlikely.

My engineering students will realize that there's a flaw in this conversation. The points on the coordinate plane are infinitesimally small. Therefore, in order for this discussion to make sense, the tip of the dart must also be infinitesimally small—the size of a single point. Due to the atomic theory of matter, the dart must be at least 1 atom across. Of course, it is much more likely to be significantly larger than that. It is a pure mathematician's game of "pretend" to say that the tip of the dart is the size of a single point on the coordinate plane.


With all of the contents of the last two boxes in mind, throwing a dart and hitting a non-zero $y$-coordinate will occur with probability $1 - 0 = 1$. This is fairly intuitive. It is also an application (though a very easy one) of the complement principle of probability. We'll see that on Page 303 of the module "Exploring Probability Through Problem Solving."

Because physical objects have some size that is measurable, and not zero, we can be very confident that we will not encounter this issue in the workplace.

On the plus side, we can use this idea of a square dartboard for other purposes. There should be an interact on the web page where you downloaded this text. It involves approximating $\pi/4$ by asking the probability of striking the unit circle when throwing a dart at this dartboard.

The app does this by using Sage. It picks a random point by generating random numbers. Then it uses the distance formula to compute the distance to the origin. If the distance is less than or equal to 1, then the point is inside the unit circle. By counting how many darts strike inside the unit circle, out of 10,000 darts, we get a crude approximation of $\pi/4$. That's because the unit circle has area equal to $\pi$ and the dartboard has area equal to 4.

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


This is the answer to the example about tracking annual stock performance on Page 283. Normally, examples have there answers in the box itself, but in this case, I really wanted you to think about the question, so I've put the answer here at the end.

It is certainly 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.

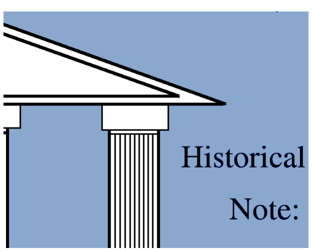
You might or might not have familiarity with the stock market, but small fluctuations happen almost constantly. Larger fluctuations are less common, but not very rare. In any case, the "equally likely assumption" does not apply.


These are answers for the question about middle initials and airline reservations from Page 284.

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.

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 clearly has several middle initials, and we'll talk about him in the next box.

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 a rare letter like "X" or "Q."




C. R. M. F. Cruttwell (1887–1941) is famous for at least three things. First, having written *A History of the Great War, (*1914–1918*)*, which is a particularly interesting book, since Cruttwell fought in that war and was severely wounded. It should be noted that Cruttwell was already teaching history at Oxford when he began military service in 1941. He later became the head of Hertford College, one of the constituent colleges of Oxford University.

Second, Cruttwell is famous 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.


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 284.

[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 almost certain to make an error.

We need to address the "equally likely assumption" next, but first, I have a funny story for you. It's in the next box.

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 somewhere around 29–35 students enrolled, in each of two sections. For the first section, 5 students showed up. For the second section, only 1 student attended, while all of the others 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.

Another faculty member once had a section where no one at all showed up, but that was for Math-010: *Beginning Algebra*, a class where the first quiz is a review of multiplying single-digit positive integers.


Continuing with the previous two boxes, while it is rare for the vast majority of a class to be absent, it is possible. Likewise, it is far more common for the majority to be present. This makes us absolutely certain that the "equally likely assumption" does not apply to this problem.


This is the answer to the question about earthquake insurance on Page 284.

[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.]

Moreover, after an earthquake, many houses might be destroyed, but most are not. I don't know what percentage of homes will fall into these percentage of damage intervals, but surely it won't be the case that each category has exactly equal probability to all the other categories. Whatever the percentages are, they will be ugly numbers with decimal points. Therefore, the "equally likely assumption" does not apply.


These are the answers to the question (from Page 285) about how people get their news: the internet, television, radio, or newspapers.

- How many outcomes will be in the sample space? [Answer: $2^4 = 16$.]
- What are they? [Answer: $\{\ \{\}; \{T\}; \{R\}; \{N\}; \{T,R\}; \{R,N\}; \{N,T\}; \{N,T,R\}; \{I\}; \{T,I\}; \{R,I\}; \{N,I\}; \{T,R,I\}; \{R,N,I\}; \{N,T,I\}; \{N,T,R,I\}\ \}$.]
- Does the "equally likely assumption" apply? [Answer: No, because the number of people who respond with $\{N,T,R,I\}$ is going to be much smaller than the number of people who respond with $\{I\}$.]