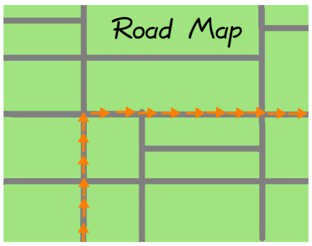
Module 3.5: The Square Root of NPQ Rule

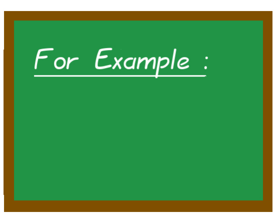


In the opening module of this chapter, I mentioned the Bernoulli-De Moivre-Laplace inequalities, more often called "the square root *npq* rule." This helped us to explain that the uncertainty in probabilistic events can be put between two bounds, forming an interval, which create promises that hold with various fairly high probabilities.

All of those vague words will now be made rigorous and mathematical. We will see that we can compute useful intervals. We will use these inequalities to solve problems in various types of insurance, environmental estimates, risk assessments, and reliability engineering. We can make predictions about the number of typos in a textbook or the percentage of students on a campus that identify as gay or lesbian. We can also describe the "Law of Large Numbers" more precisely using the inequalities. We will even model the profitability of a casino's Roulette table.

Then in the closing part of the module, we'll see why two nuisances—polling error, and the error that comes with estimating a probability in a scientific experiment—are fundamentally unavoidable. Not only that, but we will see that polling error and experimental error are mathematically the same thing. Last but not least, we will foreshadow the statistical concept of a confidence interval, which is the most important tool from statistics that scientists and engineers actually use.

We're going to start with a simple recap of the four opening examples from the module "A Formal Introduction to Probability Theory." The "square root of *npq* rule," more properly called the Bernoulli-de Moivre-Laplace inequalities, were used to compute important numbers in those examples.

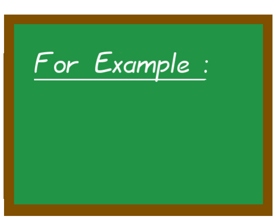


3-5-1

Here is the first example, from Page 274 of the module "A Formal Introduction to Probability Theory."

Because there is a roughly 1 in 500 chance that an ordinary American house will be destroyed in a given year, if an insurance company insurances 10,000,000 homes, then they can expect to rebuild $np = (10,000,000)(1/500) = 20,000$ homes.

It is unlikely to be exactly 20,000 homes, but it is very likely to be between 19,717 and 20,283 (probability 95.45%), and extremely likely to be between 19,576 and 20,424 (probability 99.73%).

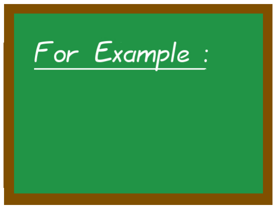


3-5-2

Here is the second example, from Page 275 of the module "A Formal Introduction to Probability Theory."

Suppose a restaurant manager discovers that 5% of the customers are unable to connect to the airport WiFi. Because they have 8000 customers per month, they can expect $np = (8000)(0.05) = 400$ customers per month will be unable to connect and get frustrated.

It is unlikely to be exactly 400 customers, but we can say it is very likely to be between 361 and 439 (probability 95.45%), and extremely likely to be between 341 and 459 (probability 99.73%).

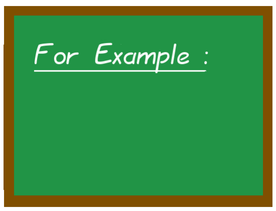


3-5-3

Here is the third example, from Page 275 of the module “A Formal Introduction to Probability Theory.”

Suppose that a person with an exceptionally large amount of patience and free time were to flip a fair coin 10,000 times. They can expect $np = (10,000)(1/2) = 5000$ tails.

It is unlikely to be exactly 5000 tails, but we can say it is very likely to be between 4900 and 5100 (probability 95.45%), and extremely likely to be between 4850 and 5150 (probability 99.73%).

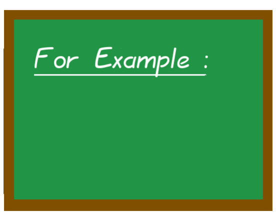


3-5-4

Here is the fourth example, from Page 275 of the module “A Formal Introduction to Probability Theory.”

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). If an ambitious satellite launching program plans to eventually have 1000 launches, then they can expect $np = (1000)(0.02) = 20$ catastrophic failures.

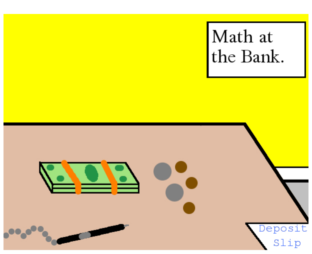
It is unlikely to be exactly 20 failures, but we can say it is very likely to be between 11 and 29 (probability 95.45%), and extremely likely to be between 6 and 34 (probability 99.73%). This is rather interesting, because if you tell a manager “between 6–34 catastrophic failures” the width of that window is huge. The 34 is 5× to 6× larger than the 6. A manager might not be able to comprehend that interval, but saying “between 11–29 catastrophic failures” might be more palatable.



3-5-5

A little bit later on Page 279 of the module “A Formal Introduction to Probability Theory,” we considered a small college. 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.



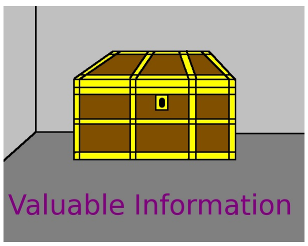
It was slightly unfair of me to mention the Bernoulli-de Moivre-Laplace inequalities at that time, and their important uses in the insurance industry, without actually showing you the formulas. Actually, these formulas come up in many different business and industrial situations, and in the social sciences—especially political science.

By the way, almost no one ever calls them the Bernoulli-de Moivre-Laplace inequalities. Instead, it is almost always called the “square root of *npq* rule.” If you look at the formulas in the next several boxes, then you’ll know why.

Besides, it is much easier to say the “square root of *npq* rule” than “the Bernoulli-de Moivre-Laplace inequalities.” Not only that, if you call it the “square root of *npq* rule,” then you are more than half way to remembering the structure of the formulas.

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, but have studied probability before. Still other readers have studied neither probability nor statistics.

With that in mind, I will state the idea four times, in four different formats. The next box contains a statement of the idea that anyone can understand.



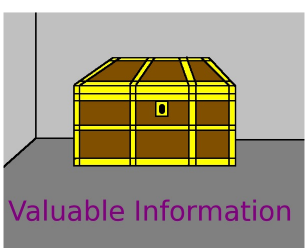
This is a version of the "square root of npq rule" which anyone in a university-level mathematics class should be able to understand.

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 "fairly likely" that
$$np - \sqrt{npq} < x < np + \sqrt{npq}$$
- ... it is "very likely" that
$$np - 2\sqrt{npq} < x < np + 2\sqrt{npq}$$
- ... it is "extremely likely" that
$$np - 3\sqrt{npq} < x < np + 3\sqrt{npq}$$
- However, this only works if $np > 5$ and $nq > 5$. Some books say $np > 10$ and $nq > 10$.

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 462.

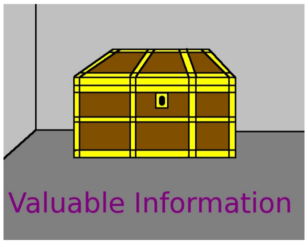


This is a version of the "square root of npq rule" which will make more sense if you had exposure to probability, but not necessarily statistics.

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.

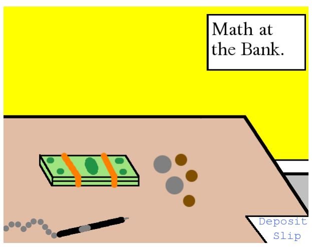
For very large n, ...

- 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 chapter.)
- 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 $nq > 5$. Some books say $np > 10$ and $nq > 10$.



Here's another way of putting the idea:

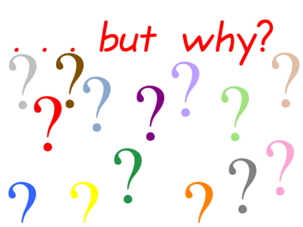
- With probability 15.86%, the observed reality will be below the "fairly likely" range; with probability 68.27%, the observed reality will be inside the "fairly likely" range; with probability 15.86%, the observed reality will be above the "fairly likely" range. (This is sometimes called the one-sigma range.)
- With probability 2.27%, the observed reality will be below the "very likely" range; with probability 95.45%, the observed reality will be inside the "very likely" range; with probability 2.27%, the observed reality will be above the "very likely" range. (This is sometimes called the two-sigma range.)
- With probability 0.13%, the observed reality will be below the "extremely likely" range; with probability 99.73%, the observed reality will be inside the "extremely likely" range; with probability 0.13%, the observed reality will be above the "extremely likely" range. (This is sometimes called the three-sigma range.)



Looking at the previous box, it seems reasonable that almost no one in business uses the "fairly likely" interval, except perhaps investment companies in explaining risk to customers. It is too weak of a guarantee. After all, 68.27% is just a bit larger than 2/3, and that is not an impressive guarantee.

It also seems reasonable that the "extremely likely" is a bit overboard. That's a solid guarantee, but it isn't solid enough if human life is at stake. (Since 0.26% is about 1 in 385—too high of a risk of death.) However, for ordinary situations when a business is trying to make a profit, the guarantees provided by the "very likely" interval seem entirely sufficient, and the "extremely likely" interval seems to be wasteful.

After all, there is only a 2.27% chance of being below the "very likely" range, and that's about 1 in 44. If you were told that a business plan has a 43 in 44 chance of being profitable (and if you actually believed that this probability is roughly correct), wouldn't you invest?



Some of my readers have taken a course in statistics. This is a version of the "square root of npq rule" which will make sense if you had statistics, because you will know what the normal distribution is. (It is sometimes called the Gaussian distribution.) If you have not yet had a course in statistics, then skip the rest of this box, jumping to the next box.

Remember, if you haven't had a course in statistics, then you should skip this box. 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.

In the limit as n grows to infinity, the variable x is normally distributed, with mean equal to np and standard deviation equal to $\sigma = \sqrt{npq}$. (Note, σ is the Greek letter "sigma.")

For very large n, this provides a good approximation only if $np > 5$ and $nq > 5$. Some books say $np > 10$ and $nq > 10$.

Looking back at the box before the previous box, you can see now where those numbers (68.27%, 95.45%, and 99.73%) come from! However, I don't want to frighten your classmates who have not yet had statistics, so I won't dwell on that.

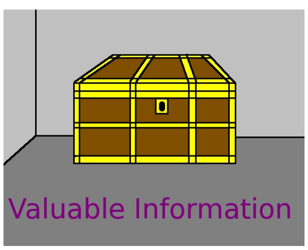
This is one of the most important of all the theorems in probability and statistics—and some would say the most important. It is called "the central limit theorem," and it has many uses, far beyond those that I can show you at this time. In some ways, the central limit theorem is the power plant at the heart of the commonly used parts of *inferential statistics*. By inferential statistics, we mean the act of drawing conclusions from data, rather than merely describing the data (known as descriptive statistics) or visualizing it.

The "square root of npq rule" is a shortcut that allows people who have not yet learned about the normal distribution to be able to write down some useful inequalities.

Here's the big summary:

- There are n independent attempts at some event that occurs with probability p, and does not occur with probability $q = 1 - p$.
- The one-sigma range is
$$np - \sqrt{npq} < x < np + \sqrt{npq}$$
and this inequality will be true with probability 68.27%. (Sometimes this is called "fairly likely.")
- The two-sigma range is
$$np - 2\sqrt{npq} < x < np + 2\sqrt{npq}$$
and this inequality will be true with probability 95.45%. (Sometimes this is called "very likely.")
- The three-sigma range is
$$np - 3\sqrt{npq} < x < np + 3\sqrt{npq}$$
and this inequality will be true with probability 99.73%. (Sometimes this is called "extremely likely.")
- However, this only works if $np > 5$ and $nq > 5$. Some books say $np > 10$ and $nq > 10$.

The tricks in the next box will make it much easier to remember the formulas given above.



I'd like to guide you on a structural tour of the above inequalities.

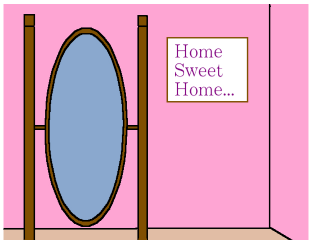
Computing the expected value, np, is easy. This can be remembered because "N P" is a classic slang abbreviation for "no problem," especially in emails. It is "no problem" to compute the expected value, because the expected value is np.

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.

You've seen the following pair of sentences four times now. "However, this only works if $np > 5$ and $nq > 5$. Some books say $np > 10$ and $nq > 10$."

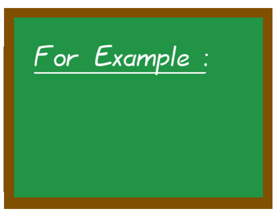
This is nothing more than it seems. Suppose I am analyzing a situation, such as a new textbook having a typo on a random page, and the probability is around $p = 0.02$. If I have $n = 200$ pages to analyze, then $np = 0.02(200) = 4$, and I must not use the square-root of npq rule—it would give me inaccurate answers if I did. If I have $n = 600$ pages to analyze, then $np = 0.02(600) = 12$, and then all textbooks that I've seen would say that the square-root of npq rule is safe.

If I have between 250 and 500 pages to analyze, then the published textbooks are in disagreement. Mostly older ones would say that it is okay, and newer ones would say that it is not okay. This is a little odd, but that's that.



A Pause for Reflection...
The following problems deal with car insurance and reference crime rates. These problems were written when I was teaching at Fordham University, in The Bronx, NY. They reflect the crime rates that were experienced in The Bronx during those years (2007–2011), and those rates were rather high. Most cities in the USA have lower crime rates than The Bronx, and even The Bronx has gotten a heck of a lot better after I left in 2011.

Accordingly, the estimates in the next few problems will seem rather high. This does show how even a discrete mathematics textbook can be influenced by the author's life experiences.



3-5-6

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 information does "the square root of *npq* rule" tell us about this situation? (Compute just the two-sigma version, also called the 95.45% range, or the "very likely" range.)

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

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

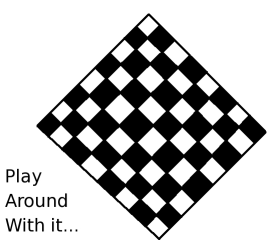
which simplifies to

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

and further simplifies to

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

In conclusion, the inequality is telling us that with probability 95.45%, the number of cars stolen will be between $171.857\cdots$ and $228.142\cdots$. It seems as though we should report an integer number of cars, discarding the 0.857 and 0.142. I'll say more about that on Page 352 of this module.

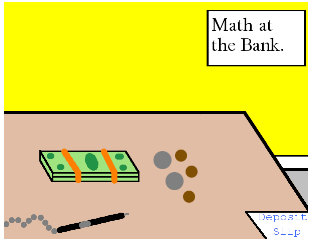


3-5-7

Let's try the previous example with a larger n. Suppose in a nearby city, 1% is again the probability that a car will be stolen during a given year, and 5% is again the probability that a car will be vandalized. However, the city has 250,000 cars, instead of 20,000 in the town of the previous box.

- What inequality does "the square root of *npq* rule" provide, to tell us the two-sigma or 95.45%-probable range 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: $12,282.0 < x < 12,717.9$, which can also be written $12,717.9 > x > 12,282.0$.]
- What inequality does "the square root of *npq* rule" provide, to tell us the two-sigma or 95.45%-probable range 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: $2599.49 > x > 2400.50$, which can also be written $2400.50 < x < 2599.49$.]

Again, 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. This is an industry standard, and we should be careful to follow it. We'll see precisely why this is done later, in the module "The Binomial Distribution."

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. In the example before that, we would say between 171 and 229 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."

I've actually glossed over a minor technicality. Engineers and financiers tend to use the 1-sigma, 2-sigma, and 3-sigma ranges, which signify 68.27%, 95.45%, and 99.73%. However, social scientists tend to prefer to use 95.00% in place of 95.45%.

Of course, $0.9545 \neq 0.9500$, so we have to make some adjustment. It turns out that to get 0.9500 instead of 0.9545, you have to replace the 2 in front of the square root signs with a 1.96 instead. This doesn't really matter at all for our purposes. To demonstrate why it doesn't matter, consider the previous checkerboard problem.

The number of vandalized cars (before rounding) would change from

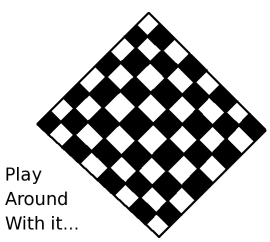
$$12,282.0\cdots < x < 12,717.9\cdots \text{ into instead } 12,286.4\cdots < x < 12,713.5\cdots$$

which is a difference of about four cars per year—nothing to get excited about.

Similarly, the number of stolen cars (before rounding) would change from

$$2400.50\cdots < x < 2599.49\cdots \text{ into instead } 2402.49\cdots < x < 2597.50\cdots$$

which is a difference of about 2–3 cars per year—even less exciting.

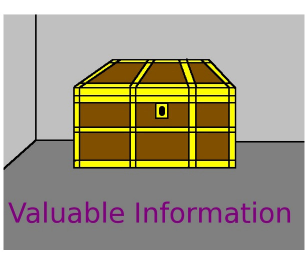


3-5-8

Let's repeat the previous checkerboard box, but with the 99.73% intervals, also known as the three-sigma intervals.

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

- With probability 99.73%, how many cars will be stolen?
  [Answer: between 2350 and 2650.]
- With probability 99.73%, how many cars will be vandalized?
  [Answer: between 12,173 and 12,827.]



By the way, there's a really cool way to check your work with the "square root of *npq* rule." 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 two-sigma (95.45% probability) interval, the three-sigma (99.73% probability) interval, or even the rarely-used one-sigma (68.27% probability) interval. For each interval, the center or midpoint (before rounding) is always exactly the expected value.

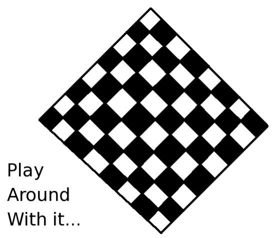
Let's check our work for the problem about stolen and vandalized cars. The expected values are

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

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

$$\frac{2400+2600}{2} = 2500 \qquad \leftarrow \text{two-sigma, stolen cars}$$
$$\frac{2350+2650}{2} = 2500 \qquad \leftarrow \text{three-sigma, vandalized cars}$$
$$\frac{12,173+12,827}{2} = 12,500 \qquad \leftarrow \text{two-sigma, stolen cars}$$
$$\frac{12,282+12,718}{2} = 12,500 \qquad \leftarrow \text{three-sigma, vandalized cars}$$

It looks like we got it right!

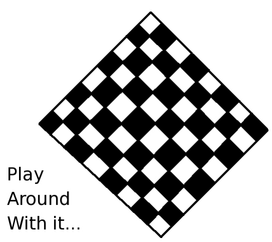


3-5-9

You might get bored at some point in this module, so I have a challenge problem for you. The most basic way of stating the question is as follows: "For what kinds of probabilities is the uncertainty in "the square-root npq rule" the largest? the smallest?"

We can state it better, as follows. Since the uncertainty in "the square-root npq rule" is $2\sqrt{npq}$ (or $3\sqrt{npq}$), for any fixed value of n, what value (or values) of p would make this uncertainty the largest possible? the smallest possible?

This problem is fairly hard and a little long. Don't be discouraged if you don't get it. The full solution is given on Page 375.

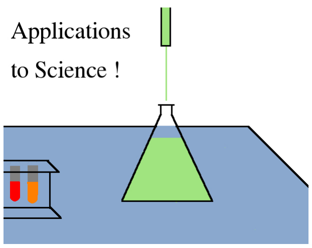


3-5-10

Suppose a garbage barge gets into an accident, dumping all the trash into the sea. A large fishery, somewhat nearby, is concerned that mercury and other heavy metals from the trash might be poisoning the fish. Therefore, they arrange for an environmental firm to visit and test 400 fish for mercury content. Further suppose it were the case that 10% of the fish were affected.

- How many fish would we expect the firm to find are affected?
  [Answer: The expected value is 40 affected fish.]
- In what range is the number of affected fish from the 400 fairly likely to fall, in the one-sigma sense (probability 68.27%)?
  [Answer: 34 to 46 affected fish.]
- In what range is the number of affected fish from the 400 very likely to fall, in the two-sigma sense (probability 95.45%)?
  [Answer: 28 to 52 affected fish.]
- In what range is the number of affected fish from the 400 extremely likely to fall, in the three-sigma sense (probability 99.73%)?
  [Answer: 22 to 58 affected fish.]

Usually around this time, someone asks the following question: "What if I want to know about some other probability, like 90% or 99%? What do I do then?" We will mostly address this issue on Page 357 of this module.



Mercury and other heavy metals are extremely poisonous. It is particularly known to collect inside fish and seafood. Mercury can cause developmental disorders, cognitive and behavioral problems, or even brain damage in young children (including while *in utero* if the fish or seafood is consumed by a pregnant mother).

One common way that mercury enters the air we breathe (or our drinking water) are the small batteries that are shaped like buttons. You'll find those in calculators, small electronic devices, and most especially toys. If a child's toy (or a college student's calculator) is broken, it seems natural to just throw it in the garbage. Actually, this is extremely bad—it should be recycled instead. Mercury is one major problem, but there are other dangers too, such as the flame retardants in printed circuit boards (PCBs).

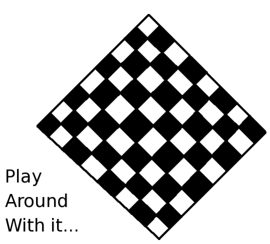
The issue of small electronic devices getting thrown into the trash is a particularly large problem in California, where small electronic devices are exceedingly common, and have been for a longer period of time than in most other places, and where seafood, fish, sushi and sashimi are consumed in enormous quantities.

The following is a warning taken from the website of the United States Environmental Protection Agency (EPA).

> Button cell batteries are miniature batteries in the shape of a coin or button. They are used in small portable electronic devices such as watches, cameras, digital thermometers, calculators and toys. Zinc air, alkaline, and silver oxide button cell batteries contain small amounts of mercury. These batteries do not pose a health risk when in use since the chances of the mercury leaking out are small.
>
> The mercury in button cell batteries can escape into the environment after they have been thrown away and are either incinerated or end up in landfills. Though there are no federal regulations prohibiting throwing button cell batteries in the regular garbage, they should be recycled instead. If they are not recycled, almost all of this mercury in them can end up in waste that gets incinerated or landfilled. If incinerated, the mercury can end up back in the air; if landfilled, it could end up in ground water, and potentially in sources of drinking water.

The bottom line is that if you have a small electronic device, and wish to throw it away, you should instead seek information about the local options for safely recycling electronics. Throwing away a thermometer, barometer, or an old thermostat can be harmful also.

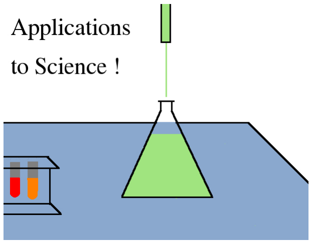
You can read more about this at the following websites:
`https://www.epa.gov/mercury/mercury-batteries`
`https://en.wikipedia.org/wiki/Mercury_in_fish`



3-5-11

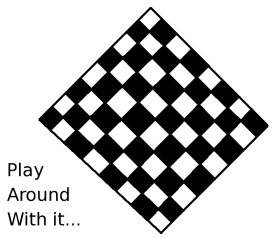
Let's suppose that there's a new medical procedure on the eyes that does wonderful things when it works, but makes 1% of the patients go blind. (That rate of a problem doesn't sound very large, does it? Only 1%?) It turns out that 1% is actually a huge rate of catastrophic failure. Generally, you'd like it to be below 0.0001. To grasp this more precisely, let's consider a county of 80,000 and suppose that a quarter of the population of the county has that procedure performed.

- What interval has a 95.45% chance of containing the number of people blinded by the procedure? [Answer: 229 to 171.]
- What interval has a 99.73% chance of containing the number of people blinded by the procedure? [Answer: 243 to 157.]



As you can see from the previous box, a medical procedure having a catastrophic failure rate of 1% is horrifically high. People who have studied Science, Technology, Engineering, and Mathematics understand this, but ordinary people do not.

Imagine if a pharmaceutical leaves 0.9% of the patients blind. A pharmaceutical salesman can say, “Yes, there is some possibility this will make you go blind, but the probability is less than one percent.” Having taught for a while—the Spring of 2018 is my 25th semester teaching—I can tell you that most ordinary people would hear “less than one percent,” and dismiss the risk as being essentially zero.



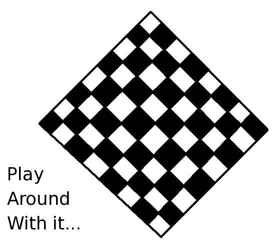
3-5-12

Imagine a calculus textbook that is being translated into Spanish. Like most calculus books, it is about 1100 pages. (This is about the right size, by the way, because of some details of how book binding works. Almost all massive math texts, like calculus books or Sullivan and Sullivan’s *College Mathematics* are between 1050 and 1150 pages.)

An amazingly good translator has been hired, one who makes a mistake on a given page with probability only 2%.

- What is the expected value of the number of typos? [Answer: 22.]
- What is the 68.27% probability-range for the number of typos? [Answer: 17 to 27.]
- What is the 95.45% probability-range for the number of typos? [Answer: 12 to 32.]
- What is the 99.73% probability-range for the number of typos? [Answer: 8 to 36.]

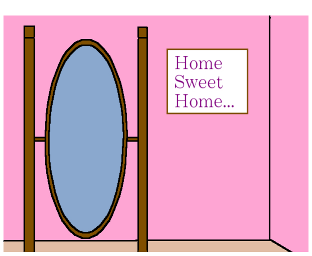
This is startling, is it not? Surely 8 typos and 36 typos are not identical outcomes.



3-5-13

Let’s suppose that on a particular college campus, 5% of the student body is openly gay or lesbian. The campus newspaper is doing an article for National Coming Out Day, and they are going to survey 250 students, asking each if he or she is openly gay or lesbian.

- What is the expected number of yes’s? [Answer: 12.5.]
- What is the two-sigma (95.45% probability) interval for the number of students who say “yes,” before rounding? [Answer: $5.60797\cdots$ to $19.3920\cdots$ students.]
- After rounding? [Answer: 5 to 20 students.]
- What do we get if we convert this to a percentage? [Answer: 2% to 8%.]

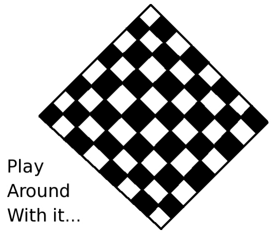


A Pause for Reflection. . .

Looking at the previous box, the uncertainty is huge.

That’s why polling error is a serious matter. Among other things, if the 5% were to stay absolutely constant, and the survey were to be repeated semester after semester, the observed responses could easily be 3%, 7%, 6.5%, 2.5%, 7.5%.

Those fluctuations could cause all sorts of editorials to be written and published—each semester—about what political or sociological phenomena could be driving such enormous changes in campus life, even if the underlying rate is staying rock solid at exactly 5%. We’ll explore polling error in more detail at several points in this module.



3-5-14

Let's suppose that something has gone wrong at the manufacturing line of a plant that makes keyboards for laptops. In particular, it turns out that 5% of the laptop keyboards are defective. The staff at the factory have not yet computed this percentage—they are going to calculate that momentarily. First, they will inspect the last 1000 laptop keyboards and check each one.

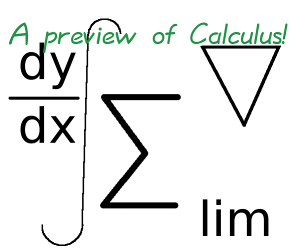
- What is the expected value of the number of defective keyboards that they'll find? [Answer: 50.]
- What is the 95.45% probability-range for the number of defective keyboards? [Answer: 36 to 64.]
- What is the 99.73% probability-range for the number of defective keyboards? [Answer: 29 to 71.]

Let's re-examine the data in the previous box. As noted on Page 276 of the module "A Formal Introduction to Probability Theory," the "Law of Large Numbers" is not called the law of somewhat-biggish numbers. We have fairly large windows of uncertainty in this case.

If we repeated the problem substituting 1,000,000 for 1000, then we obtain

- What is the expected value of the number of defective keyboards? [Answer: 50,000.]
- What is the 95.45% probability-range for the number of defective keyboards? [Answer: 49,564 to 50,436.]
- What is the 99.73% probability-range for the number of defective keyboards? [Answer: 49,346 to 50,654.]

As you can see, the uncertainty is considerably reduced when we check 1,000,000 keyboards instead of 1000 keyboards. This is why it is important, especially in safety studies, to have a large n. We'll return to the circumstances of this box and the previous box later, on Page 366.

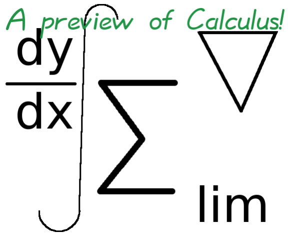


For some reason, there is a tradition in discrete mathematics textbooks to avoid talking about calculus. That seems very strange to me, but I think partly this is to recognize that discrete mathematics is an independent branch of mathematics, separate from the precalculus/calculus/differential equations "trunk line" that forms the foundation of engineering degrees, the math minor, and the first half of a Bachelor's Degree in mathematics. Partly, this might also be the case because (in theory) some students might not have had any calculus before reaching discrete mathematics—yet, that does not appear to be possible at any university that I know of.

The reason that I'm bringing up calculus is that I want to tell you where the numbers 68.27%, 95.45%, and 99.73% came from. They certainly are not arbitrary. This box and the next two boxes will probably only make sense if you've had a university-level statistics course. However, please read them whether you've had such a course or not.

As it turns out, if a random variable X has "the Gaussian Distribution" or "the normal distribution," with a mean of 0 and a standard deviation of 1, then the probability that $a < X < b$ is given by the integral

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx = Pr\{a < X < b\}$$

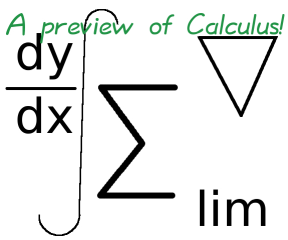


The integral in the previous box can be calculated numerically, but not symbolically, just like Fresnel's integral and a few other famous non-symbolic integrals. After all, that's why numerical integration exists as a subject.

Permit me to share with you three numerical computations:

$$\begin{aligned}
\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx &= 0.6827\cdots = Pr\{-1 < X < 1\} \\
\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx &= 0.9545\cdots = Pr\{-2 < X < 2\} \\
\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx &= 0.9973\cdots = Pr\{-3 < X < 3\}
\end{aligned}$$

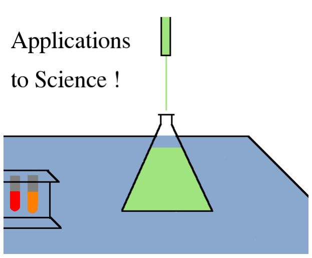
This is where the 68.27%, the 95.45%, and the 99.73% come from. Often, students ask me if other percentages are possible. They are, and we'll explore that in the next box.



As I mentioned in the previous box, you can choose other values as well. For example, it turns out that

$$\begin{aligned}
\int_{-2.58\cdots}^{2.58\cdots} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx &= 0.9900\cdots = Pr\{-2.58\cdots < X < 2.58\cdots\} \\
\int_{-1.96\cdots}^{1.96\cdots} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx &= 0.9500\cdots = Pr\{-1.96\cdots < X < 1.96\cdots\} \\
\int_{-1.645\cdots}^{1.645\cdots} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}\, dx &= 0.9000\cdots = Pr\{-1.645\cdots < X < 1.645\cdots\}
\end{aligned}$$

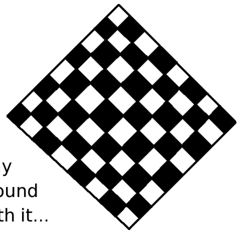
Therefore, if your boss asks you for 99.00% in place of 95.45% or 99.73%, then you can put a 2.58 in place of the 2 or 3, and produce the correct answer to your boss's question. While I am glad to have shared this connection between integrals and statistics with you, as it comes to pass, we will have no further use of integrals in this chapter.



There's another huge application of what we are talking about here to manufacturing. That's six-sigma engineering. The whole "six-sigma movement" is of enormous importance in manufacturing today.

A firm that uses six-sigma engineering will expect 3.4 defects per million manufactured items. This manufacturing concept has been championed by many large corporations, but most famously Toyota and General Electric. Sadly, I don't have time to go into that now. However, you can learn more from the following two articles:

- `https://www.isixsigma.com/new-to-six-sigma/getting-started/what-six-sigma/`
- `https://www.isixsigma.com/new-to-six-sigma/getting-started/six-sigma-qa/`



Play
Around
With it...

3-5-15

I'd like to turn away from manufacturing now, and turn toward another use of "the square root of npq rule." Back on Page 276 of the module "A Formal Introduction to Probability Theory," in order to explain the "Law of Large Numbers," we considered a thought experiment: a coin is flipped a really large number of times, and we count the number of heads.

Suppose a fair coin is flipped a trillion times, and let x be the number of heads.

- What interval will contain x with probability 68.27%?
  [Answer: $499,999,500,000 < x < 500,000,500,000$.]
- What interval will contain x with probability 95.45%?
  [Answer: $499,999,000,000 < x < 500,001,000,000$.]
- What interval will contain x with probability 99.73%?
  [Answer: $499,998,500,000 < x < 500,001,500,000$.]

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 two-sigma version of "the square-root of npq rule" is telling us that we will get an answer within 1,000,000 of that, with probability 95.45%.

Meanwhile, the one-sigma version is telling us that we will get an answer within 500,000 of that, with probability 68.27%, and the three-sigma version is telling us that we will get an answer within 1,500,000 of that, with probability 99.73%.

Back on Page 276 of the module "A Formal Introduction to Probability Theory," I wrote that I had used Sage and the Bernouli-DeMoivre-Laplace inequalities (i.e. Sage and "the square root of npq rule") to compute the data in the next box. I used the three-sigma, or 99.73% probability, versions.

This was neat, because we were able to see that 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. (By the way, did you notice that the last bullet-point of the previous checkerboard box is precisely and exactly the last data entry in our table?)

The reason that I'm bringing this up again is that "the square root of npq rule" is the main component of the Sage code which made that table below. You might be interested to see the code—it is given after the table.

Here is the data mentioned in the previous box. These bounds hold with probability 99.73%.

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

```
p=1/2

for k in range(1, 13):
    # the above line means that k will be 1, 2, 3, ..., 12

    n = 10^k

    lower = floor( n*p - 3*sqrt( n*p*(1-p) ) )
    upper = ceil( n*p + 3*sqrt( n*p*(1-p) ) )

    print "For n =", n,
    print "it will be between", lower,
    print "and", upper, "heads."
```

If you look carefully at that code, you will see the "square root of npq rule" in two places. Namely, the computation of `lower` and `upper`. The commands `floor` and `ceil` simply mean "round down" and "round up," respectively.

Since we wanted the 99.73%-version (the 3-sigma version) of the intervals, we put a 3 in front of the square root sign using:

$$np - 3\sqrt{np(1-p)} < x < np + 3\sqrt{np(1-p)}$$

Instead of flipping a coin a billion times, if I were to roll an ordinary six-sided die exactly six billion times, then I would expect one billion sixes. What if I roll the die six million times? Six thousand times? Six hundred times? Sixty times? Let's find out.

Looking at that Sage code above (which you can cut-and-paste into your web browser), let's make the following changes.

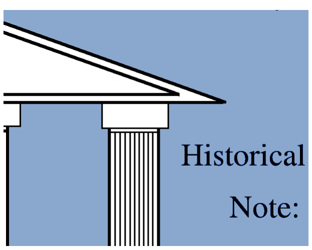
- Change `p=1/2` into `p=1/6`.
- Change `n = 10^k` into `n = 6*10^k`.
- Change `print "and", upper, "heads."` into `print "and", upper, "sixes."`
- Now click "Evaluate."

If you do the above correctly, you should receive the output given below this box.

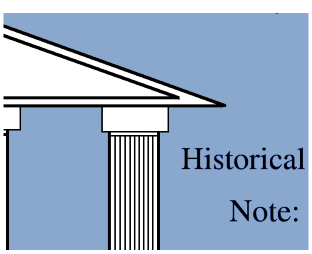
```
For n = 60 it will be between 1 and 19 sixes.
For n = 600 it will be between 72 and 128 sixes.
For n = 6000 it will be between 913 and 1087 sixes.
For n = 60000 it will be between 9726 and 10274 sixes.
For n = 600000 it will be between 99133 and 100867 sixes.
For n = 6000000 it will be between 997261 and 1002739 sixes.
For n = 60000000 it will be between 9991339 and 10008661 sixes.
For n = 600000000 it will be between 99972613 and 100027387 sixes.
For n = 6000000000 it will be between 999913397 and 1000086603 sixes.
For n = 60000000000 it will be between 9999726138 and 10000273862 sixes.
For n = 600000000000 it will be between 99999133974 and 100000866026 sixes.
For n = 6000000000000 it will be between 999997261387 and 1000002738613 sixes.
```

Looking at the data above, you can see that somewhere around 600,000 dice rolls will cause the uncertainty to be very tiny. Even at $n = 6000$ the uncertainty is significant, and the uncertainty is very large for $n = 600$ and $n = 60$.

Once again, this is why we call it the “Law of Large Numbers,” and not “the law of somewhat biggish numbers.”

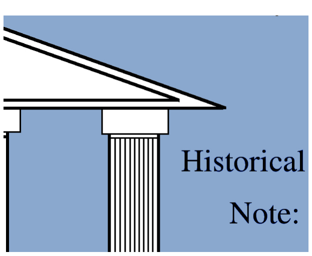


Many textbooks omit Bernoulli’s name and refer to this as the “de Moivre-Laplace Central Limit Theorem.” However, that’s not very fair, as Bernoulli was the first to develop this idea. In spoken English, it is common for both students and faculty to refer to this result as “the Central Limit Theorem” to avoid awkwardly mispronouncing the French, or as the “square root of npq rule.” Technically, calling it “the Central Limit Theorem” is wrong, because that theorem says (see Page 349) that the number of observed outcomes is normal/Gaussian, and the inequalities we’ve been discussing are corollaries of that theorem. Also, we’ve only spoken about events that either happen, or don’t, whereas “the Central Limit Theorem” applies in many, many other places, such as adding or averaging numbers.



If you’d like to pronounce the foreign 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—the Binomial Distribution formula—a formula that lies at the heart of the science of reliability. We’ll learn about that in the module called “The Binomial Distribution.” We also talked about the idea of a “Bernoulli Random Variable” in the module “A Formal Introduction to Probability Theory.”

Meanwhile, we will discuss Abraham de Moivre in the next box.



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. (In case you would like to know the pronunciation: ah-BRAH-am duh MWAHV-ruh. However, the “ruh” at the end is guttural, de-emphasized, and almost silent.)

I will summarize Abraham de Moivre’s mathematical accomplishments in the next box.

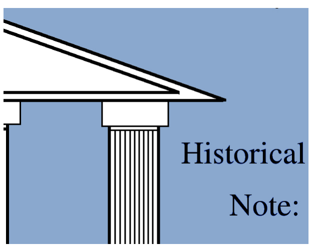
The contributions of Abraham de Moivre are numerous. He's primarily known among mathematicians for a formula which allows for the roots and exponents of complex-imaginary numbers to be computed, such as

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

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.

Abraham de Moivre contributed to financial mathematics by writing 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.

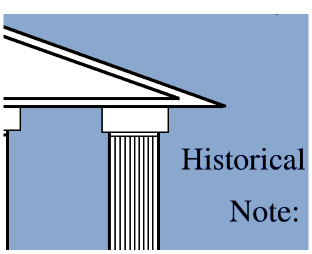
Professionally, Abraham de Moivre had major contributions. Yet, he was not able to earn a proper living, for reasons that we'll now explore in the next few boxes.



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.

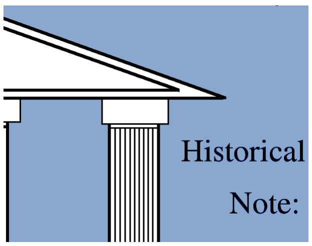
By now, you've surely noticed a bunch of French names in the history of mathematics, but also in other scholarly subjects like science and philosophy. That's because France, in stark contrast to the rest of Europe, had religious freedom for a critical window of time. People from all over Europe would immigrate to France, making Paris an intellectual melting pot. You might think that religious freedom and mathematics are unrelated, but the fact of the matter is that when religious freedom was granted, the censorship of books was instantaneously abandoned. Overnight, the reality went from "all books published must be officially checked by royal censors," to the complete absence of any such requirement.

Permit me to offer some more information about this 87-year golden age in the next box.



The period of religious freedom 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 revoked religious freedom in France.

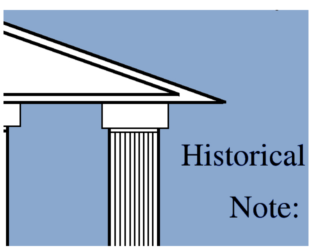


Continuing with the previous box, Louis XIV is not the Louis that got his head cut off in the guillotine.

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). (You might remember that we discussed the 87 years between these two edicts back on Page 165 in the module "Intermediate Set Theory and Irrationality," and on Page 231 as well as Page 256 in the module "Fermat's Last Theorem, Set Theory, and Number Theory.")

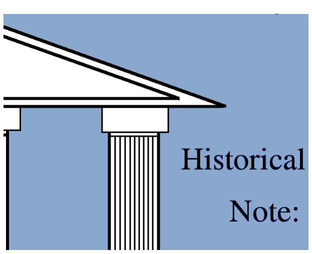
The abrupt termination of religious freedom and the consequent trials and persecutions helped inspire both the American Revolution and the French Revolution, making this story a crucial turning point in the history of both Europe and the USA. In fact, the town New Rochelle, NY, which is a few miles from 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 a 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.



Abraham de Moivre earned supplementary money by playing and teaching chess in coffee houses. He 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.

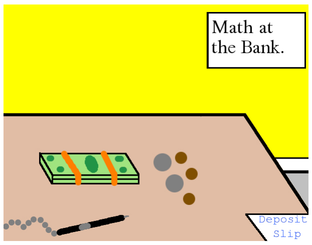
If you've taken calculus, then you've probably heard of the furious debate over who invented it: Isaac Newton (1642–1726) or Gottfried Leibnitz (1646–1716). Abraham de Moivre was appointed to the committee for deciding the Newton versus Leibnitz dispute, next to many of these famous mathematicians. Yet, he never held a university position.

Having learned about the troubles 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 history, see *Abraham De Moivre: Setting the Stage for Classical Probability and Its Applications* by David R. Bellhouse, published by CRC Press in 2011.

So far, we've applied "the square root of *npq* rule" to its most common applications: insurance-related problems, manufacturing engineering or reliability engineering problems, and explaining the "Law of Large Numbers." Now we're going to apply it to understanding how casinos work, and why they are so profitable!

Generally, I have avoided problems about dice and cards throughout the probability modules of this textbook. The reason is that such problems are extremely common in high school probability lessons, and therefore students have seen more than enough of these problems by the time they arrive on a university campus.

I'm going to make one exception now. We're going to examine why casinos are so profitable. In particular, while each spin of a Roulette wheel is unpredictable, the sum of the incomes and payouts over an entire year (for one Roulette wheel) is fairly predictable. This summing is an excellent showcase of the "square root of npq rule."



Have you ever wondered why gangsters and mafia are often associated with casinos? Anyone who has visited Las Vegas can testify to the enormous wealth of the hotels and casinos, with their flagrant displays of lights and fountains on "the strip." It seems like a risky business, because the casino cannot predict the outcome of individual spins of the roulette wheel, or hands of poker or blackjack.

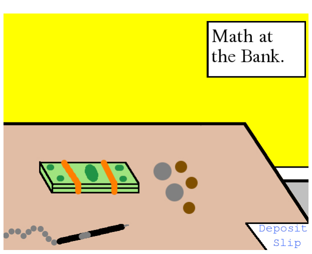
As it turns out, while the individual spins of the roulette wheel are unpredictable, their sum is predictable enough over long periods of time. We will analyze this now, with the help of "the square root of npq rule." We will analyze Roulette, as it is the simplest game.

In the game of Roulette, there is a wheel that can spin. After bets are placed, the wheel is spun, and an attendant (called a "croupier") will drop a white ball onto the wheel. The white ball bounces around, and lands on one of the wheel's little boxes on the perimeter.

The boxes are numbered 0–36, and in most of the world, there is also a 00. The 0 (and 00 if present) are green. Half the other numbers are red, and half are black, giving 18 of each color.

In France and Monaco, they do not have the 00, in contrast to most of the world. Since most of the world uses both 0 and 00, we will use them in our exploration. It turns out that these two lonely boxes are how the Casino derives profit from Roulette. As you can see, the photo to the left is of a roulette wheel from France or Monaco.

The above image was generated by Toni Lozano, who uploaded it to the WikiMedia Commons in 2006. Like this textbook, the image is shared under "The Creative Commons," and I am happy to offer this academic citation.



Before we begin, we need to estimate how many bets per year might be made at an individual Roulette table.

Let's suppose that a particular Roulette wheel in a large and busy casino gets 200,000 bets per year. This is not hard to imagine, because there are

$$(365)(24)(60) = 525,600 \text{ minutes}$$

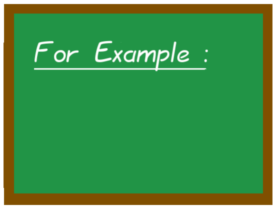
in most years, or 527,040 minutes in a leap year.

Most casinos are open 24/7 and it easy for six or seven bets to be made per spin, and a spin every other minute. If the casino were constantly busy, then we might estimate

$$(6.5)(525,600)/2 = 1,708,200 \text{ bets/year}$$

but surely the casino is not constantly busy. There can't be as many bets at 9 am on a Tuesday in late January as there would be during the peak season on a Friday night.

With all that in mind, 200,000 bets per year seems like a good but conservative estimate.



3-5-16

The favorite bet of experienced gamblers in Roulette is to bet on red or bet on black. Let's assume there are 200,000 bets on red or black per year, each of a single dollar. What is the 2-sigma (95.45% probability) interval for the casino's profits?

As mentioned in the previous box, there are 18 red spots, 18 black spots, and 2 green spots. Since the little white ball is equally likely to fall on any of the 38 spots, a gambler who bets on red will win with probability $p = 18/38 = 9/19$. The probability is the same if the gambler bets on black. Of course, $n = 200,000$ and $q = 1 - p = 1 - 9/19 = 10/19$.

We will do the computation in the next box.

For the number of bets where the gambler wins, we have

$$\begin{array}{rcccl} np - 2\sqrt{npq} & < & x & < & np + 2\sqrt{npq} \\ (200,000)(9/19) - 2\sqrt{200,000(9/19)(10/19)} & < & x & < & (200,000)(9/19) + 2\sqrt{200,000(9/19)(10/19)} \\ 94,736.8\cdots - 446.593\cdots & < & x & < & 94736.8\cdots + 446.593\cdots \\ 94,290.2\cdots & < & x & < & 95,183.4\cdots \end{array}$$

With the special rounding rules, we have that x is between 94,290 and 95,184. We will continue the problem in the next box, after a brief check.

Before we continue with the calculation of the previous box, let's pause—just for a moment—and check our work.

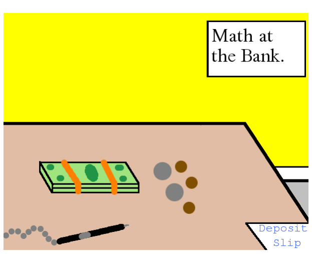
The expected value is

$$np = (200,000)\left(\frac{9}{19}\right) = 94,736.8\cdots$$

The average of the lower and upper bounds of the 2-sigma range, before rounding, comes to

$$\frac{94,290.2\cdots + 95,183.4\cdots}{2} = 94,736.8\cdots$$

which is an exact match. Now we can continue the calculation of the previous box with confidence, and we'll do that in the next box.



Continuing with the previous box, in the event of a win, the payout is \$ 2, being a return of the \$ 1 bet, and \$ 1 of winnings. So the total cost c paid out would be obtained by multiplying by 2. We have

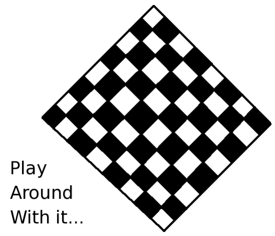
$$\$\ 188,580 < c < \$\ 190,368$$

Next, the revenue is \$ 200,000, since there were 200,000 bets of \$ 1 each. The profit is the revenue minus the cost. We get

$$\$\ 11,420 > 200,000 - c > \$\ 9632$$

which means that the casino will, with probability 95.45%, make between \$ 9632 and \$ 11,420 from this category of bet at one table. It is important to keep in mind that the casino will have many tables.

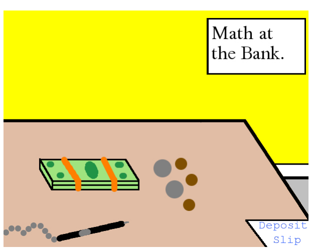
As you can see, they simply aren't taking a risk in any meaningful sense. Their profit is assured. Naturally, any casino has many, many tables.



3-5-17

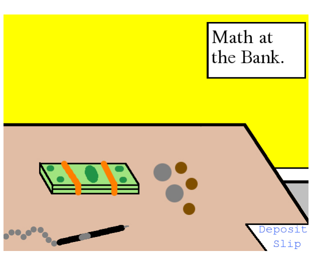
The most common bet at the Roulette table is to bet on your favorite number. For example, since my birthday is December 20th, I might bet on 20 or on 12. Let's perform the same analysis on bets of a particular number, but let's estimate that there are 100,000 bets per year of this type. Of course, since the probability of the little white ball landing upon the specific number that you want is only 1/38, this bet must be incentivized. If the number is selected, the casino must payout \$ 36 for a \$ 1 bet, which represents a return of the \$ 1 bet, and \$ 35 in winnings.

- What is the expected number of winning bets? [Answer: $100,000/38 = 2631.57\cdots$.]
- What is the expected cost of the payouts? [Answer: \$ 94,736.84.]
- What is the expected profit? [Answer: \$ 5263.15.]
- What is the 2-sigma range for the number of winning bets? [Answer: 2530 to 2733.]
- What is the 2-sigma range for the total cost of the payouts? [Answer: \$ 91,080 to \$ 98,388.]
- What is the 2-sigma range for the profit? [Answer: \$ 8920 down to \$ 1612.]



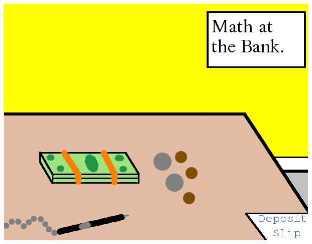
As you can see from the previous box, the uncertainty of the amount of the profit is larger the case of betting on a single number, but there is no meaningful risk of losing money.

Moreover, the total profit from these two categories of bets, from each Roulette table, will be very likely to be inside \$ 11,244 to \$ 20,340. (I obtained these by adding \$ 1612 and \$ 9632 as well as \$ 8920 and \$ 11,420, which is a crude approximation—combining two intervals is actually a bit of a tricky calculation, but exploring that point is not useful for us.) There are other categories of bets at a Roulette table, but since these are the most common bets, the profits will be only slightly larger. As you can see, there is a steady flow of money from the set of all gamblers, into the casino, every year. And of course, any casino has many tables.



We have now established that this is a profitable business, but there are also many other types of business that are profitable. Why are casinos the favorite of gangsters and the mafia? Because of the variability! It is very easy to hide income from illegal activities by pretending that the proceeds from the gambling were a little bit higher than expected. After all, there is a large interval into which the profits are very likely to fall.

Those of you who don't gamble are probably thinking that the previous several boxes, studying Roulette in detail, were a waste of your time. However, the model that I portrayed over the last few boxes is actually very similar to the model that we would use for simple insurance products.

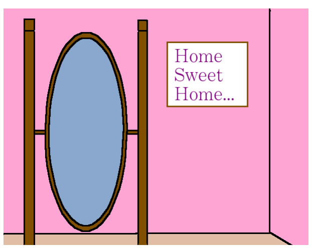


For example, whenever I buy airplane tickets, I am offered travelers insurance as an optional (but highly recommended) purchase. The insurance company knows the historical probabilities of flight passengers getting hospitalized, into car accidents, or otherwise becoming too sick to fly, and these were obtained by analyzing data. I simply don't know what that probability is, but we could estimate it at perhaps 2%. The Roulette wheel lands on my favorite number with probability 1/38. These are actually the same concept mathematically.

It is, once again, a Bernoulli Random Variable, a weighted coin.

- For the traveler's insurance:
  - With some probability (maybe 2%), I'm too sick to fly, and the insurance company must refund me the cost of my flight.
  - It is far more probable (maybe 98%) that I am not too sick to fly, and they don't pay out.
- For Roulette:
  - With probability 1/38, the little white ball will land on my favorite number, and the casino must pay me my winnings.
  - It is far more probable that the little white ball will land on some other number, and the casino doesn't have to pay me anything.

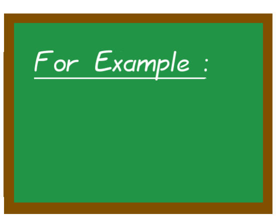
This is mathematically a very unfair coin that lands on heads with probability 1.5% or 1/38, and lands on tails with probability 98.5% or 37/38.



A Pause for Reflection...

The previous box shows you why it is such a pleasure to teach a course like *Discrete Mathematics*. When you talk to ordinary people, they would be surprised to learn that these situations—the game of Roulette, flight insurance, and a very unfair coin—are similar. In fact, they are not only similar, but they are mathematically modeled exactly the same way. If you speak with ordinary people, they would probably disbelieve my previous sentence.

That's why it is such a pleasure to teach STEM students (Science, Technology, Engineering, and Mathematics) students instead of ordinary people. The way that you think is spectacularly different to the way that ordinary people "think."



3-5-18

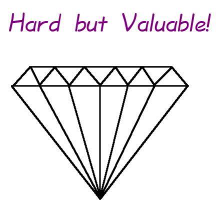
Let's return to the problem about the factory making laptop keyboards. We know, but the staff doesn't know, that 5% of the keyboards are flawed. They take a sample of 1000 keyboards, and test them. They probably won't get 50 on the dot. The staff might get 56, 44, or 51 for example. After counting the number of defective keyboards, the staff probably would compute the probability of a defective keyboard. They might obtain something like

$$\frac{56}{1000} = 0.056 \qquad \text{or} \qquad \frac{44}{1000} = 0.044 \qquad \text{or} \qquad \frac{51}{1000} = 0.051$$

depending on what the count of defective keyboards actually turns out to be.

We can generalize this, using the intervals we found before. The estimate of the staff will, with 95.45% probability, be between 0.036 and 0.064, because with 95.45% probability they obtain 36 to 64 defective keyboards. The estimate will, with 99.73% probability, be between 0.029 and 0.071, because with 99.73% probability they obtain 29 to 71 defective keyboards. Note, telling your bosses 3.6%–6.4% or 2.9%–7.1% will leave them feeling fairly uncertain—but that's too bad, because the situation is fundamentally uncertain.

As you can see, it is very hard to accurately estimate probabilities by observation. There will always be an interval that you obtain from an experiment. The interval will contain the true probability with probability 95.45% or 99.73%.



We are about to explore something very advanced, but for a good reason. Just as a novelist uses foreshadowing to prepare the reader for what comes later, good coursework should also foreshadow advanced topics. We're going to take a sneak peak at the inherent uncertainty that is entirely unavoidable when estimating probabilities. In fact, we're going to see that estimating probabilities is very hard unless you have a very large n.

In the physical sciences, including engineering, this experimental error is routinely encountered, and resolved by using a computer to compute confidence intervals. However, the underlying mathematics and statistical processes used to compute those confidence intervals are fairly complicated (especially compared to the rest of probability). For this reason, this is usually only taught at the end of a course on statistics—probably in a hurry, to students who are stressed out about impending final exams. Nonetheless, we have the tools now to learn why this uncertainty is unavoidable, and where it comes from—extremely valuable insights.

What follows should probably be read on a different evening/afternoon as the body of the module above, so that you will have had time to digest what you've learned before treading in to more advanced territory.

Let's re-examine the previous example. The symbol $\hat{p}$ is used to denote the estimate of p, where as the symbol p represents the true value of p. When reading equations or inequalities aloud, we say "p hat" for $\hat{p}$.

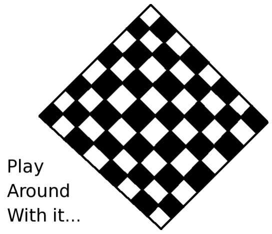
How did we obtain that $\hat{p}$ is to fall inside the interval $0.036 < \hat{p} < 0.064$ with 95.45% probability? This was obtained by taking the 95.45% probability (two-sigma) interval for the number of defective keyboards, and dividing it by n. Therefore, we can do the same thing to the original formula, in order get a general inequality.

We can obtain this general inequality by taking the inequality for the 95.45% probability (two-sigma) interval, and dividing it by n.

$$\begin{array}{rcccl}
pn + 2\sqrt{npq} & > & x & > & pn - 2\sqrt{npq} \\
\frac{pn+2\sqrt{npq}}{n} & > & \frac{x}{n} & > & \frac{pn-2\sqrt{npq}}{n} \\
\frac{pn}{n} + \frac{2\sqrt{npq}}{n} & > & \frac{x}{n} & > & \frac{pn}{n} - \frac{2\sqrt{npq}}{n} \\
p + \frac{2\sqrt{npq}}{n} & > & \frac{x}{n} & > & p - \frac{2\sqrt{npq}}{n} \\
p + \frac{2\sqrt{npq}}{\sqrt{n^2}} & > & \frac{x}{n} & > & p - \frac{2\sqrt{npq}}{\sqrt{n^2}} \\
p + 2\sqrt{\frac{npq}{n^2}} & > & \frac{x}{n} & > & p - 2\sqrt{\frac{npq}{n^2}} \\
p + 2\sqrt{\frac{pq}{n}} & > & \frac{x}{n} & > & p - 2\sqrt{\frac{pq}{n}} \\
p + 2\sqrt{\frac{pq}{n}} & > & \hat{p} & > & p - 2\sqrt{\frac{pq}{n}}
\end{array}$$

To see the power of this inequality, let's solve a problem without it, and then see how much of a shortcut the inequality can provide.

In political science, exactly the same phenomenon occurs. A poll is taken to consult the opinions of n people. Unless n is huge, there is actually a surprising amount of uncertainty in the result. The result should always be an interval of percentages, not a percentage. However, I read polling data extremely often, usually produced by very prestigious polling firms like Gallup Polls or the Pew Research Center. They do not report intervals, but instead give percentages. Yes, at the bottom of the webpage is a remark like "this survey had a margin of error of $\pm 3\%$," but as we will see shortly, such a remark in fine print does not properly capture the full picture of what is going on.



3-5-19

Let's suppose that 80% of the population of a state is in favor of the Senate candidate John Q. Public. The state newspaper is going to carry out a poll every Monday, Wednesday, and Friday, and publish the results. Each poll is of 1000 residents. For the purposes of this problem, let's assume that the 80% stays constant over the week.

- What is the expected number of surveyed people who will say that they are voting for John Q. Public? [Answer: 800.]
- What interval will contain the number of surveyed people who said they are voting for John Q. Public, with probability 95.45%? [Answer: 774 to 826.]
- What interval will contain the newspaper's prediction for the percentage of people in the state who will vote for John Q. Public, with probability 95.45%? [Answer: $774/1000 = 77.4\%$ to $826/1000 = 82.6\%$.]

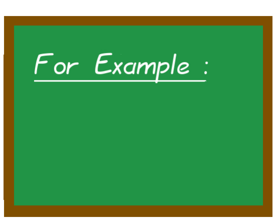
Let's look again at the problem in the previous box.

As those polls come in, three times a week for several weeks, even if the true percentage stays locked at 80% for some reason, the observed percentage—the newspaper's prediction—will fluctuate. Despite that fluctuation, it is very likely to fall inside the interval 77.4% to 82.6%. Each particular observation will be inside that interval with probability 95.45%.

For this reason, we would say that "the margin of error of this poll is $\pm 2.6\%$," because $77.4\% = 80\% - 2.6\%$ and $82.6\% = 80\% + 2.6\%$. Note that the margin of error can be obtained directly with

$$2\sqrt{pq/n} = 2\sqrt{(0.8)(0.2)/1000} = 2\sqrt{0.00016} = 0.0252982\cdots$$

which, because of rounding, ended up being 2.6%. Political polls always use the two-sigma interval, so far as I can tell.



3-5-20

Now let's use our new inequality for $\hat{p}$, using the given data that $p = 0.8$ and $n = 1000$, and see what we get.

Of course, $q = 1 - p = 1 - 0.8 = 0.2$. Plugging in p, q, and n, we get the following:

$$\begin{array}{rcccl} p + 2\sqrt{\frac{pq}{n}} & > & \hat{p} & > & p - 2\sqrt{\frac{pq}{n}} \\ 0.8 + 2\sqrt{\frac{0.8(0.2)}{1000}} & > & \hat{p} & > & 0.8 - 2\sqrt{\frac{0.8(0.2)}{1000}} \\ 0.8 + 2\sqrt{0.00016} & > & \hat{p} & > & 0.8 - 2\sqrt{0.00016} \\ 0.8 + 0.0252982\cdots & > & \hat{p} & > & 0.8 - 0.0252982\cdots \\ 0.825298\cdots & > & \hat{p} & > & 0.774701\cdots \end{array}$$

The only discrepency comes from the fact that the computation of the two-sigma interval (774 to 826) involved rounding.

For Example :

\# 3-5-21

Now let's return to the hypothetical case of a factory checking 1,000,000 keyboards. (They wouldn't actually do that of course, but my purpose here is not to make a realistic example, but to show an effect of the "Law of Large Numbers.") Earlier, we computed that they would find 49,564 to 50,436 defective keyboards, so their $\hat{p}$ would fall into the interval 0.049564 to 0.050436 with probability 95.45%. Let's ask our new inequality, "in what interval is the observed $\hat{p}$ going to fall into, with probability 95.45%?"

We plug in $p = 0.05$, $q = 1 - 0.05 = 0.95$, and $n = 1,000,000$. We obtain the following

$$\begin{array}{rcccl} p + 2\sqrt{\frac{pq}{n}} & > & \hat{p} & > & p - 2\sqrt{\frac{pq}{n}} \\ 0.05 + 2\sqrt{\frac{0.05(0.95)}{1{,}000{,}000}} & > & \hat{p} & > & 0.05 - 2\sqrt{\frac{0.05(0.95)}{1{,}000{,}000}} \\ 0.05 + 2\sqrt{0.0000000475} & > & \hat{p} & > & 0.05 - 2\sqrt{0.0000000475} \\ 0.05 + 0.000435889894354 & > & \hat{p} & > & 0.05 - 0.000435889894354 \\ 0.0504358\cdots & > & \hat{p} & > & 0.0495641\cdots \end{array}$$

It is almost disturbing how accurate this is! In any case, we can see that with a large n, like $n = 1,000,000$, these experimental computations are fairly accurate.

For Example :

\# 3-5-22

Let's imagine a town where a school board has written a new proposal that must be approved by a referendum. Of course, some members of the board are campaigning in favor of the proposal, and perhaps some others are campaigning against it. Suppose a local newspaper interviews 500 people exiting a local grocery store, and 55% of the people surveyed say they are in favor of the new proposal. What is the percentage of the town in favor of the new proposal? Of course, this must be an interval, not a number, so let's use the 95.45% (two-sigma) interval.

We plug in $p = 0.55$, $q = 1 - 0.55 = 0.45$, and $n = 500$. We obtain the following

$$\begin{array}{rcccl} p + 2\sqrt{\frac{pq}{n}} & > & \hat{p} & > & p - 2\sqrt{\frac{pq}{n}} \\ 0.55 + 2\sqrt{\frac{0.55(0.45)}{500}} & > & \hat{p} & > & 0.55 - 2\sqrt{\frac{0.55(0.45)}{500}} \\ 0.55 + 2\sqrt{0.000495} & > & \hat{p} & > & 0.55 - 2\sqrt{0.000495} \\ 0.55 + 0.0444971\cdots & > & \hat{p} & > & 0.55 - 0.0444971\cdots \\ 0.5944971\cdots & > & \hat{p} & > & 0.505502\cdots \end{array}$$

Therefore, we can conclude that between 50.55% and 59.45% of the town is in favor of the new proposal. That's actually a reasonably narrow interval, but that's because we used 500 people. Realistically, it is far more common to use a much smaller sample size, which we will explore in the next box.

Play Around With it...

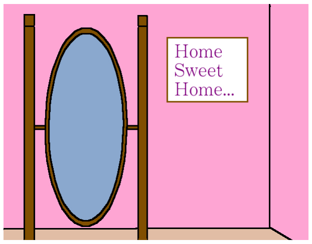
\# 3-5-23

Let's repeat the example of the previous box, but with $n = 100$, which is a far more realistic sample size. Use the formula

$$p + 2\sqrt{\frac{pq}{n}} > \hat{p} > p - 2\sqrt{\frac{pq}{n}}$$

to compute the percentage of the entire district that is in favor of the proposal, with a 95.45% margin of error. (Hint: the words "95.45% margin of error" just mean that we are using the two-sigma interval, nothing more.)

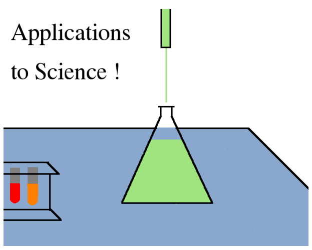
[Answer: $64.9498\cdots\% > \hat{p} > 45.0501\cdots\%$.]



A Pause for Reflection...
Let's look at that interval from the previous box, and analyze its political consequences.

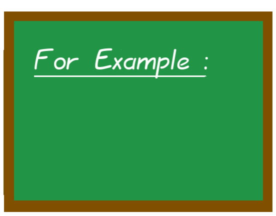
- At the high end, 64.94%, someone might think that the proposal will glide through the election, and possibly cause a nearly 2:1 outcome. (Note, 66.6667% would be exactly a 2:1 outcome.) Someone might imagine that the proposal has enormous momentum behind it, and that there is no reason to keep campaigning for it.
- At the low end, 45.05%, someone might think that the majority is against the proposal, even though that's the opposite of the truth. They might think that the proposal is unlikely to pass, and stop campaigning for that reason.

What it really comes down to is that surveys and polls are really much less effective research tools than most of us would imagine. There is too much uncertainty.



Let's return to the problem about the contaminated fish from Page 353. In particular, the environmental firm draws 400 fish from the lake. With probability 68.72%, they will detect between 34 to 46 contaminated fish, according to the one-sigma version of the "square root of *npq* rule." However, according to the two-sigma version of the "square root of *npq* rule," they will find that 28 to 52 fish are contaminated. That comes to between 7% and 13% of the 400-fish sample, because $28/400 = 0.07$ and $52/400 = 0.13$.

In the next box, we will compute what our shortcut formula would predict in this situation.



\# 3-5-24

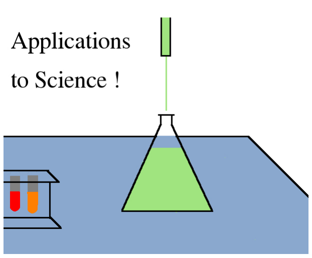
Let's see what the formula

$$p + 2\sqrt{\frac{pq}{n}} > \hat{p} > p - 2\sqrt{\frac{pq}{n}}$$

will tell us about $\hat{p}$. We have $p = 0.1$, $q = 0.9$, and $n = 400$. We obtain

$$\begin{array}{rcccl} 0.1 + 2\sqrt{\frac{0.1(0.9)}{400}} & > & \hat{p} & > & 0.1 + 2\sqrt{\frac{0.1(0.9)}{400}} \\ 0.1 + 2\sqrt{0.000225} & > & \hat{p} & > & 0.1 + 2\sqrt{0.000225} \\ 0.1 + 2(0.015) & > & \hat{p} & > & 0.1 - 2(0.015) \\ 0.1 + 0.03 & > & \hat{p} & > & 0.1 - 0.03 \\ 0.13 & > & \hat{p} & > & 0.07 \end{array}$$

As you can see, it is pretty neat that we can jump directly to the answer, skipping several intermediate steps. This is a useful formula. Moreover, the range of $\hat{p}$ is rather large. After all 13% is just slightly less than double 7%.



I'm going to talk a bit about realism now. Because it was a problem in this textbook, you and I knew that 10% of the fish (in the problem of the previous box) were contaminated. However, the environmental firm hired by the fishery would not have known that at the start. Similarly, with the laptop-keyboard problem, you and I knew that 5% of the keyboards were defective, but the factory can't possibly know that before beginning their quality-control checks. Yet a third time, you and I knew that the underlying population of the town was 80% in favor of John Q. Public, but the polling firm doesn't know that before they start their poll.

Now that the textbook problems are finished, let me tell you how it works in reality.

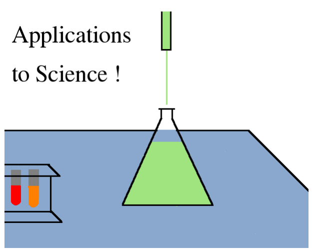
There is a little bait-and-switch. This slight of hand has been done by statisticians across the world, for over a century, and is used in disciplines as far apart as biology, political science, and quality engineering. In place of

$$p + 2\sqrt{\frac{p(1-p)}{n}} > \hat{p} > p - 2\sqrt{\frac{p(1-p)}{n}}$$

we're going to write

$$\hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} > p > \hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

instead. The first inequality is one that we've worked with several times, except we replaced the q with $1 - p$. To reach the second inequality, we interchange p and $\hat{p}$. I'm going to tell you how this is used now, and you'll see from that why it is done.



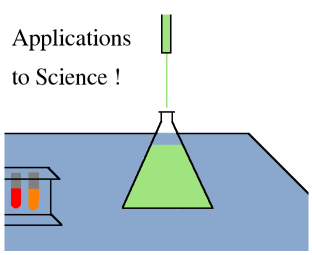
The environmental firm comes to the fishery, and they pull out 400 fish, testing them for contamination. Let's suppose that 44 fish turn out to be contaminated. Then the estimate is $44/400 = 0.11 = \hat{p}$. This means that their best guess is 11%. Next, they will compute

$$0.11 + 2\sqrt{\frac{0.11(1-0.11)}{400}} > p > 0.11 - 2\sqrt{\frac{0.11(1-0.11)}{400}}$$

which simplifies to

$$0.141288\cdots > p > 0.0787110\cdots$$

causing them to tell the boss that, with 95.45% confidence, the percentage of contaminated fish is between 7.87% and 14.13%. You and I know that the true rate is 10%, so we are not surprised by this announcement.



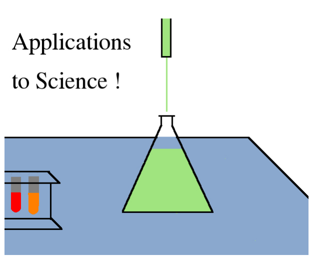
Similarly, the laptop-keyboard factory will test 1000 keyboards. Let's suppose that they find 43 of them are faulty. Then the estimate is $43/1000 = 0.043 = \hat{p}$. This means that their best guess is 4.3%. Next, they will compute

$$0.043 + 2\sqrt{\frac{0.043(1-0.43)}{1000}} > p > \hat{p} - 2\sqrt{\frac{0.043(1-0.43)}{1000}}$$

which simplifies to

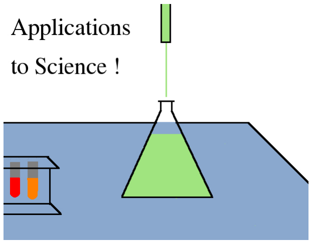
$$0.0743113\cdots > p > 0.0116886\cdots$$

causing them to tell the boss that, with 95.45% confidence, the percentage of flawed keyboards is between 1.17% and 7.43%. You and I know that the true rate is 5%, so we are not surprised by this announcement.



The previous four boxes are a tiny taste of the concept of a confidence interval. You can already see that confidence intervals are tremendously important for communicating the accuracy (or lack thereof) that occurs when making experimental measurements. Often, a proper single-semester university-level statistics course will end with confidence intervals as the capstone topic, or maybe second-to-last topic. When you take your statistics course (which might be in the past for some of my readers, but probably in the future for most of my readers), then be sure to wake up and play close attention to the chapter about confidence intervals.

Out of everything in the entire statistics course, it is one of the most useful things that you will learn.



To recap the last five boxes, let's be honest with each other.

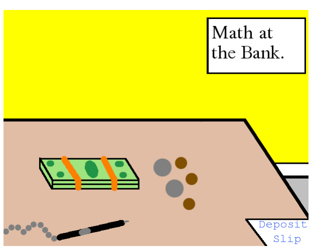
Surely it is better to say "We detect that 4.30% of the keyboards in our sample of 1000 are flawed, but with 95.45% confidence we can only conclude that the underlying rate of flawed keyboards is between 1.17% and 7.43%" than it is to say "We detect that roughly 4.3% of the keyboards are flawed." (Especially because the truth, known to you and I but not the factory, is that 5% are flawed.)

Similarly, it is better to say "We detect that 11.00% of the fish in our sample of 400 are contaminated, but with 95.45% confidence we can only conclude that the underlying rate of contaminated fish is between 7.87% and 14.13%" than it is to say "We detect that roughly 11% of the fish are contaminated." (Especially because the truth, known to you and I but not the environmental firm, is that 10% are contaminated.)

It is even more honest to say "between 1.17%–7.43% of keyboards" and "between 7.87%–14.13% of fish," not mentioning the 4.30% and the 11.00%.

The previous box demonstrates two examples of a "confidence interval." Computing confidence intervals is often the capstone concept of a university-level statistics course, or very close to the end. Scientists use confidence intervals in scientific papers very often, and engineers do often enough, but probably should use them more often. For example, when I was a graduate student in computer engineering (2001–2004) many papers about computer architecture did not include confidence intervals for their results, but should have.

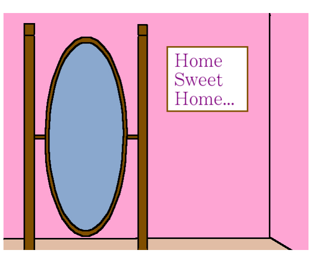
Once I was the anonymous peer-reviewer for a paper in cryptanalysis, and I rejected the paper for publication because they gave only averages, without going further to express the uncertainty of their results.



When proofreading the last few boxes, one of my proofreaders (Russell Chamberlain) made an interesting comment. He has his *Professional Science Masters in Industrial & Applied Mathematics* from the department where I teach, at the University of Wisconsin—Stout, and he works at Thrivent Financial.

"Just to add on to this point, I use confidence intervals daily at work."

Indeed, I have often stated in the classroom that point estimation (giving a single number as an approximation) is the work of amateurs and semi-professionals. Interval estimation, such as saying "between 1.17%–7.43% of keyboards are defective" is the work of a professional.

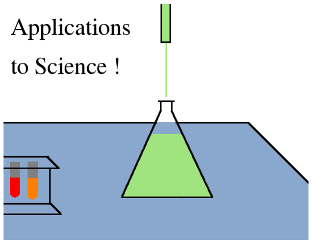


A Pause for Reflection. . .

Clearly, these ideas are also very important for sociologists and political scientists, and those who get PhDs in political science are forced to take a course in these matters.

However, I do not believe that undergraduates from that discipline are genuinely forced to *master* topics such as these—though hopefully they are exposed to these ideas. It is a touchy point, because many undergraduates are drawn to the social sciences because the math prerequisites are considerably lower than in STEM subjects. Therefore, when faculty from mathematics or statistics bring up mathematical/statistical illiteracy in conversation with faculty from the social sciences, the latter have a lot of fear. In my experience, they worry that they'll lose the majority of their students if the standards are raised.

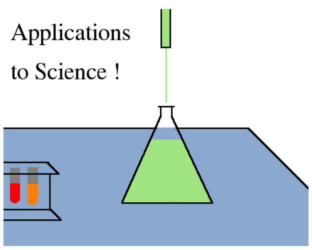
I'd like to mention one bold exception: at American University, in Washington DC, they require such a course for undergraduates in political science. It is GOVT-310: *Introduction to Political Research* or STAT-202: *Basic Statistics*. (Checked on February 6th, 2018.)



One of the cornerstone tools of archeology, geology, and evolutionary biology is radio-carbon dating. There are other forms of dating objects, animals, and people, but radio-carbon dating is the most familiar to most of us. When people are alive, they are exchanging carbon with their environment, as they eat and drink, as well as exhale, urinate, and defecate. But when death occurs, there is no more eating, drinking, exhaling, et cetera…

While alive, the ratio between Carbon-14 and Carbon-12 will be very close to those of the surroundings. However, upon death, no new carbon is coming in. Some Carbon-14 will decay, unbelievably slowly, and it simply won't be replaced.

That process of radioactive decay follows mathematical laws that you have no doubt been shown when you learned about exponential and logarithmic functions. A measurement of the ratio of Carbon-14 to Carbon-12 in a human or animal corpse, or in an object made of wood, can be used mathematically to determine how long it has been since that human or animal died, or the tree was cut down.

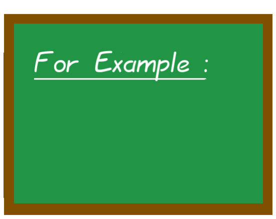


Continuing with the previous box, all this is no doubt familiar to you. You might be wondering what it has to do with probability.

In reality, in each slice of time, be it a century, a year, a month, a day, or an hour, there is some probability that a Carbon-14 atom will decay. If the time-slice is 5730 years, that probability is $1/2$, and that's why we say, "The Half Life of Carbon-14 is 5730 years."

If the time-slice is 1 year, that probability is $0.000125889\cdots$. Similarly, if the time-slice is 1 month, that probability is $0.0000104908\cdots$. (These two numbers were computed with Poisson's Rare Events Theorem, and I don't have time to go into that just at the moment.) Therefore, given that the probabilities are so low, how do we know that probability will not ruin, or add great uncertainty to, any radio-carbon dating measurement?

I will explain with a concrete example in the next box.



3-5-25

Suppose that the femur of some caveman has been found, and it is found to have 200 grams of carbon. Further suppose that, at the moment of death, 1.2×10^{-12} of that carbon is Carbon-14, because that's the environmental density of Carbon-14 among all carbon. (A chemist would say 1.2 "parts per trillion.") So that's 2.4×10^{-10} grams of Carbon-14.

As it turns out, the mass of a Carbon-14 atom is $2.32534\cdots \times 10^{-23}$ grams/atom. This means there are

$$\frac{2.4 \times 10^{-10} \text{ grams}}{2.32534\cdots \times 10^{-23} \text{ grams/atom}} = 1.03210\cdots \times 10^{13} \text{ atoms}$$

which is pretty amazing. After all, only 1.2 atoms per trillion are Carbon-14 (all the others are Carbon-13 and Carbon-12), and we're talking about only 200 grams, yet we have 10.321 trillion atoms of Carbon-14.

Now we know we have $n = 1.03210\cdots \times 10^{13}$ atoms, and the probability that a Carbon-14 atom has not decayed is $p = 1/4$. Let's use a 3-sigma (99.73%) interval to determine how many atoms of Carbon-14 are left. We'll do that in the next box.

Continuing with the previous box, we obtain the following:

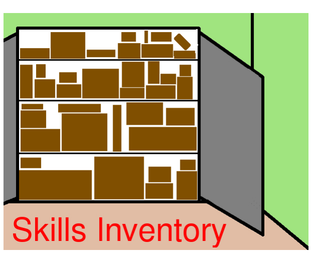
$$np - 3\sqrt{npq} = (1.03210 \times 10^{13})(1/4) - 3\sqrt{(1.03210 \times 10^{13})(1/4)(3/4)} = 2.58025 \times 10^{12} - 1.39111 \times 10^{6} = 2.58024 \times 10^{12}$$
$$np + 3\sqrt{npq} = (1.03210 \times 10^{13})(1/4) + 3\sqrt{(1.03210 \times 10^{13})(1/4)(3/4)} = 2.58025 \times 10^{12} + 1.39111 \times 10^{6} = 2.58025 \times 10^{12}$$

As you can see, to six significant figures, we cannot measure this uncertainty. The uncertainty is so small, that it cannot be observed in five decimal places, and barely in six.

After all, if I told you that a mutual fund had 2.58025×10^{12} cents in it, with an uncertainty of 1.39111×10^{6} cents, that would be 25.8025 billion dollars with an uncertainty of \$ 13,911. That's rather certain.

In summary, because *there are just so many atoms* even in a small sample, the $3\sqrt{npq}$ becomes very tiny compared to the np, and there is very little uncertainty.

The previous example was recommended to the author by Dr. Gabriel Hanna, a physicist who now works in the health insurance industry.



Here is a summary of what we have learned in this module.

- We saw several definitions of "the square root of npq rule."
- We applied "the square root of npq rule" to give us intervals that tell us how often something is actually going to happen, based on the probability p and the number of attempts n.
- We learned the special rounding technique used for such intervals.
- We learned how to check our work, by averaging the endpoints of the intervals, and comparing that to the expected value.
- We applied these techniques to a wide variety of problems, including environmental, political, medical, and industrial situations.
- We learned why we should never throw small electronic devices into the trash.
- We learned about polling error, and how it can really damage political science research.
- We learned about how the "Law of Large Numbers" affects factory investigations of defects, and we also explored the "Law of Large Numbers" further, using some Sage code.
- We learned where the numbers 68.27%, 95.45%, and 99.73% come from—they come from some fairly cool integrals.
- We analyzed the profitability of a casino's Roulette tables.

This module is now complete. Thank you for reading. What follows are the solutions to various problems from inside the module.

On Page 353, I asked you to try to figure out the solution to the following question: "For what kinds of probabilities is the uncertainty in "the square-root npq rule" the largest? the smallest?"

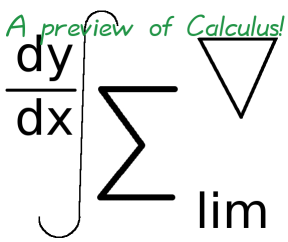
As I mentioned at that time, by "uncertainty," I mean $2\sqrt{npq}$ (or $3\sqrt{npq}$). As it turns out, the 2 or the 3 won't matter, so let's go with the 2. We are trying to find the minima and maxima of the function

$$f(p) = 2\sqrt{npq}$$

which is just an easy first-semester calculus problem.

To find the minima and maxima of a function over a closed interval (in this case, the interval is $0 \leq p \leq 1$), we all know what to do. We have to take the first derivative, and then the set of candidates will be the end points of the interval, the points where the first derivative is undefined, and the points where the first derivative is zero.

We'll do this in the next box.



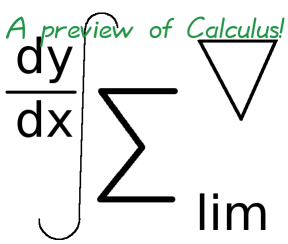
Before we start, we need to make our function a single-variable function by plugging in $q = 1 - p$. Then we have

$$f(p) = 2\sqrt{npq} = 2\sqrt{np(1-p)} = 2\left(np(1-p)\right)^{1/2} = 2\left(np - np^2\right)^{1/2}$$

and it is easy to take the derivative of that

$$\begin{aligned} f'(p) &= \\ &= 2(1/2)\left(np - np^2\right)^{1/2-1}\frac{d}{dp}\left(np - np^2\right) \\ &= 2(1/2)\left(np - np^2\right)^{-1/2}(n1 - 2np) \\ &= \left(np - np^2\right)^{-1/2}(n - 2np) \\ &= \frac{n - 2np}{\sqrt{np - np^2}} \end{aligned}$$

First, the endpoints of the interval are $p = 0$ and $p = 1$. We will continue in the next box.



Second, $f'(p)$ will be undefined if $f'(p)$ attempts to divide by zero, or if the junk under the square root sign is negative. Dividing by zero will occur if $np - np^2 = 0$. We can divide both sides by n, and we get $p - p^2 = 0$, and that factors into $p(1-p) = 0$. That means $p = 0$ or $1 - p = 0$. The latter is just $p = 1$. (Of course, we have tacitly assumed $n \neq 0$, but since it is the number of times that you are trying something, then that's going to be a positive integer, not zero.)

As it turns out, $np - np^2$ is never negative for $0 \leq p \leq 1$ and $n > 0$, but it would be tedious to investigate that now.

Third, $f'(p)$ will equal zero if and only if its numerator is zero while the denominator is nonzero. When is $n - 2np = 0$? We divide both sides by n and get $1 - 2p = 0$, which means $1 = 2p$ and therefore $p = 1/2$. As a mathematics professor, I feel obligated to check that the denominator is not zero. Actually, the denominator turns out to be

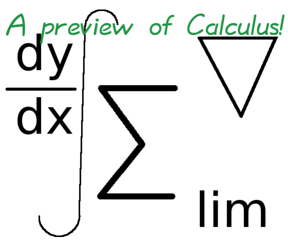
$$\sqrt{np - np^2} = \sqrt{n(1/2) - n(1/2)^2} = \sqrt{n(1/2) - n(1/4)} = \sqrt{n(1/2 - 1/4)} = \sqrt{n(1/4)} = \sqrt{n/4}$$

which is not zero.

Looking at the previous box, our candidates are $\{0, 1/2, 1\}$.

$$\begin{aligned} f(0) &= 2\sqrt{np(1-p)} = 2\sqrt{n(0)(1-0)} = 2\sqrt{0} = 0 \\ f(1/2) &= 2\sqrt{n(1/2)(1-1/2)} = 2\sqrt{n(1/2)(1/2)} = 2\sqrt{n(1/4)} = 2\sqrt{n/4} \\ f(1) &= 2\sqrt{n(1)(1-1)} = 2\sqrt{n(1)(0)} = 2\sqrt{n(0)} = 2\sqrt{0} = 0 \end{aligned}$$

Now, we have our answers. The function $f(p)$ takes its maximum at $p = 1/2$ where $f(1/2) = 2\sqrt{n/4}$, and its minima are at $p = 1$ and $p = 0$, both of which make $f(p) = 0$. The translation into ordinary english is that the most uncertain type of probability event is a 50-50 chance ($p = 1/2$), but the least uncertain type of probability event is something that is either certain to happen ($p = 1$), or certain not to happen ($p = 0$).



The reader who is very good at calculus might have realized that I could have solved the problem in an easier way. The key realization is that $\sqrt{np(1-p)}$ will have its minima and maxima at the same places as $np(1-p)$.

This often comes up in related-rates problems in a first semester calculus course. It is easier to find the minimum or maximum of the square of the distance, than it is to find the minimum or maximum of the distance itself. That's because the distance formula contains a square root, and the derivative gets a little nasty.

However, most of my students would not have noticed this immediately, so I thought I should provide the longer solution.