CONTENTS

1. THE GAMMA FUNCTION

In this chapter we'll explore some of the strange and wonderful properties of the Gamma function $\Gamma(s)$, defined by

For $s > 0$ (or actually $\Re(s) > 0$), the **Gamma function** $\Gamma(s)$ is

$$\Gamma(s) \;=\; \int_0^\infty e^{-x}x^{s-1}dx \;=\; \int_0^\infty e^{-x}\frac{dx}{x}.$$

There are countless integrals or functions we can define. Just looking at it, there is nothing to make you think it will appear throughout probability and statistics, but it does. We'll see where it occurs and why, and discuss many of its most important properties.

1.1. Existence of $\Gamma(s)$.

Looking at the definition of $\Gamma(s)$, it's natural to ask: *Why do we have restrictions on s?* Whenever you are given an integrand, you must make sure it is well-behaved before you can conclude the integral exists. Frequently there are two trouble points to check, near $x = 0$ and near $x = \pm\infty$ (okay, three points). For example, consider the function $f(x) = x^{-1/2}$ on the interval $[0,\infty)$. This function blows up at the origin, but only mildly. Its integral is $2x^{1/2}$, and this is integrable near the origin. This just means that

$$\lim_{\epsilon\to 0}\int_\epsilon^1 x^{-1/2}dx$$

exists and is finite. Unfortunately, even though this function is tending to zero, it approaches zero so slowly for large x that it is not integrable on $[0,\infty)$. The problem is that integrals such as

$$\lim_{B\to\infty}\int_1^B x^{-1/2}dx$$

is infinite. Can the reverse problem happen, namely our function decays fast enough for large x but blows up too rapidly for small x? Sure – the following is a standard, albeit initially strange looking, example. Consider

$$f(x) \;=\; \begin{cases} \frac{1}{x\log^2 x} & \text{if } x > 0 \\ 0 & \text{otherwise.}\end{cases}$$

Our function has a nice integral:

$$\int \frac{1}{x\log^2 x}dx \;=\; \int (\log x)^{-2}\frac{1}{x}dx \;=\; \int (\log x)^{-2} d\log x \;=\; -(\log x)^{-1}.$$

We check the two limits:

$$\lim_{B\to\infty}\int_1^B \frac{dx}{x\log^2 x} \;=\; \lim_{B\to\infty}\left(-\frac{1}{\log x}\right)\Bigg|_1^B \;=\; -\lim_{B\to\infty}\frac{1}{\log B} \;=\; 0.$$

What about the second limit? We have

$$\lim_{\epsilon\to 0}\int_\epsilon^1 \frac{1}{x\log^2 x}dx \;=\; \lim_{\epsilon\to 0}\left(-\frac{1}{\log x}\right)\Bigg|_\epsilon^1 \;=\; \lim_{\epsilon\to 0}\frac{1}{\log\epsilon} \;=\; -\infty$$

(to see the last, write ϵ as $1/n$ and send $n \to \infty$).

So it's possible for a positive function to fail to be integrable because it decays too slowly for large x, or it blows up too rapidly for small x. As a rule of thumb, if as $x \to \infty$ a function is decaying faster than $1/x^{1+\epsilon}$ for any epsilon, then the integral at infinity will be finite. For small x, if as $x \to 0$ the function is blowing up slower than $x^{-1+\epsilon}$ then the integral at 0 will be okay near zero. You should always do tests like this, and get a sense for when things will exist and be well-defined.

Returning to the Gamma function, let's make sure it's well-defined for any $s > 0$. The integrand is $e^{-x}x^{s-1}$. As $x \to \infty$, the factor x^{s-1} is growing polynomially but the term e^{-x} is decaying exponentially, and thus their product decays rapidly. If we want to be a bit more careful and rigorous, we can argue as follows: choose some integer $M > s + 1701$ (we put in a large number to alert you to the fact that the actual value of our number does not matter). We clearly have $e^x > x^M/M!$, as this is just one term in the Taylor series expansion of e^x. Thus $e^{-x} < M!/x^M$, and the integral for large x is finite and well-behaved, as it's bounded by

$$\begin{aligned}
\int_1^B e^{-x}x^{s-1}dx \;&\le\; \int_1^B M!x^{-M}x^{s-1}dx \\
&=\; \int M!\int_1^B x^{s-M-1} \\
&=\; M!\frac{x^{s-M}}{s-M}\Bigg|_1^B \\
&=\; \frac{M!}{s-M}\left[\frac{1}{B^{M-s}} - 1\right].
\end{aligned}$$

It was very reasonable to try this. We know the e^x grows very rapidly, so e^{-x} decays quickly. We need to borrow some of the decay from e^{-x} to handle the x^{s-1} piece.

What about the other issue, near $x = 0$? Well, near $x = 0$ the function e^{-x} is bounded; it's largest value is when $x = 0$ so it is at most 1. Thus

$$\begin{aligned}
\int_0^1 e^{-x}x^{s-1}dx \;&\le\; \int_0^1 1\cdot x^{s-1}dx \\
&=\; \frac{x^s}{s}\Bigg|_0^1 \;=\; \frac{1}{s}.
\end{aligned}$$

We've shown everything is fine for $s > 0$; what if $s \le 0$? Could these values be permissible as well? The same type of argument as above shows that there are no problems when x is large. Unfortunately, it's a different story for small x. For $x \le 1$ we clearly have $e^{-x} \ge 1/e$. Thus our integrand is at least as large as x^{s-1}/e. If $s \le 0$, this is no longer integrable on $[0, 1]$. For definiteness, let's do $s = -2$ Then we have

$$\int_0^\infty e^{-x} x^{-3} dx \quad \ge \quad \int_0^\infty \frac{1}{e}\, x^{-3} dx$$
$$= \quad \frac{1}{e} x^{-2} \Big|_0^1,$$

and this blows up.

The arguments above can (and should!) be used every time you meet an integral. Even though our analysis hasn't suggested a reason why anyone would *care* about the Gamma function, we at least know that it is well-defined and exists for all $s > 0$. In the next section we'll show how to make sense of Gamma for all values of s. This should be a bit alarming – we've just spent this section talking about being careful and making sure we only use integrals where they are well-defined, and now we want to talk about putting in values such as $s = -3/2$? Obviously, whatever we do, it won't be anything as simple as just plugging $s = -3/2$ into the formula.

If you're interested, $\Gamma(-3/2) = 4\sqrt{\pi}/3$ – we'll prove this soon!

1.2. **The Functional Equation of** $\Gamma(s)$**.** We turn to one of the most important property of $\Gamma(s)$. In fact, this property allows us to make sense of *any* value of s as input, such as the $s = -3/2$ of the last section. Obviously this can't mean just naively throwing in any s in the definition, though many good mathematicians have accidentally done so. What we're going to see is the the **Analytic (or Meromorphic) Continuation**.. The gist of this is that we can take a function f that makes sense in one region and extend its definition to a function g defined on a larger region in such a way that our new function g agrees with f where they are both defined, but g is defined for more points.

The following absurdity is a great example. What is

$$1 \;+\; 2 \;+\; 4 \;+\; 8 \;+\; 16 \;+\; 32 \;+\; 64 \;+\; \cdots ?$$

Well, we're adding all the powers of 2, thus it's clearly zero, right? Wrong – the "natural" meaning for this sum is -1! A sum of infinitely many positive terms is negative? What's going on here?

This example comes from something you've probably seen many times, the geometric series. If we take the sum

$$1 \;+\; r \;+\; r^2 \;+\; r^3 \;+\; r^4 \;+\; r^5 \;+\; r^6 \;+\cdots$$

then, *so long as* $|r| < 1$, the sum is just $\frac{1}{1-r}$. There are many ways to see this. **ADD REF TO THE SHOOTING GAME**. The most common, as well as one of the most boring, is to let

$$S_n \;=\; 1 + r + \cdots + r^n.$$

If we look at $S_n - rS_n$, almost all the terms cancels; we're left with

$$S_n - rS_n \;=\; 1 - r^{n+1}.$$

We factor the left hand side as $(1 - r)S_n$, and then dividing both sides by $1 - r$ gives

$$S_n \;=\; \frac{1 - r^{n+1}}{1 - r}.$$

If $|r| < 1$ then $\lim_{n\to\infty} r^n = 0$, and thus taking limits gives

$$\sum_{m=0}^{\infty} r^m \;=\; \lim_{n\to\infty} S_n \;=\; \lim_{n\to\infty} \frac{1 - r^{n+1}}{1 - r} \;=\; \frac{1}{1 - r}.$$

This is known as the **geometric series formula**, and is used in a variety of problems.

Let's rewrite the above. The summation notation is nice and compact, but that's not what we want right now – we want to really see what's going on. We have

$$1 \;+\; r \;+\; r^2 \;+\; r^3 \;+\; r^4 \;+\; r^5 \;+\; r^6 \;+\cdots \;=\; \frac{1}{1 - r}, \quad |r| < 1.$$

Note the left hand side makes sense only for $|r| < 1$, but the right hand side makes sense for *all* values of r other than 1! We say the right hand side is an analytic continuation of the left, with a pole at $s = 1$ (poles are where our functions blow-up).

Let's define the function

$$f(x) \;=\; 1 \,+\, x \,+\, x^2 \,+\, x^3 \,+\, x^4 \,+\, x^5 \,+\, +x^6 \,+\, \cdots.$$

For $|x| < 1$ we also have

$$f(x) \;=\; \frac{1}{1-x}.$$

And now the big question: what is $f(2)$? If we use the second definition, it's just $\frac{1}{1-2} = -1$, while if we use the first definition its that strange sum of all the powers of 2. *THIS* is the sense in which we mean the sum of all the powers of 2 is -1. We do not mean plugging in 2 for the series expansion; instead, we evaluate the extended function at 2.

It's now time to apply these techniques to the Gamma function. We'll show, using integration by parts, that Gamma can be extended for all s (or at least for all s except the negative integers and zero). Before doing the general case, let's do a few representative examples to see why integration by parts is such a good thing to do. Recall

$$\Gamma(s) \;=\; \int_0^\infty e^{-x} x^{s-1} dx, \quad s > 0.$$

The easiest value of s to take is $s = 1$, as then the x^{s-1} term becomes the harmless $x^0 = 1$. In this case, we have

$$\Gamma(1) \;=\; \int_0^\infty e^{-x} dx \;=\; -e^{-x}\Big|_0^\infty \;=\; -0 + 1 \;=\; 1.$$

Building on our success, what is the next easiest value of s to take? A little experimentation suggests we try $s = 2$. This will make x^{s-1} equal x, a nice, integer power. We find

$$\Gamma(2) \;=\; \int_0^\infty e^{-x} x dx.$$

Now we can begin to see why integration by parts will play such an important role. If we let $u = x$ and $dv = e^{-x}dx$, then $du = dx$ and $v = -e^{-x}$, then we'll see great progress – we start with needing to integrate xe^{-x} and after integration by parts we're left with having to do e^{-x}, a wonderful savings. Putting in the details, we find

$$\begin{aligned} \Gamma(2) \;&=\; uv\Big|_0^\infty - \int_0^\infty v du \\ &=\; -xe^{-x}\Big|_0^\infty + \int_0^\infty e^{-x} dx. \end{aligned}$$

The boundary term vanishes (it's clearly zero at zero; use L'Hopital to evaluate it at ∞, giving $\lim_{x\to\infty} \frac{x}{e^x} = \lim_{x\to\infty} \frac{1}{e^x} = 0$), while the other integral is just $\Gamma(1)$. We've thus shown that

$$\Gamma(2) \;=\; \Gamma(1);$$

however, it is more enlightening to write this in a slightly different way. We took $u = x$ and then said $du = dx$; let's write it as $u = x^1$ and $du = 1dx$. This leads us to

$$\Gamma(2) \;=\; 1\Gamma(1).$$

At this point you should be skeptical – does it really matter? Anything times 1 is just itself! It does matter. If we were to calculate $\Gamma(3)$, we would find it equals $2\Gamma(2)$, and if we then progressed to $\Gamma(4)$ we would see it's just $3\Gamma(3)$. This pattern suggests $\Gamma(s+1) = s\Gamma(s)$, which we now prove.

We have

$$\Gamma(s+1) \;=\; \int_0^\infty e^{-x}x^{s+1-1}dx \;=\; \int_0^\infty e^{-x}x^s dx.$$

We now integrate by parts. Let $u = x^s$ and $dv = e^{-x}$; we're basically forced to do it this way as e^{-x} has a nice integral, and by setting $u = x^s$ when we differentiate the power of our polynomial goes down, leading to a simpler integral. We thus have

$$u \;=\; x^s, \quad du \;=\; sx^{s-1}dx, \quad dv \;=\; e^{-x}dx, \quad v \;=\; -e^{-x},$$

which gives

$$\begin{aligned}\Gamma(s+1) &= -x^s e^{-x}\Big|_0^\infty + \int_0^\infty e^{-x}sx^{s-1}dx \\ &= 0 + s\int_0^\infty e^{-x}x^{s-1}dx \;=\; s\Gamma(s),\end{aligned}$$

completing the proof. This relation is so important its worth isolating it, and giving it a name.

Functional equation of $\Gamma(s)$: The Gamma function satisfies

$$\Gamma(s+1) \;=\; s\Gamma(s).$$

This allows us to extend the Gamma function to all s. We call the extension the Gamma function as well, and it is well-defined and finite for all s save the negative integers and zero.

Let's return to the example from the previous section. Later we'll prove that $\Gamma(1/2) = \sqrt{\pi}$. For now we assume we know this, and show how we can figure out what $\Gamma(-3/2)$ should be. From the functional equation, $\Gamma(s+1) = s\Gamma(s)$. We can rewrite this as $\Gamma(s) = s^{-1}\Gamma(s+1)$, and we can now use this to 'walk up' from $s = -3/2$, where we don't know the value, to $s = 1/2$, where we assume we do. We have

$$\Gamma\left(-\frac{3}{2}\right) \;=\; -\frac{2}{3}\Gamma\left(-\frac{1}{2}\right) \;=\; -\frac{2}{3}\cdot(-2)\,\Gamma\left(\frac{1}{2}\right) \;=\; \frac{4\sqrt{\pi}}{3}.$$

This is the power of the functional equation – it allows us to define the Gamma function essentially everywhere, so long as we know its values for $s > 0$. Why are zero and the negative integers special? Well, let's look at $\Gamma(0)$:

$$\Gamma(0) \;=\; \int_0^\infty e^{-x}x^{0-1}dx \;=\; \int_0^\infty e^{-x}x^{-1}dx.$$

The problem is that this is not integrable. While it decays very rapidly for large x, for small x it looks like $1/x$. The details are:

$$\lim_{\epsilon\to 0}\int_\epsilon^1 e^{-x}x^{-1}dx \;\geq\; \frac{1}{e}\lim_{\epsilon\to 0}\int_\epsilon^1 \frac{dx}{x} \;=\; \frac{1}{e}\lim_{\epsilon\to 0}\log x\Big|_\epsilon^1 \;=\; \lim_{\epsilon\to 0} -\log\epsilon \;=\; \infty.$$

Thus $\Gamma(0)$ is undefined, and hence by the functional equation it is also undefined for all the negative integers.

1.3. **The Factorial Function and $\Gamma(s)$.** In the last section we showed that $\Gamma(s)$ satisfies the functional equation $\Gamma(s+1) = s\Gamma(s)$. This is reminiscent of a relation obeyed by a better known function, the factorial function. Remember

$$n! \;=\; n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1;$$

we write this in a more suggestive way as

$$n! \;=\; n\cdot(n-1)!.$$

Note how similar this looks to the relationship satisfied by $\Gamma(s)$. It's not a coincidence – the Gamma function is a generalization of the factorial function!

We've shown that $\Gamma(1) = 1$, $\Gamma(2) = 1$, $\Gamma(3) = 2$, and so on. We can interpret this as $\Gamma(n) = (n-1)!$ for $n \in \{1, 2, 3\}$; however, applying the functional equation allows us to extend this equality to all n. We proceed by induction. Proofs by induction have two steps, the base case (where you show it holds in some special instance) and the inductive step (where you assume it holds for n and then show that implies it holds for $n + 1$).

We've already done the base case, as we've checked $\Gamma(1) = 0!$. We checked a few more cases then we needed to. Typically that's a good strategy when doing inductive proofs. By getting your hands dirty and working out a few cases in detail, you often get a better sense of what's going on, and you can see the pattern. Remember, we initially wrote $\Gamma(2) = \Gamma(1)$, but after some thought (as well as years of experience) we rewrote it as $\Gamma(2) = 1 \cdot \Gamma(1)$.

We now turn to the inductive step. We assume $\Gamma(n) = (n-1)!$, and we must show $\Gamma(n+1) = n!$. From the functional equation, $\Gamma(n + 1) = n\Gamma(n)$; but by the inductive step $\Gamma(n) = (n - 1)!$. Combining gives $\Gamma(n + 1) = n(n - 1)!$, which is just $(n + 1)!$, or what we needed to show. This completes the proof. $\Box$

We now have two different ways to calculate say 1020!. The first is to do the multiplications out: $1020 \cdot 1019 \cdot 1018 \cdots$. The second is to look at the corresponding integral:

$$1020! \ = \ \Gamma(1021) \ = \ \int_0^\infty e^{-x} x^{1020} dx.$$

There are advantages to both methods; we wanted to discuss some of the benefits of the integral approach, as this is definitely not what most people have seen. Integration is hard; most people don't see it until late in high school or college. We all know how to multiply numbers – we'be been doing this since grade school. Thus, why make our lives difficult by converting a simple multiplication problem to an integral?

The reason is a general principle of mathematics – often by looking at things in a different way, from a higher level, new features emerge that you can exploit. Also, once we write it as an integral we have a lot more tools in our arsenal; we can use results from integration theory and from analysis to study this. We do this in Chapter 2, and see just how much we can learn about the factorial function by recasting it as an integral.

1.4. **Special Values of $\Gamma(s)$.** We know that $\Gamma(n+1) = n!$ whenever n is a non-negative integer. Are there choices of s that are important, and if so, what are they? In other words, we've just generalized the factorial function. What was the point? It may be that the non-integral values are just curiosities that don't really matter, and the entire point might be to have the tools of calculus and analysis available to study $n!$. This, however, is most emphatically *not* the case. Some of these other values are very important in probability; in a bit of foreshadowing, we'll say they play a central role in the subject.

So, what are the most important Because of the functional equation, once we know $\Gamma(1)$ we know the Gamma function at all non-negative integers, which gives us all the factorials. So 1 is an important choice of s. We'll now see that $s = 1/2$ is also very important.

One of the most important, if not the most important, distribution is the normal distribution. We say X is normally distributed with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, if the density function is

$$f_{\mu,\sigma}(x) \ = \ \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}.$$

Looking at this density, we see there are two parts. There's the exponential part, and the constant factor of $1/\sqrt{2\pi\sigma^2}$. Because the exponential function decays so rapidly, the integral will be finite and thus, if appropriately normalized, we will have a probability density. The hard part is determining just what this integral is. Another way of formulating this question is: Let $g(x) = e^{-(x-\mu)^2/2\sigma^2}$. As it decays rapidly and is never negative, it can be rescaled to integrate to one and hence become a probability density. That scale factor is just $1/c$, where

$$c \;=\; \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}.$$

In Chapter **ADD REF** we'll see numerous applications and uses of the normal distribution. It's not hard to make an argument that it is important, and thus we *need* to know the value of this integral. That said, why is this in the Gamma function chapter?

The reason is that, with a little bit of algebra and some change of variables, we'll see that this integral is just $\sqrt{2}\Gamma(1/2)$. We might as well assume $\mu = 0$ and $\sigma = 1$ (if not, then step 1 is just to change variables and let $t = \frac{x-\mu}{\sigma}$. So let's look at

$$\begin{aligned} I \;&:=\; \int_{-\infty}^{\infty} e^{-x^2/2} dx \\ &=\; 2\int_{0}^{\infty} e^{-x^2/2} dx. \end{aligned}$$

This only vaguely looks related to the Gamma function. The Gamma function is the integral of e^{-x} times a polynomial in x, while here we have the exponential of $-x^2/2$. Looking at this, we see that there's a natural change of variable to try to make our integral look like the Gamma function at some special point. We have to try $u = x^2/2$, as this is the only way we'll end up with the exponential of the negative of our variable. We want to find dx in terms of u and du for the change of variables, thus we rewrite $u = x^2/2$ as $x = (2u)^{1/2}$, which gives $dx = (2u)^{-1/2}du$. Plugging all of these in, we see

$$\begin{aligned} I \;&=\; 2\int_{0}^{\infty} e^{-u}(2u)^{-1/2} du \\ &=\; \sqrt{2}\int_{0}^{\infty} e^{-u}u^{-1/2} du. \end{aligned}$$

We're almost done – this does look very close to the Gamma function; there are just two issues, one trivial and one minor. The first is that we're using the letter u instead of x, but that's fine as we can use whatever letter we want for our variable. The second is that $\Gamma(s)$ involves a factor of u^{s-1} and we have $u^{-1/2}$. This is easily fixed; we just write

$$u^{-\frac{1}{2}} \;=\; u^{\frac{1}{2}-\frac{1}{2}-\frac{1}{2}} \;=\; u^{\frac{1}{2}-1};$$

we just **added zero**, one of the most useful things to do in mathematics. (It takes awhile to learn how to 'do nothing' well, which is why we keep pointing this out.) Thus

$$I \;=\; \sqrt{2}\int_{0}^{\infty} e^{-u}u^{\frac{1}{2}-1} du \;=\; \sqrt{2}\Gamma(1/2).$$

We did it – we've found another value of s that is important. Now we just need a way to find out what $\Gamma(1/2)$ equals! We could of course just go back to the standard normal's density and do the polar coordinate trick (see **ADD REF**); however, it is possible to evaluate this directly, and a lot can be gained by doing so. We'll give a few different proofs.

1.4.1. *The Cosecant Identity: First Proof.* Books have entire chapters on the various identities satisfied by the Gamma function. In this section we'll concentrate on one that is particularly well-suited to our investigation of $\Gamma(1/2)$, namely the cosecant identity.

The cosecant identity We have

$$\Gamma(s)\Gamma(1-s) \;=\; \pi\csc(\pi s) \;=\; \frac{\pi}{\sin(\pi s)}.$$

Before proving this, let's take a moment, as it really is just a moment, to use this to finish our study. For almost all s the cosecant identity relates two values, Gamma at s and Gamma at $1-s$; if you know one of these values, you know the other. Unfortunately, this means that in order for this identity to be useful, we have to know at least one of the two values. Unless, of course, we make the *very* special choice of taking $s = 1/2$. As $1/2 = 1 - 1/2$, the two values are the same, and we find

$$\Gamma(1/2)^2 \ = \ \Gamma(1/2)\Gamma(1/2) \ = \ \frac{\pi}{\sin(\pi/2)} \ = \ \pi;$$

taking square-roots gives $\Gamma(1/2) = \sqrt{\pi}$.

In this and the following subsections, we'll give various proofs of the cosecant identity. If all you care about is using it, you can of course skip this; however, if you read on you'll get some insight as to how people come up with formulas like this, and how they prove them. The arguments will become involved in places, but we'll try to point out why we are doing what we're doing, so that if you come across a situation like this in the future, a new situation where you are the first one looking at a problem and there is no handy guidebook available, you'll have some tools for your studies.

Proof of the cosecant identity. We've seen the cosecant identity is useful; now let's see a proof. How should we try to prove this? Well, one side is $\Gamma(s)\Gamma(1-s)$. Both of these numbers can be represented as integrals. So this quantity is really a double integral. Whenever you have a double integral, you should start thinking about changing variables or changing the order of integration, or maybe even both! The point is using the integral formulations gives us a starting point. This argument might not work, but it's something to try (and, for many math problems, one of the hardest things is just figuring out where to begin).

What we are about to write looks like it does what we have decided to do, but there's ***two*** subtle mistakes:

$$\begin{aligned}\Gamma(s)\Gamma(1-s) &= \int_0^\infty e^{-x}x^{s-1}dx \cdot \int_0^\infty e^{-x}x^{1-s-1}dx \\ &= \int_0^\infty e^{-x}x^{s-1} \cdot e^{-x}x^{1-s-1}dx. \end{aligned} \tag{1}$$

Why is this wrong? The first expression is the integral representation of $\Gamma(s)$, the second expression is the integral representation of $\Gamma(1-s)$, so their product is $\Gamma(s)\Gamma(1-s)$ and then we just collecting terms? Unfortunately, ***NO!***. The problem is that we used the same dummy variable for both integrations. We cannot write it as one integral – we had two integrations, each with a dx, and then ended up with just one dx. This is one of the most common mistakes students make. By not using a different letter for the variables in each integration, we accidentally combined them and went from a double integral to a single integral.

We should use two different letters, which in a fit of creativity we'll take to be x and y. Then

$$\begin{aligned}\Gamma(s)\Gamma(1-s) &= \int_0^\infty e^{-x}x^{s-1}dx \cdot \int_0^\infty e^{-y}y^{1-s-1}dy \\ &= \int_{y=0}^\infty \int_{x=0}^\infty e^{-x}x^{s-1}e^{-y}y^{-s}dxdy. \end{aligned}$$

While the result we're gunning for, the cosecant formula, is beautiful and important, even more important (and far more useful!) is to learn how to attack problems like this. There aren't that many options for dealing with a double integral. You can integrate as given, but in this case that would be a bad idea as we would just get back the product of the Gamma functions. What else can we do? We can switch the orders of integration. Unfortunately, that too isn't any help; switching orders can only help us if the two variables are mingled in the integral, and that isn't the case now. Here, the two variables aren't seeing each other; if we switch the order of integration, we haven't really changed anything. Only one option remains: we need to change variables.

This is the hardest part of the proof. We have to figure out a good change of variables. Let's look at the first possible choice. We have $x^{s-1}y^{-s} = (x/y)^{s-1}y^{-1}$; perhaps a good change of variables would be to let $u = x/y$? If we are to do this, we fix y, and then for fixed y we set $u = x/y$, giving $du = dx/y$. The $1/y$ is encouraging, as we had an extra y earlier. This leads to

$$\Gamma(s)\Gamma(1-s) \ = \ \int_{y=0}^\infty e^{-y}\left[\int_{u=0}^\infty e^{-uy}u^{s-1}du\right]dy.$$

Now switching orders of integration is non-trivial, as u and y appear together. That gives

$$\begin{aligned}\Gamma(s)\Gamma(1-s) &= \int_{u=0}^{\infty} u^{s-1}\left[\int_{y=0}^{\infty} e^{-(u+1)y} dy\right] du \\ &= \int_{u=0}^{\infty} u^{s-1}\left[-\frac{e^{-(u+1)y}}{u+1}\Big|_{0}^{\infty}\right] dy \\ &= \int_{u=0}^{\infty} u^{s-1}\frac{1}{u+1}\, dy \;=\; \int_{u=0}^{\infty} \frac{u^{s-1}}{u+1}\, du.\end{aligned}$$

Warning: we have to be very careful above, and make sure the interchange is justified. Remember earlier in the chapter when we had a long discussion about the importance of making sure an integral makes sense? The integrand above is $\frac{u^{s-1}}{u+1}$. It has to decay sufficiently rapidly as $u \to \infty$ and it cannot blow up too quickly as $u \to 0$ if the integral is to be finite. If you work out what this entails, it forces $s \in (0,1)$; if $s \le 0$ then it blows up too rapidly near 0, while if $s \ge 1$ it doesn't decay fast enough at infinity.

In hindsight, this restriction is not surprising, and in fact we should have expected it. Why? Remember earlier in the proof we remarked that there were ***two*** mistakes in (1); if you were really alert, you would have noticed we only mentioned ***one*** mistake! What is the missing mistake? We used the integral representation of the Gamma function. That is only valid when the argument is positive. Thus we need $s > 0$ and $1 - s > 0$; these two inequalities force $s \in (0,1)$. If you didn't catch this mistake this time, don't worry about it; just be aware of the danger in the future. This is one of the most common errors made (by both students and researchers). It's so easy to take a formula that works in some cases and accidentally use it in a place where it is not valid.

Alright. For now, let's restrict ourselves to taking $s \in (0,1)$. We leave it as an exercise to show that if the relationship holds for $s \in (0,1)$ then it holds for all s. *Hint:* keep using the functional equation of the Gamma function. It's easy to see how the $\csc(\pi s)$ or the $\sin(\pi s)$ changes if we increase s by 1; the Gamma pieces follow with a bit more work.

Now we really can say

$$\Gamma(s)\Gamma(1-s) \;=\; \int_{0}^{\infty} \frac{u^{s-1}}{u+1}\, du. \tag{2}$$

What next? Well, we have two factors, u^{s-1} and $\frac{1}{u+1}$. Note the second looks like the sum of a geometric series with ratio $-u$. Admittedly, this is not going to be an obvious identification at first, but the more math you do, the more experience you gain and the easier it is to recognize patterns. We know $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$, so all we have to do is take $r = -u$.

We must be careful – we're about to make the same mistake again, namely using a formula where it isn't applicable. It's very easy to fall into this trap. Fortunately, there's a way around it. We split the integral into two parts, the first part is when $u \in [0,1]$ and the second when $u \in [1,\infty]$. In the second part we'll then change variables by setting $v = 1/u$ and do a geometric series expansion there. **Splitting an integral** is another useful technique to master. It allows us to break a complicated problem up into simpler ones, ones where we have more results at our disposal to attack it.

For the second integral, we'll make the change $v = 1/u$. This gives $dv = -du/u^2$ or $du = -v^2 dv$ (since $1/u^2 = v^2$), and the bounds of integration go from being $u : 1 \to \infty$ to $v : 1 \to 0$ (we'll then use the negative sign to switch the order of integration to the more common $v : 0 \to 1$). Continuing onward, we have

$$\begin{aligned}\Gamma(s)\Gamma(1-s) &= \int_{0}^{1} \frac{u^{s-1}}{u+1}\, du + \int_{1}^{\infty} \frac{u^{s-1}}{u+1}\, du \\ &= \int_{0}^{1} \frac{u^{s-1}}{u+1}\, du + \int_{1}^{\infty} - \int_{1}^{0} \frac{(1/v)^{s-1}}{(1/v)+1}\, v^2 dv \\ &= \int_{0}^{1} \frac{u^{s-1}}{u+1}\, du + \int_{1}^{\infty} + \int_{0}^{1} \frac{v^{-s}}{v+1}\, dv.\end{aligned}$$

Note how similar the two expressions are. We now use the geometric series formula, and then we'll interchange the integral and the sum. Everything can be justified because $s \in (0, 1)$, so all the integrals exist and are well behaved, giving

$$\begin{aligned}
\Gamma(s)\Gamma(1-s) &= \int_0^1 u^{s-1} \sum_{n=0}^{\infty} (-1)^n u^n du + \int_0^1 v^{-s} \sum_{m=0}^{\infty} (-1)^m v^m dv \\
&= \sum_{n=0}^{\infty} (-1)^n \int_0^1 u^{s-1+n} du + \sum_{m=0}^{\infty} (-1)^m \int_0^1 v^{m-s} dv \\
&= \sum_{n=0}^{\infty} (-1)^n \frac{u^{s+n}}{n+s}\bigg|_0^1 + \sum_{m=0}^{\infty} (-1)^m \frac{v^{m+1-s}}{m+1-s}\bigg|_0^1 \\
&= \sum_{n=0}^{\infty} (-1)^n \frac{1}{n+s} + \sum_{m=0}^{\infty} (-1)^m \frac{1}{m+1-s}.
\end{aligned}$$

Note we used two different letters for the different sums. While we could have used the letter n twice, it's a good habit to use different letters. What happens now is that we'll adjust the counting a bit to easily combine them.

The two sums look very similar. They both look like a power of negative one divided by either $k + s$ or $k - s$. Let's rewrite both sums in terms of k. The first sum has one extra term, which we'll pull out. In the first sum we'll set $k = n$, while in the second we'll set $k = m+1$ (so $(-1)^m$ becomes $(-1)^{k-1} = (-1)^{k+1}$. We get

$$\begin{aligned}
\Gamma(s)\Gamma(1-s) &= \frac{1}{s} + \sum_{k=1}^{\infty} (-1)^k \frac{1}{k+s} + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k-s} \\
&= \frac{1}{s} + \sum_{k=1}^{\infty} (-1)^k \left[\frac{1}{k+s} - \frac{1}{k-s}\right] \\
&= \frac{1}{s} + \sum_{k=1}^{\infty} (-1)^k \frac{2s}{k^2-s^2} \\
&= \frac{1}{s} - \sum_{k=1}^{\infty} (-1)^k \frac{-2s}{k^2-s^2}.
\end{aligned}$$

It may not look like it, but we've just finished the proof. The problem is recognizing the above is $\pi \csc(\pi s) = \pi / \sin(\pi s)$. This is typically proved in a complex analysis course; see for instance **ADD REF**.

We can at least see it's reasonable. We're claiming

$$\frac{\pi}{\sin(\pi s)} = \frac{1}{s} - \sum_{k=1}^{\infty} (-1)^k \frac{2s}{k^2-s^2}.$$

If s is an integer then $\sin(\pi s) = 0$ and thus the left hand side is infinite, while exactly one of the terms on the right hand side blows up. This at least shows our answer is reasonable. Or mostly reasonable. It seems likely that our sum is $c/\sin(\pi s)$ for some c, but it isn't clear that that constant is π. Fortunately, there's even a way to get that, but it involves knowing a bit more about certain special sums. If we take $s = 1/2$ then the sum becomes

$$\begin{aligned} \frac{1}{1/2} - \sum_{k=1}^{\infty} (-1)^k \frac{1}{k^2 - (1/2)^2} &= 2 - \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2 - 1/4} \\ &= 2 - \sum_{k=1}^{\infty} \frac{(-1)^k 4}{4k^2 - 1} \\ &= 2 - 4 \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} \end{aligned}$$

Now add something about this – a non-standard formula!

1.4.2. *The Cosecant Identity: Second Proof.* We already have a proof of the cosecant identity for the Gamma function – why do we need another? For us, the main reason is educational. The goal of this book is not to teach you have to answer one specific problem at one moment in your life, but rather to give you the tools to solve a variety of new problems whenever you encounter them. Because of that, it's worth seeing multiple proofs as different approaches emphasize different aspects of the problem, or generalize better for other questions.

Let's go back to the set-up. We had $s \in (0, 1)$ and

$$\begin{aligned} \Gamma(s)\Gamma(1-s) &= \int_0^{\infty} e^{-x} x^{s-1} dx \cdot \int_0^{\infty} e^{-y} y^{1-s-1} dy \\ &= \int_{y=0}^{\infty} \int_{x=0}^{\infty} e^{-x} x^{s-1} e^{-y} y^{-s} dx dy. \end{aligned}$$

We've already talked about what our options are. We can't integrate it as is, or we'll just get back the two Gamma functions. We can't change the order of integration, as the x and y variables are not mingled and thus changing the order of integration won't really change the problem. The only thing left to do is change variables.

Before we set $u = x/y$. We were led to this because we saw $x^{s-1} y^{-s} = (x/y)^{s-1} y^{-1}$, and thus it's not unreasonable to set $u = x/y$. Are there any other 'good' choices for a change of variable? There is, but it's not surprising if you don't see it. It's our old friend, polar coordinates.

It should seem a little strange to use polar coordinates here. After all, we use those for problems with radial and angular symmetry. We use them for integrating over circular regions. **NONE** of this is happening here! That said, we think a good case can be made for trying this.

- First, we don't know that many change of variables; we do know polar coordinates, so we might as well try it.
- Second, we're trying to show the answer is $\pi \csc(\pi s) = \pi / \sin(\pi s)$. The answer involves the sine function, so perhaps this suggests we should try polar coordinates.

At the end of the day, a method either works or it doesn't. We hope the above at least motivates why we're trying this here, and can provide guidance for you in the future.

Recall for polar coordinates we have the following relations:

$$x = r\cos\theta, \quad y = r\sin\theta, \quad dxdy = r dr d\theta.$$

What are the bounds of integration? We're integrating over the upper right quadrant, $x, y : 0 \to \infty$. In polar coordinates it becomes $r : 0 \to \infty$ and $\theta : 0 \to \pi/2$. Our integral now becomes

$$\begin{aligned}
\Gamma(s)\Gamma(1-s) &= \int_{\theta=0}^{\pi/2}\int_{r=0}^{\infty} e^{-r\cos\theta}(r\cos\theta)^{s-1}e^{-r\sin\theta}(r\sin\theta)^{-s} r\,dr\,d\theta \\
&= \int_{\theta=0}^{\pi/2}\int_{r=0}^{\infty} e^{-r(\cos\theta+\sin\theta)}\left(\frac{\cos\theta}{\sin\theta}\right)^{s-1}\frac{1}{\sin\theta}dr\,d\theta \\
&= \int_{\theta=0}^{\pi/2}\left(\frac{\cos\theta}{\sin\theta}\right)^{s-1}\frac{1}{\sin\theta}\left[\int_{r=0}^{\infty} e^{-r(\cos\theta+\sin\theta)}dr\right]d\theta \\
&= \int_{\theta=0}^{\pi/2}\left(\frac{\cos\theta}{\sin\theta}\right)^{s-1}\frac{1}{\sin\theta}\left[-\frac{e^{-r(\cos\theta+\sin\theta)}}{\cos\theta+\sin\theta}\right]_0^{\infty}d\theta \\
&= \int_{\theta=0}^{\pi/2}\left(\frac{\cos\theta}{\sin\theta}\right)^{s-1}\frac{1}{\sin\theta}\frac{1}{\cos\theta+\sin\theta}\,d\theta.
\end{aligned}$$

It doesn't look like we've made much progress, but we're just one little change of variables away from a great simplification. Note that a lot of the integrand only depends on $\cos\theta/\sin\theta = \text{ctan}\theta$ (the cotangent of θ). If we do make the change of variables $u = \text{ctan}\theta$ then $du = -\csc^2\theta = 1/\sin^\theta$; if you don't remember this formula, you can get it by the quotient rule:

$$\begin{aligned}
\text{ctan}'(\theta) = \left(\frac{\cos\theta}{\sin\theta}\right)' &= \frac{\cos'\theta\sin\theta - \sin'\theta\cos\theta}{\sin^2\theta} \\
&= \frac{-\sin^2\theta - \cos^2\theta}{\sin^2\theta} = -\frac{1}{\sin^2\theta}.
\end{aligned}$$

Now things are looking really promising; our proposed change of variables needs a $1/\sin^2\theta$, and we already have a $1/\sin\theta$ in the integrand. We get the other by writing

$$\frac{1}{\cos\theta+\sin\theta} = \frac{1}{\sin\theta}\frac{1}{(\cos\theta/\sin\theta)+1} = \frac{1}{\sin\theta}\frac{1}{\text{ctan}\theta+1}.$$

All that remains is to find the bounds of integration. If $u = \text{ctan}\theta = \cos\theta/\sin\theta$, then $\theta : 0 \to \pi/2$ corresponds to $u : \infty \to 0$ (don't worry that we're integrating from infinity to zero – we have a minus sign floating around, and that will flip the order of integration).

Putting all the pieces together, we find

$$\begin{aligned}
\Gamma(s)\Gamma(1-s) &= \int_{\theta=0}^{\pi/2}\frac{\text{ctan}^{s-1}\theta}{\text{ctan}\theta+1}\frac{d\theta}{\sin^2\theta} \\
&= \int_{u=\infty}^{0}\frac{u^{s-1}}{u+1}(-du) = \int_0^{\infty}\frac{u^{s-1}}{u+1}du.
\end{aligned}$$

This integral should look familiar – it is exactly the integral we saw in the previous section, in equation (2). Thus from here onward we can just follow the steps in that section.

A lot of students freeze when they first see a difficult math problem. Why varies from student to student, but a common refrain is. "*I didn't know where to start.*" For those who feel that way, this should be comforting. There are (at least!) two different change of variables we can do, both leading to a solution for the problem. As you continue in math you'll see again and again that there are many different approaches you can take. Don't be afraid to try something. Work with it for awhile and see how it goes. If it isn't promising you can always backtrack and try something else.

1.4.3. *The Cosecant Identity: Special Case* $s = 1/2$. While obviously we would like to prove the cosecant formula for arbitrary s, *the* most important choice of s is clearly $s = 1/2$. We need $\Gamma(1/2)$ in order to write down the density functions for normal distributions. Thus, while it would be nice to have a formula for any s, it's still cause for celebration if we can handle just $s = 1/2$.

Remember in (2) that we showed

$$\Gamma(s)\Gamma(1-s) \;=\; \int_0^\infty \frac{u^{s-1}}{u+1}\,du.$$

Taking $s = 1/2$ gives

$$\Gamma(1/2)^2 \;=\; \int_0^\infty \frac{u^{-1/2}}{1+u}\,du.$$

We're going to solve this with a highly non-obvious change of variable. Let's state it first, see how it works, and then discuss why this is a reasonable thing to try. Here it is: take $u = z^2$, so $z = u^{1/2}$ and $dz = du/2\sqrt{u}$. Note how beautifully this fits with our integral. We have a $u^{-1/2}du$ term already, which becomes $2dz$. Substituting gives

$$\Gamma(1/2)^2 \;=\; \int_0^\infty \frac{2dz}{1+z^2} \;=\; 2\int_0^\infty \frac{dz}{1+z^2}.$$

Looking at this integral, you should think of the trigonometric substitutions from calculus. Whenever you see $1-z^2$ you should try $z = \sin\theta$ or $z = \cos\theta$; when you see $1+z^2$ you should try $z = \tan\theta$. Let's make this change of variables. The reason it's so useful is the Pythagorean formula

$$\sin^2\theta + \cos^2\theta \;=\; 1$$

becomes, on dividing both sides by $\cos^2\theta$,

$$\tan^2\theta + 1 \;=\; \frac{1}{\cos^2\theta} \;=\; \sec^2\theta.$$

Letting $z = \tan\theta$ means we replace $1+z^2$ with $\sec^2\theta$. Further, $dz = \sec^2\theta d\theta$ (if you don't remember this, just use the quotient rule: $\tan\theta = \sin\theta/\cos\theta$). As $z : 0 \to \infty$, we have $\theta : 0 \to \pi/2$. Collecting everything gives

$$\begin{aligned}\Gamma(1/2)^2 \;&=\; 2\int_0^{\pi/2} \frac{1}{\sec^2\theta}\,\sec^2\theta d\theta\\ &=\; 2\int_0^{\pi/2} d\theta \;=\; 2\frac{\pi}{2} \;=\; \pi,\end{aligned}$$

and if $\Gamma(1/2)^2 = \pi$ then $\Gamma(1/2) = \sqrt{\pi}$ as claimed!

And there we have it: a correct, elementary proof that $\Gamma(1/2) = \sqrt{\pi}$. You should be able to follow the proof line by line, but that's not the point of mathematics. The point is to *see* why the author is choosing to do these steps so taht you too could create a proof like this.

There were two changes of variables. The first was replacing u with z^2, and the second was replacing z with $\tan\theta$. The two changes are related. How can anyone be expected to think of these? To be honest, when writing this chapter one of us had to consult their notes from teaching a similar course several years ago. We remembered that somehow tangents came into the problem, but couldn't remember the exact trick we thought of so long ago. It's not easy. It takes time, but the more you do, the more patterns you can detect. We have a $1+u$ in the denominator; we know how to handle terms such as $1+z^2$ through trig substitution. As the cosecant identity involves trig functions, that suggests this could be a fruitful avenue to explore. It's not a guarantee, but we might as well try it and see where it leads.

Flush with our success, the most natural thing to try next are these substitutions for general s. If we do this, we would find

$$\begin{aligned}\Gamma(s)\Gamma(1-s) &= \int_0^\infty \frac{z^{2s-2}}{1+z^2}\, 2z dz \\ &= 2\int_0^\infty \frac{z^{2s-1}}{1+z^2}\, dz \\ &= 2\int_0^{\pi/2} \frac{\tan^{2s-1}\theta}{\sec^2\theta}\, \sec^2\theta d\theta \\ &= 2\int_0^{\pi/2} \tan^{2s-1}\theta d\theta.\end{aligned}$$

We now see how special $s = 1/2$ is. For this, and only for this, value does the integrand collapse to just being the constant function 1, which is easily integrated. Any other choice of s forces us to have to find integrals of powers of the tangent function, which is no easy task! Formulas do exist; for example,

$$\begin{aligned}\int \tan\theta d\theta &= \frac{1}{2\sqrt{2}}\Bigg[-2\arctan\left(1-\sqrt{2}\sqrt{\tan\theta}\right) + 2\arctan\left(1+\sqrt{2}\sqrt{\tan\theta}\right) \\ &\quad + \log\left(1-\sqrt{2}\sqrt{\tan\theta}+\tan\theta\right) - \log\left(1+\sqrt{2}\sqrt{\tan\theta}+\tan\theta\right)\Bigg].\end{aligned}$$

1.5. **The Beta Function and the Gamma Function.** The **Beta function** is defined by

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt, \quad a,b>0.$$

This is a very slight resemblance with the Gamma function; both involve the integration variable raised to a parameter minus 1. It turns out this is not just a coincidence or a stretch of the imagination, but rather these two functions are intimately connected by:

Fundamental Relation of the Beta Function: For $a, b > 0$ we have

$$B(a,b) := \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

With a little bit of algebra, we can rearrange the above and find

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 t^{a-1}(1-t)^{b-1}dt = 1;$$

this means that we've discovered a new density, the density of the **Beta distribution**:

$$f_{a,b} = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}dt & \text{if } 0 \le t \le 1 \\ 0 & \text{otherwise.}\end{cases}$$

We'll discuss this distribution in greater detail in **ADD REF**. For now we'll just say briefly that it's an important family of densities as a lot of times our input is between 0 and 1, and the two parameters a and b give us a lot of freedom in creating 'one-hump' distributions (namely densities that go up and then go down). We plot several of these densities in Figure 1

1.5.1. *Proof of the Fundamental Relation.* We prove the fundamental relation of the Beta function. While this is an important result, remember that our purpose in doing so is to help you see how to attack problems like this. Multiplying both sides by $\Gamma(a+b)$, we see that we must prove

$$\Gamma(a)\Gamma(b), \quad \Gamma(a+b)\int_0^1 t^{a-1}(1-t)^{b-1}dt$$

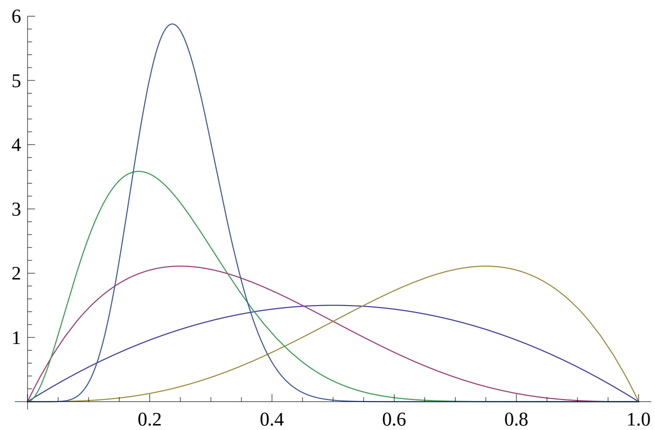

FIGURE 1. Plots of Beta densities for (a, b) equal to (2, 2), (2, 4), (4, 2), (3, 10), and (10, 30).

are equal. There are two ways to do this; we can either work with the product of the Gamma functions, or expand the $\Gamma(a+b)$ term and combine it with the other integral.

Let's try working with the product of the Gamma functions. Note that we can use the integral representation freely, as we've assumed $a, b > 0$. Note how different this is than the last section, where we had to restrict to $s \in (0,1)$. Anyway, we'll argue along the lines of our first proof of the cosecant identity, and we find

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_0^\infty e^{-x}x^{a-1}dx \int_0^\infty e^{-y}y^{b-1}dy \\ &= \int_{y=0}^\infty \int_{x=0}^\infty e^{-(x+y)}x^{a-1}y^{b-1}dxdy.\end{aligned}$$

Remember, we can't change the order of integration, as that won't gain us anything as the two variables are not mixed. Our only remaining option is to change variables. We've fixed y and are integrating with respect to x. Let's try $x = yu$ so $dx = ydu$; we've seen changes of variables like this helped in the previous section, and this will at least mix things up. We find

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_{y=0}^\infty \left[\int_{u=0}^\infty e^{-(1+y)u}(yu)^{a-1}y^{b-1}ydu\right]dy \\ &= \int_{y=0}^\infty \int_{u=0}^\infty y^{a+b-1}u^{a-1}e^{-(1+u)y}dudy \\ &= \int_{u=0}^\infty \int_{y=0}^\infty y^{a+b-1}u^{a-1}e^{-(1+u)y}dydu.\end{aligned}$$

We've changed variables and then switched the order of integration. So right now we're fixing u and then integrating with respect to y. For u fixed, consider the change of variables $t = (1+u)y$. This is a good choice, and a somewhat reasonable one to try. We need to get a $\Gamma(a+b)$ arising somehow. For that, we want something like the exponential of the negative of one of our variables. Right now we have $e^{-(1+u)y}$, which isn't of the desired form. By letting $t = (1+u)y$, however, it now becomes e^{-t}. Again, what drives this change of variables is trying to get something looking like $\Gamma(a+b)$; note how useful it is to have a sense of what the answer is!

Anyway, if $t = (1+u)y$ then $dy = dt/(1+u)$ and our integral becomes

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_{u=0}^{\infty}\int_{t=0}^{\infty}\left(\frac{t}{1+u}\right)^{a+b-1} u^{a-1}e^{-t}\frac{1}{1+u}\,dtdu \\ &= \int_{u=0}^{\infty}\left(\frac{u}{1+u}\right)^{a-1}\left(\frac{1}{1+u}\right)^{b-1}\left[\int_{t=0}^{\infty} e^{-t}t^{a+b-1}dt\right]du \\ &= \Gamma(a+b)\int_{u=0}^{\infty}\left(\frac{u}{1+u}\right)^{a-1}\left(\frac{1}{1+u}\right)^{b-1}du,\end{aligned}$$

where we used the definition of the Gamma function to replace the t-integral with $\Gamma(a+b)$. We're definitely making progress – we've found the $\Gamma(a+b)$ factor.

We should also comment on how we wrote the algebra above. We combined everything that was to the $a-1$ power together, and what was left was to the $b+1$ power. Again, this is a promising sign; we're trying to show that this equals $\Gamma(a+b)$ times an integral involving x^{a-1} and $(1-x)^{b-1}$; it's not exactly this, but it is close. (You might be a bit worried that we have a $b+1$ and not a $b-1$ – it'll work out after yet another change of variables). So, looking at what we have and again comparing it with where we want to go, what's the next change of variables? Let's try $\tau = \frac{u}{1+u}$, so $1-\tau = \frac{1}{1+u}$ and $d\tau = \frac{du}{(1+u)^2}$ (by the quotient rule), or $du = (1+u)^2 d\tau = \frac{d\tau}{(1-\tau)^2}$. Since $u: 0 \to \infty$, $\tau: 0 \to 1$. Thus

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \Gamma(a+b)\int_0^1 \tau^{a-1}(1-\tau)^{b+1}\frac{d\tau}{(1-\tau)^2} \\ &= \Gamma(a+b)\int_0^1 \tau^{a-1}(1-\tau)^{b-1}d\tau,\end{aligned}$$

which is what we needed to show!

As always, after going through a long proof we should stop, pause, and think about what we did and why. There were several change of variables and an interchange of orders of integration. As we've already discussed why these changes of variables are reasonable, we won't rehash that here. Instead, we'll talk one more time about how useful it is to know the answer. If you can guess the answer somehow, that can provide great insight as to what to do. For this problem, knowing we wanted to find a factor of $\Gamma(a+b)$ helped us make the change of variables to fix the exponential. And knowing we wanted factors like a variable to the $a-1$ power suggested the change of variables $\tau = \frac{u}{1-u}$.

1.5.2. *The Fundamental Relation and* $\Gamma(1/2)$. We give yet another derivation of $\Gamma(1/2)$, this time using properties of the Beta function. Taking $a = b = 1/2$ gives

$$\begin{aligned}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) &= \Gamma\left(\frac{1}{2}+\frac{1}{2}\right)\int_0^1 t^{1/2-1}(1-t)^{1/2-1}dt \\ &= \Gamma(1)\int_0^1 t^{-1/2}(1-t)^{-1/2}dt.\end{aligned}$$

As always, the question becomes: what's the right change of variables? If we think back to our studies of the Gamma function, we have $\Gamma(1/2)^2$ was supposed to be $\pi/\sin(\pi/2)$. This is telling us that trig functions should play a big role, so perhaps we want to do something to facilitate using trig functions or trig substitution. If so, one possibility is to take $t = u^2$. This makes the factor $(1-t)^{-1/2}$ equal to $(1-u^2)^{-1/2}$, which is ideally suited for a trig substitution.

Now for the details. We set $t = u^2$ or $u = t^{1/2}$, so $du = dt/2t^{1/2}$ or $t^{-1/2}dt = 2du$; we write it like this as we have a $t^{-1/2}dt$ already! The bounds of integration are still 0 to 1, and we have

$$\Gamma\left(\frac{1}{2}\right)^2 = \int_0^1 (1-u^2)^{-1/2}2du.$$

We now use trig substitution. Take $u = \sin\theta$, $du = \cos\theta d\theta$, and $u : 0 \to 1$ becomes $\theta : 0 \to \pi/2$ (we chose $u = \sin\theta$ over $u = \cos\theta$ as this way the bounds of integration become 0 to $\pi/2$ and not $\pi/2$ to 0, though of course either approach is fine). We now have

$$\begin{aligned}\Gamma\left(\frac{1}{2}\right)^2 &= 2\int_0^{\pi/2}(1-\sin^2\theta)^{-1/2}\cos\theta d\theta \\ &= 2\int_0^{\pi/2}\frac{\cos\theta d\theta}{(\cos^2\theta)^{1/2}} \\ &= 2\int_0^{\pi/2} d\theta \ = \ 2\cdot\frac{\pi}{2} \ = \ \pi,\end{aligned}$$

which gives us yet another way to see $\Gamma(1/2) = \sqrt{\pi}$.

2. Stirling's Formula

It is possible to look at probabilities without seeing factorials, but you have to do a lot of work to avoid them. Remember that $n!$ is the product of the first n positive integers, with the convention that $0! = 1$. Thus $3! = 3\cdot 2\cdot 1 = 6$ and $4! = 4\cdot 3! = 24$. There is a nice combinatorial interpretation: $n!$ is the number of ways to arrange n people in a line when order matters (there are n choices for the first person, then $n-1$ for the second, and so on). With this point of view, we interpret $0! = 1$ as meaning there is just one way to do arrange nothing!

Factorials arise everywhere, sometimes in an obvious manner and sometimes in a hidden one. The first instance of factorials encountered in probability is with the binomial coefficients, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Remember this is the number of ways to choose k objects from n when order does not matter.

There are, however, less obvious occurrences of the factorial function. Perhaps the best hidden one occurs in the density function of the standard normal, $\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$. It turns out that $\sqrt{\pi} = (-1/2)!$. You should be hearing alarm bells upon reading this; after all, we've defined the factorial function for integer input, and now we are taking the factorial, not just of a negative number, but of a negative rational! What does this mean? How do we interpret the factorial of $-1/2$? What does it mean to ask about the number of ways of order -1/2 people?

The answer to this is through the Gamma function, $\Gamma(s)$, where

$$\Gamma(s) \ = \ \int_0^\infty e^{-x}x^{s-1}dx, \quad \Re(s) > 0.$$

Some authors write this as

$$\Gamma(s) \ = \ \int_0^\infty e^{-x}x^s\frac{dx}{x}.$$

There is no difference in the two expressions, though they look different. Writing dx/x emphasizes how nicely the measure transforms under rescaling: if we send x to $u = ax$ for any fixed a, then $dx/x = du/u$.

It turns out the Gamma function generalizes the factorial function: if n is a non-negative integer then $\Gamma(n+1) = n!$. We described this and other properties of the Gamma function in **ADD REF**.

The purpose of this chapter is to describe Stirling's formula for approximating $n!$ for large n, and discuss some applications. The key fact is

Stirling's formula: As $n \to \infty$, we have

$$n! \;\approx\; n^n e^{-n}\sqrt{2\pi n};$$

by this we mean

$$\lim_{n\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} \;=\; 1.$$

More precisely, we have the following series expansion:

$$n! \;=\; n^n e^{-n}\sqrt{2\pi n}\left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \cdots\right).$$

Whenever you see a formula, you should always try some simple tests to see if it is reasonable. What kind of tests can we apply? Well, Stirling's formula claims that $n! \approx n^n e^{-n}\sqrt{2\pi n}$. As $n! = n(n-1)\cdots 1$, clearly $n! \le n^n$, which is consistent with Stirling though it is too large by approximately e^{-n}. What about a lower bound? Well, clearly $n! \ge n(n-1)\cdots \frac{n}{2}$ (we'll assume $n/2$ is even for convenience), so $n! \ge (n/2)^{n/2}$. While this is a lower bound, it is a poor one; it looks like $n^{n/2}2^{-n/2}$; the power of n should be n and not $n/2$ according to Stirling. There's a bit of an art to finding good, elementary upper and lower bounds. This is a more advanced topic and requires a 'feel' for how to proceed; nevertheless, it's a great skill to develop. So as not to interrupt the flow, we'll hold off on these more elementary bounds for now, and wait till §2.5 to show you how close one can get to Stirling just by knowing how to count.

What other checks can we do? Well, $(n+1)!/n! = n+1$; let's see what Stirling gives:

$$\begin{aligned}
\frac{(n+1)!}{n!} &= \frac{(n+1)^{n+1}e^{-(n+1)}\sqrt{2\pi(n+1)}}{n^n e^{-n}\sqrt{2\pi n}} \\
&= (n+1)\cdot\left(\frac{n+1}{n}\right)^n \cdot \frac{1}{e}\cdot\sqrt{\frac{n+1}{n}} \\
&= (n+1)\left(1+\frac{1}{n}\right)^n \frac{1}{e}\sqrt{1+\frac{1}{n}}.
\end{aligned}$$

As $n \to \infty$, $(1+1/n)^n \to e$ (this is the definition of e; see **ADD REF**) and $\sqrt{1+1/n} \to 1$; thus our approximation above is basically $(n+1)\cdot e \cdot \frac{1}{e}\cdot 1$, which is $n+1$ as needed for consistency!

While the arguments above are not proofs, they are almost as valuable. It is essential to be able to look at a formula and get a feel for what it is saying, and for whether or not it is true. By some simple inspection, we can get upper and lower bounds that sandwich Stirling's formula. Moreover, Stirling's formula is consistent with $(n+1)! = (n+1)n!$. This gives us reason to believe we're on the right track, especially when we see how close our ratio was to $n+1$.

A full proof of Stirling's formula is beyond the scope of a first course in probability. The purpose of this chapter is to give several arguments supporting it, and discuss some of its applications. Our hope is that the time spent on these arguments will give you a better sense of what's going on.

2.1. **Stirling's Formula and Probabilities.** Before getting bogged down in the technical details of the proofs of Stirling's formula, let's first spend some time seeing how it can be used. Our first problem is inspired by a phenomenon that surprises many students. Imagine we have a fair coin, so it lands on heads half the time and tails half the time. If we flip it $2n$ times, we expect to get n heads; thus if we flip it two million times we expect one million heads. It turns out that as $n \to \infty$, the probability of getting *exactly* n heads in $2n$ tosses of the fair coin tends to zero!

This result can seem quite startling. The expected value from $2n$ tosses is n heads, and the probability of getting n heads tends to zero? What's going on? If this is the expected value, shouldn't it be very likely! The key to understanding this problem is to note that while the expected value is n heads in $2n$ tosses, the standard deviation is $\sqrt{n/2}$; returning to the case of two million tosses, we expect one million heads with fluctuations of the size 700. If now we tossed the coin two trillion times, we would expect one trillion heads and fluctuations on the order of 700,000. As n increases, the 'window' about the mean where outcomes are likely is growing like the standard deviation, i.e., it is growing like $\sqrt{N}$ (up to some constants). Thus the probability is being shared among more and more values, and thus it makes sense that the probabilities of individual, specific outcomes (like exactly n tosses) is going down. In summary: the probability of exactly n heads in $2n$ tosses is going down as the probability is being spread over a wider and wider set. We now use Stirling's formula to quantify just how rapidly this probability decays.

For large n, we can painlessly approximate $n!$ with Stirling's formula. Remember we are trying to answer: *What is the probability of getting exactly n heads in $2n$ tosses of a fair coin?*

It is very easy to write down the answer: it is simply

$$\text{Prob(exactly } n \text{ heads in } 2n \text{ tosses)} \;=\; \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n \;=\; \frac{1}{2^{2n}}.$$

The reason is that each of the 2^{2n} strings of heads and tails are equally likely, and there are $\binom{2n}{n}$ strings with exactly n heads. How big is $\binom{2n}{n}$?

It's now Stirling's formula to the rescue. We have

$$\begin{aligned}\binom{2n}{n} &= \frac{(2n)!}{n!n!}\\ &\approx \frac{(2n)^{2n}e^{-2n}\sqrt{2\pi\cdot 2n}}{n^ne^{-n}\sqrt{2\pi n}\cdot n^ne^{-n}\sqrt{2\pi n}}\\ &= \frac{2^{2n}}{\sqrt{\pi n}};\end{aligned}$$

thus the probability of exactly n heads is

$$\binom{2n}{n}\frac{1}{2^{2n}} \;\approx\; \frac{1}{\sqrt{\pi n}}.$$

This means that if $n = 100$ there is a little less than a 6% chance of getting half heads, while if $n = 1,000,000$ the probability falls to less than .06%.

The above exercise was actually discussed in the 2008 presidential primary season. Obama and Clinton each received 6,001 votes in the democratic primary in Syracuse, NY. While there were 12,346 votes cast in the primary, for simplicity let's assume that there were just 12,002, and ask what is the probability of a tie. If we assume each candidate is equally likely to get any vote, the answer is just $\binom{12{,}002}{6{,}001}/2^{12{,}002}$. The exact answer is approximately 0.00728; using our approximation from above we would estimate it as

$$\frac{1}{\sqrt{\pi\cdot 12,002}} \;\approx\; 0.00514,$$

which is fairly close to the true answer.

While we found the probability is a little less than 1%, some news outlets reported that the probability was about one in a million, with some going so far as to say it was "almost impossible". Why are there such different answers? It all depends on how you model the problem. If we assume the two candidates are equally likely to get each vote, then we get an answer of a little less than 1%. If, however, we take into account other bits of information, then the story changes. For example, Clinton was then a senator from NY, and it should be expected that she'll do better in her current home state. In fact, she won the state overall with 57.37% of the vote to Obama's 40.32%. Again for ease of analysis, let's say Clinton had 57% and Obama 43%. If we use these numbers, we no longer assume each voter is equally likely to vote for Clinton or Obama, and we find the probability of a tie is just $\binom{12002}{6001}.57^{6001}.43^{6001}$. Using Stirling's formula, we found $\binom{2n}{n} \approx 2^{2n}/\sqrt{\pi n}$. Thus, under these assumptions, the probability of a tie is approximately

$$\frac{2^{12002}}{\sqrt{\pi \cdot 6001}} \cdot .57^{6001} .43^{6001} \approx 1.877 \cdot 10^{-54}.$$

Note how widely different the probabilities are depending on our assumption of what is to be expected!

2.2. **Stirling's Formula and Convergence of Series.** Another great application of Stirling's formula is to help us determine for what x certain series converge. We have lots of powerful tests from calculus for this purpose, such as the ratio, root and integral test; we review these in **ADD REF**; however, we can frequently avoid having to use these tests and instead use the simpler comparison test by applying Stirling's formula.

For our first example, let's take

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

We want to find out for which x does it converge. Using the ratio test we would see that it converges for *any* choice of x. We could use the root test, but that requires us to know a little bit about the growth of $n!$; fortunately Stirling's formula gives us this information. We have

$$n!^{1/n} \approx \left(n^n e^{-n} \sqrt{2\pi n}\right)^{1/n} = \frac{n}{e};$$

as this tends to infinity we find $(1/n!)^{1/n}$ tends to zero, and thus by the root test the radius of convergence is infinity (i.e., the series converges for all x).

While the above is a method to determine that the series for e^x converges for all x, it is quite unsatisfying. In addition to using Stirling's formula we also had to use the powerful root test. Is it possible to avoid using the root test, and determine that the series always converges *just* by using Stirling's formula? The answer is yes, as we now show.

We wish to show the series for e^x converges for all x. The idea is that $n!$ grows so rapidly that, no matter what x we take, $x^n/n!$ rapidly tends to zero. If we can show that, for all n sufficiently large, $|x^n/n!| < r(x)$ for some $r(x)$ less than 1, then the series for e^x converges by the comparison test; we wrote $r(x)$ to emphasize that the bound may depend on x. Explicitly, let's say our inequality holds for all $n \geq N(x)$. To determine if a series converges, it is enough to study the tail, as the finite number of summands in the beginning don't affect convergence. We have

$$\left| \sum_{n=N(x)}^{\infty} \frac{x^n}{n!} \right| \leq \sum_{n=N(x)}^{\infty} r(x)^n = \frac{r(x)^{N(x)}}{1 - r(x)}.$$

We are thus reduced to proving that $|x^n/n!| \leq r(x) < 1$ for all n large for some $r(x) < 1$. Plugging in Stirling's formula, we see that $x^n/n!$ looks like

$$\frac{x^n}{n^n e^{-n} \sqrt{2\pi n}} = \frac{1}{\sqrt{2\pi n}} \left(\frac{ex}{n}\right)^n \leq \left(\frac{ex}{n}\right)^n.$$

For any fixed x, once $n \geq 2ex + 1$ then $|ex/n| \leq 1/2$, which completes the proof.

MOVE THIS TO ANOTHER CHAPTER, SAY THE GENERATING FUNCTION CHAPTER. We can use this type of argument to prove many important series converge. Let's consider the moment generating function of the standard normal (for the definition and properties of the moment generating function, see §**ADD REF**). The moments of the standard normal are readily determined; the n^{th} moment is

$$\mu_n = \int_{-\infty}^{\infty} x^n \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \begin{cases} (2m-1)!! & \text{if } n = 2m \text{ is even} \\ 0 & \text{if } n = 2m+1 \text{ is odd,} \end{cases}$$

where the double factorial means take every other term until you reach 2 or 1. Thus $4!! = 4 \cdot 2$, $5!! = 5 \cdot 3 \cdot 1$, $6!! = 6 \cdot 4 \cdot 2$ and so on. The moment generating function, $M_X(t)$, is

$$M_X(t) \;=\; \sum_{n=0}^{\infty} \frac{\mu_n}{n!} t^n \;=\; \sum_{m=0}^{\infty} \frac{(2m-1)!!}{(2m)!} t^{2m}.$$

There are many ways to do the algebra. We need to understand how rapidly $(2m-1)!!t^{2m}/(2m)!$ is decaying. We have

$$\frac{(2m-1)!!}{(2m)!} \;=\; \frac{(2m-1)!!}{(2m)!! \cdot (2m-1)!!} \;=\; \frac{1}{2m \cdot (2m-2) \cdots 2} \;=\; \frac{1}{2^m m!}.$$

Thus

$$M_X(t) \;=\; \sum_{m=0}^{\infty} \frac{t^{2m}}{2^m m!} \;=\; \frac{(t^2/2)^m}{m!} \;=\; e^{t^2/2}.$$

2.3. **From Stirling to the Central Limit Theorem.** This section becomes a bit technical, and can safely be skipped; however, if you spend the time mastering it you'll learn some very useful techniques for attacking and estimating probabilities, and see a very common pitfall and learn how to avoid it.

The Central Limit Theorem is one of the gems of probability, saying the sum of nice independent random variables converges to being normally distributed as the number of summands grows. As a powerful application of Stirling's formula, we'll show it implies the Central Limit Theorem for the special case when the random variables $X_1, \dots, X_{2N}$ are all binomial random variables. It is technically easiest if we normalize these by

$$\text{Prob}(X_i = n) \;=\; \begin{cases} 1/2 & \text{if } n = 1 \\ 1/2 & \text{if } n = -1 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Think of this as a 1 for a head and a -1 for a tail; this normalization allows us to say that the expect value of the sum $X_1 + \cdots + X_{2N} = 0$. As we may interpret X_i as whether or not the i^{th} toss is a head or a tail, this is the expected number of heads (which is zero).

Let $X_1, \dots, X_{2N}$ be independent binomial random variables with probability density given by (3). Then the mean is zero as $1 \cdot (1/2) + (-1) \cdot (1/2) = 0$, and the variance of each is

$$\sigma^2 \;=\; (1-0)^2 \cdot \frac{1}{2} + (-1-0)^2 \cdot \frac{1}{2} \;=\; 1.$$

Finally, we set

$$S_{2N} \;=\; X_1 + \cdots + X_{2N}.$$

Its mean is zero. This follows from

$$\mathbb{E}[S_{2N}] \;=\; \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_{2N}] \;=\; 0 + \cdots + 0 \;= 0.$$

Similarly, we see the variance of S_{2N} is $2N$. We therefore expect S_{2N} to be on the order of 0, with fluctuations on the order of $\sqrt{2N}$.

Let's consider the distribution of S_{2N}. We first note that the probability that $S_{2N} = 2k + 1$ is zero. This is because S_{2N} equals the number of heads minus the number of tails, which is always even: if we have k heads and $2N - k$ tails then S_{2N} equals $2N - 2k$.

The probability that S_{2N} equals $2k$ is just $\binom{2N}{N+k}(\frac{1}{2})^{N+k}(\frac{1}{2})^{N-k}$. This is because for S_{2N} to equal $2k$, we need $2k$ more 1's (heads) than -1's (tails), and the number of 1's and -1's add to $2N$. Thus we have $N+k$ heads (1's) and $N-k$ tails (-1's). There are 2^{2N} strings of 1's and -1's, $\binom{2N}{N+k}$ have exactly $N+k$ heads and $N-k$ tails, and the probability of each string is $(\frac{1}{2})^{2N}$. We have written $(\frac{1}{2})^{N+k}(\frac{1}{2})^{N-k}$ to show how to handle the more general case when there is a probability p of heads and $1-p$ of tails.

We now use Stirling's Formula to approximate $\binom{2N}{N+k}$. We find

$$\begin{aligned}\binom{2N}{N+k} &\approx \frac{(2N)^{2N}e^{-2N}\sqrt{2\pi\cdot 2N}}{N^N e^{-N}\sqrt{2\pi N}N^N e^{-N}\sqrt{2\pi N}}\\ &= \frac{(2N)^{2N}}{(N+k)^{N+k}(N-k)^{N-k}}\sqrt{\frac{N}{\pi(N+k)(N-k)}}\\ &= \frac{2^{2N}}{\sqrt{\pi N}}\frac{1}{(1+\frac{k}{N})^{N+\frac{1}{2}+k}(1-\frac{k}{N})^{N+\frac{1}{2}-k}}.\end{aligned}$$

The rest of the argument is just doing some algebra to show that this converges to a normal distribution. There is, unfortunately, a very common trap people frequently fall into when dealing with factors such as these. To help you avoid these in the future, we'll describe this common error first and then finish the proof.

We would like to use the definition of e^x from **ADD REF** to deduce that as $N\to\infty$, $\left(1+\frac{w}{N}\right)^N \approx e^w$; unfortunately, we must be a little more careful as the values of k we consider grow with N. For example, we might believe that $(1+\frac{k}{N})^N \to e^k$ and $(1-\frac{k}{N})^N \to e^{-k}$, so these factors cancel. As k is small relative to N we may ignore the factors of $1/2$, and then say

$$\left(1+\frac{k}{N}\right)^k = \left(1+\frac{k}{N}\right)^{N\cdot\frac{k}{N}} \to e^{k^2/N};$$

similarly, $(1-\frac{k}{N})^{-k} \to e^{k^2/N}$. Thus we would claim (*and we shall see later in Lemma 2.1 that this claim is in error!*) that

$$\left(1+\frac{k}{N}\right)^{N+\frac{1}{2}+k}\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k} \to e^{2k^2/N}.$$

We show that $\left(1+\frac{k}{N}\right)^{N+\frac{1}{2}+k}\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k} \to e^{k^2/N}$. The importance of this calculation is that it highlights how crucial *rates* of convergence are. While it is true that the main terms of $(1\pm\frac{k}{N})^N$ are $e^{\pm k}$, the error terms (in the convergence) are quite important, and yield large secondary terms when k is a power of N. What happens here is that the secondary terms from these two factors reinforce each other. Another way of putting it is that one factor tends to infinity while the other tends to zero. Remember that $\infty\cdot 0$ is one of our undefined expressions; it can be anything depending on how rapidly the terms grow and decay; we'll say more about this at the end of the section.

The short of it is that we cannot, sadly, just use $\left(1+\frac{w}{N}\right)^N \approx e^w$. We need to be more careful. The correct approach is to take the logarithms of the two factors, Taylor expand the logarithms, and then exponentiate. This allows us to better keep track of the error terms.

Before doing all of this, we need to know roughly what range of k will be important. As the standard deviation is $\sqrt{2N}$, we expect that the only k's that really matter are those within a few standard deviations from 0; equivalently, k's up to a bit more than $\sqrt{2N}$. We can carefully quantify exactly how large we need to study k by using Chebyshev's Inequality (see **ADD REF**). From this we learn that we need only study k where $|k|$ is at most $N^{\frac{1}{2}+\epsilon}$. This is because the standard deviation of S_{2N} is $\sqrt{2N}$. We then have

$$\text{Prob}(|S_{2N}-0|\geq(2N)^{1/2+\epsilon}) \leq \frac{1}{(2N)^{2\epsilon}},$$

because $(2N)^{1/2+\epsilon}=(2N)^{\epsilon}\text{StDev}(S_{2N})$. Thus it suffices to analyze the probability that $S_{2N}=2k$ for $|k|\leq N^{\frac{1}{2}+\frac{1}{9}}$.

We now come to the promised lemma which tells us what the right value is for the product; the proof will show us how we should attack problems like this in general.

Lemma 2.1. *For any* $\epsilon \le \frac{1}{9}$, *for* $N \to \infty$ *with* $k \le (2N)^{1/2+\epsilon}$, *we have*

$$\left(1+\frac{k}{N}\right)^{N+\frac{1}{2}+k}\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k} \longrightarrow e^{k^2/N}e^{O(N^{-1/6})}.$$

Proof: Recall that for $|x| < 1$,

$$\log(1+x) \;=\; \sum_{n=1}^{\infty}\frac{(-1)^{n+1}x^n}{n}.$$

As we are assuming $k \le (2N)^{1/2+\epsilon}$, note that any term below of size k^2/N^2, k^3/N^2 or k^4/N^3 will be negligible. Thus if we define

$$P_{k,N} \quad := \quad \left(1+\frac{k}{N}\right)^{N+\frac{1}{2}+k}\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k}$$

then using the big-Oh notation from **ADD REF** we find

$$\begin{aligned}
\log P_{k,N} \quad &= \quad \left(N+\frac{1}{2}+k\right)\log\left(1+\frac{k}{N}\right)+\left(N+\frac{1}{2}-k\right)\log\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k} \\
&= \quad \left(N+\frac{1}{2}+k\right)\left(\frac{k}{N}-\frac{k^2}{2N^2}+O\left(\frac{k^3}{N^3}\right)\right) \\
&\qquad +\left(N+\frac{1}{2}-k\right)\left(-\frac{k}{N}-\frac{k^2}{2N^2}+O\left(\frac{k^3}{N^3}\right)\right) \\
&= \quad \frac{2k^2}{N}-2\left(N+\frac{1}{2}\right)\frac{k^2}{2N^2}+O\left(\frac{k^3}{N^2}+\frac{k^4}{N^3}\right) \\
&= \quad \frac{k^2}{N}+O\left(\frac{k^2}{N^2}+\frac{k^3}{N^2}+\frac{k^4}{N^3}\right).
\end{aligned}$$

As $k \le (2N)^{1/2+\epsilon}$, for $\epsilon < 1/9$ the big-Oh term is dominated by $N^{-1/6}$, and we finally obtain that

$$P_{k,N} \quad = \quad e^{k^2/N}e^{O\left(N^{-1/6}\right)},$$

which completes the proof. □

We now finish the proof of S_{2N} converging to a Gaussian. Combining Lemma 2.1 with (4) yields

$$\binom{2N}{N+k}\frac{1}{2^{2N}} \quad \approx \quad \frac{1}{\sqrt{\pi N}}\,e^{-k^2/N}.$$

The proof of the Central Limit Theorem in this case is completed by some simple algebra. We are studying $S_{2N} = 2k$, so we should replace k^2 with $(2k)^2/4$. Similarly, since the variance of S_{2N} is $2N$, we should replace N with $(2N)/2$. While these may seem like unimportant algebra tricks, it is very useful to gain an ease at doing this. By doing such small adjustments we make it easier to compare our expression with its conjectured value. We find

$$\text{Prob}(S_{2N} = 2k) \;=\; \binom{2N}{N+k}\frac{1}{2^{2N}} \quad \approx \quad \frac{2}{\sqrt{2\pi\cdot(2N)}}\,e^{-(2k)^2/2(2N)}.$$

Remember S_{2N} is never odd. The factor of 2 in the numerator of the normalization constant above reflects this fact, namely the contribution from the probability that S_{2N} is even is twice as large as we would expect, because it has to account for the fact that the probability that S_{2N} is odd is zero. Thus the above looks like a Gaussian with mean 0 and variance $2N$. For N large such a Gaussian is slowly varying, and integrating from $2k$ to $2k+2$ is basically $2/\sqrt{2\pi(2N)}\cdot\exp-(2k)^2/2(2N)$. □

As our proof was long, let's spend some time going over the key points. We were fortunate in that we had an explicit formula for the probability, and that formula involved binomial coefficients. We used Chebyshev's inequality to limit which probabilities we had to investigate. We then expanded using Stirling's formula, and did some algebra to make our expression look like a Gaussian.

For a nice challenge: Can you generalize the above arguments to handle the case when $p \neq \frac{1}{2}$.

2.4. **Integral Test and the Poor Man's Stirling.** Using the integral test from calculus, we'll show that

$$n^n e^{-n} \cdot e \ \leq \ n! \ \leq \ n^n e^{-n} \cdot en.$$

As we've remarked throughout the book, whenever we see a formula our first response should be to test its reasonableness. Before going through the proof, let's compare this to Stirling's formula, which says $n! \approx n^n e^{-n}\sqrt{2\pi n}$. We correctly identify the main factor of $n^n e^{-n}$, but we miss the factor of $\sqrt{2\pi n}$. It is interesting to note how much we miss by. Our lower bound is just e while our upper bound is en. If we take the geometric mean of these two (recall the geometric mean of x and y is $\sqrt{xy}$), $\sqrt{e \cdot en}$, we get $e\sqrt{n}$. This is approximately $2.718\sqrt{n}$, which is remarkably close to the true answer, which is $\sqrt{2\pi n} \approx 2.5063\sqrt{n}$.

The gist of the above is: our answer is quite close. The arguments below are fairly elementary. The involve the integral test from calculus, and the Taylor series expansion of $\log(1+x)$, which is just $x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$. While the algebra grows a bit at the end, it's important not to let that distract from the main idea, which is that we can approximate a sum very well and fairly easily through an integral. The difficulty is when we want to quantify how close the integral is to our sum; this requires us to do some slightly tedious book-keeping. We go through the details of this proof to highlight how you can use these techniques to attack problems. We'll recap the arguments afterwards, emphasizing what you should take-away from all of this.

Now, to the proof! Let $P = n!$ (we use the letter P because *product* starts with P). It is very hard to look at a product and have a sense of what it's saying. This is in part because our experience in previous classes is always with sums, not products. For example, in calculus we encounter Riemann sums all the time; I've never seen a Riemann product!

We thus want to convert our problem on $n!$ to a related one where we have more experience. The natural thing to do is to take the logarithm of both sides. The reason is that the logarithm of a product is the sum of the logarithms. This is excellent general advice: you should have a Pavlovian response to take a logarithm whenever you run into a product.

Returning to our problem, we have

$$\log P \ = \ \log n! \ = \ \log 1 + \log 2 + \cdots + \log n \ = \ \sum_{k=1}^{n} \log k.$$

We want to approximate this sum with an integral. Note that if $f(x) = \log x$ then this function is increasing for $x \geq 1$. We claim this means

$$\int_1^n \log t dt \ \leq \ \sum_{k=1}^{n} \log k \ \leq \ \int_2^{n+1} \log t dt.$$

This follows by looking at the upper and lower sums; note this is the same type of argument as you've seen in proving the Fundamental Theorem of Calculus (the old upper and lower sum approach); see Figure 2. This is probably the most annoying part of the argument, getting the bounds for the integrals correct.

We now come to the hardest part of the argument. We need to know what is the integral of $\log t$. This is not one of the standard functions, but it turns out to have a relatively simple anti-derivative, namely $t \log t - t$. While it is very hard to find a typical anti-derivative, it's very straightforward to check and make sure it works – all we have to do is take the derivative. We can now use this to

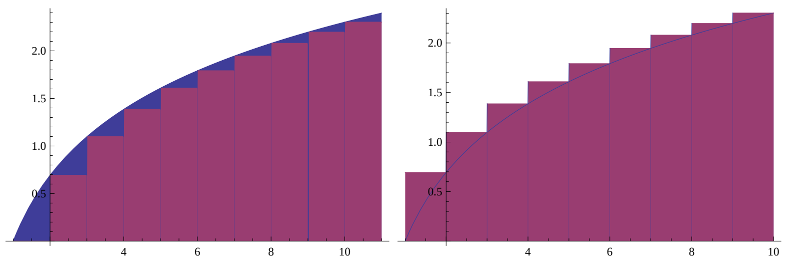

FIGURE 2. Lower and upper bound for $\log n!$ when $n = 10$.

approximate $\log n!$. Note we only make the upper integral larger if we start it at 1 and not 2. We find

$$\begin{aligned} (t\log t - t)\Big|_{t=1}^{n} \ &\leq\ \log n!\ \leq\ (t\log t - t)\Big|_{t=1}^{n+1} \\ n\log n - n + 1\ &\leq\ \log n!\ \leq\ (n+1)\log(n+1) - (n+1) + 1. \end{aligned}$$

We'll study the lower bound first. From

$$n\log n - n + 1\ \leq\ \log n!,$$

we find after exponentiating that

$$e^{n\log n - n + 1}\ =\ n^n e^{-n}\cdot e\ \leq\ n!.$$

What about the upper bound? We have

$$\begin{aligned} (n+1)\log(n+1) - n\ &=\ (n+1)\log\left(n\left(1+\frac{1}{n}\right)\right) - n \\ &=\ (n+1)\log n + (n+1)\log\left(1+\frac{1}{n}\right) - n \\ &=\ n\log n + \log n - n + (n-1)\left(\frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^2} - \cdots\right) \\ &\leq\ n\log n + \log n - n + 1, \end{aligned}$$

where the last follows from the fact that we have an alternating series, and so truncating after a positive term overestimates the sum. Exponentiating gives

$$n!\ \leq\ e^{n\log n - n + \log n + 1}\ =\ n^n e^{-n}\cdot en.$$

Putting the two bounds together, we've shown that

$$n^n e^{-n}\cdot e\ \leq\ n!\ \leq\ n^n e^{-n}\cdot en,$$

which is what we wanted to show. Note our arguments were entirely elementary, and introduce many nice and powerful techniques that can be used for a variety of problems. We replace a sum with an integral. We replace a complicated function, $\log(n+1)$, with its Taylor expansion. We note that for an alternating series, if we truncate after a positive term we over-estimate. Finally, and most importantly, we've seen how to get a handle on products – we should take logarithms and convert the product to a sum!

2.5. **Elementary Approaches towards Stirling's Formula.** In the last section we used the integral test to get very good upper and lower bounds for $n!$. While the bounds are great, we did have to use calculus twice. Once was in applying the integral test, and the other was being inspired to see that the anti-derivative of $\log t$ is $t\log t - t$; while it is easy to check this by differentiation, if you haven't seen such relations before it looks like it is a real challenge to find.

In this section we'll present an entirely elementary approach to estimates of Stirling's formula. We'll mostly avoid using calculus, instead just counting in a clever manner. *You may safely skip this section; however, the following arguments highlight a great way to look at estimation, and these ideas will almost surely be useful for a variety of other problems that you'll encounter over the years.*

2.5.1. *Lower bounds towards Stirling, I.* One of the most useful skills you can develop is the knack for how to approximate well a very complicated quantity. While we can often resort to a computer for brute force computations to get a feel for the answer, there are times when the parameter dependence is so wild that this is not realistic. Thus, it is very useful to learn how to look at a problem and glean something about the behavior as parameters change.

Stirling's formula provides a wonderful testing ground for some of these methods. Remember it says that $n! \sim n^n e^{-n}\sqrt{2\pi n}$ as n tends to infinity. We've already seen how to get a reasonable upper bound without too much work; what about a good lower bound? Our first attempt (see **ADD REF**) was quite poor; now we show a truly remarkable approach that leads to a very good lower bound with very little work.

To simplify the presentation, we assume that $n = 2^N$ for some N; if we don't make this assumption, we need to use floor functions throughout, or do a bit more arguing (which we'll do later). Using the floor function makes the formulas and the analysis look more complicated, and the key insight, the main idea, is now buried in unenlightening algebra which is required to make the statements true. We prefer to concentrate on this special case as we can then highlight the method without being bogged down in details.

The idea is to use **dyadic decompositions**. This is a powerful idea, and is used throughout mathematics. We study the factors of $n!$ in the intervals $I_1 = (n/2, n]$, $I_2 = (n/4, n/2]$, $I_3 = (n/8, n/4]$, $\dots$, $I_N = (1, 2)$. Note on I_k that each of the $n/2^k$ factors is at least $n/2^k$. Thus

$$\begin{aligned} n! &= \prod_{k=1}^{N} \prod_{m \in I_k} m \\ &\geq \prod_{k=1}^{N} \left(\frac{n}{2^k}\right)^{n/2^k} \\ &= n^{n/2+n/4+n/8+\cdots+n/2^N} 2^{-n/2} 4^{-n/4} 8^{-n/8} \cdots (2^N)^{-n/2^N}. \end{aligned}$$

Let's look at each factor above slowly and carefully. Note the powers of n *almost* sum to n; they would if we just add $n/2^N = 1$ (since we are assuming $n = 2^N$). Remember, though, that $n = 2^N$; there is thus no harm in multiplying by $(n/2^N)^{n/2^N}$ as this is just 1^1. We now have $n!$ is greater than

$$n^{n/2+n/4+n/8+\cdots+n/2^N+n/2^N} 2^{-n/2} 4^{-n/4} 8^{-n/8} \cdots (2^N)^{-n/2^N} (2)^{-n/2^N}.$$

Thus the n-term gives n^n. What of the sum of the powers of 2? That's just

$$\begin{aligned} 2^{-n/2} 4^{-n/4} 8^{-n/8} \cdots (2^N)^{-n/2^N} \cdot 2^{-n/2^N} &= 2^{-n\left(1/2+2/4+3/8+\cdots N/2^N\right)} 2^{-2^N/2^N} \\ &> 2^{-n\left(\sum_{k=0}^{N} k/2^k\right)} \\ &\leq 2^{-n\left(\sum_{k=0}^{\infty} k/2^k\right)} 2^{-1} \\ &= 2^{-2n-1} \ = \ \frac{1}{2} 4^{-n}. \end{aligned}$$

To see this, we use the following wonderful identity:

$$\sum_{k=0}^{\infty} k x^k \ = \ \frac{x}{(1-x)^2};$$

for a proof, see the section on differentiating identities **ADD REF**.

Putting everything together, we find

$$n! \ \geq \ \frac{1}{2} n^{n-1} 4^{-n},$$

which compares favorably to the truth, which is $n^n e^{-n}$.

As with many things in life, we can get a better result if we are willing to do a little more work. For example, consider the interval $I_1 = (n/2, n]$. We can pair elements at the beginning and the end: n and $n/2+1$, $n-1$ and $n/2+2$, $n-2$ and $n/2+3$ and so on until $3n/4$ and $3n/4+1$; for example, if we have the interval $(8, 16]$ then the pairs are: (16,9), (15,10), (14,11), and (13,12). We now use one of the gold standard problems from calculus: if we want to maximize xy given that $x + y = L$ then the maximum occurs when $x = y = L/2$. This is frequently referred to as the Farmer Bob (or Brown) problem, and is given the riveting interpretation that if we are trying to find the rectangular pen that encloses the maximum area for his cows to graze given that the perimeter is L, then the answer is a square. Thus of all our pairs, the one that has the largest product is $3n/4$ with $3n/4 + 1$, and the smallest is n and $n/2 + 1$, which has a product exceeding $n^2/2$. We therefore decrease the product of all elements in I_1 by replacing each product with $\sqrt{n^2/2} = n/\sqrt{2}$. Thus a little thought gives us that

$$n \cdot (n-1) \cdots \frac{3n}{4} \cdots \left(\frac{n}{2}+1\right) \cdot \frac{n}{2} \;\geq\; \left(\frac{n}{\sqrt{2}}\right)^{n/2} \;=\; \left(\frac{n\sqrt{2}}{2}\right)^{n/2},$$

a nice improvement over $(n/2)^{n/2}$, and this didn't require too much additional work!

We now do a similar analysis on I_2; again the worst case is from the pair $n/2$ and $n/4 + 1$ which has a product exceeding $n^2/8$. Arguing as before, we find

$$\prod_{m \in I_2} m \;\geq\; \left(\frac{n}{\sqrt{8}}\right)^{n/4} \;=\; \left(\frac{n}{2\sqrt{2}}\right)^{n/4} \;=\; \left(\frac{n\sqrt{2}}{4}\right)^{n/4}.$$

At this point hopefully the pattern is becoming clear. We have almost exactly what we had before; the only difference is that we have a $n\sqrt{2}$ in the numerator each time instead of just an n. This leads to very minor changes in the algebra, and we find

$$n! \;\geq\; \frac{1}{2}(n\sqrt{2})^n 4^{-n} \;=\; n^n (2\sqrt{2})^{-n}.$$

Notice how close we are to $n^n e^{-n}$, as $2\sqrt{2} \approx 2.82843$, which is just a shade larger than $e \approx 2.71828$. It's amazing how close our analysis has brought us to Stirling; we're in striking distance of it in fact!

We end this section on elementary questions with a few things for you to try.

Can you modify the above argument to get a reasonably good upper bound for $n!$?

After reading the above argument, you should be wondering exactly how far can we push things. What if we didn't do a dyadic decomposition; what if instead we did say a triadic: $(2n/3, n]$, $(4n/9, 2n/3]$, ... Or perhaps fix a r and look at $(rn, n]$, $(r^2n, rn]$, ... for some universal constant r. Using this and the pairing method described above, what is the largest lower bound attainable. In other words, what value of r maximizes the lower bound for the product.

2.5.2. *Lower bounds towards Stirling, II.* We continue seeing just how far we can push elementary arguments. Of course, in some sense there is no need to do this; there are more powerful approaches that yield better results with less work. As this is true, we're left with the natural, nagging question: *why spend time reading this?*

There are several reasons for giving these arguments. Even though they are weaker than what we can prove, they need less machinery. To prove Stirling's formula, or good bounds towards it, requires results from calculus, real and complex analysis; it's nice to see what we can do just from basic properties of the integers. Second, there are numerous problems where we just need some simple bound. By carefully going through these pages, you'll get a sense of how to generate such elementary bounds, which we hope will help you in something later in life.

Again, the rest of the material in this subsection is advanced and not needed in the rest of the book. You may safely skip it, but we urge you to at least skim these arguments.

We now generalize our argument showing that $n! > (n/4)^n$ for $n = 2^N$ to any integer n; in other words, it was harmless assuming n had the special form $n = 2^N$. Suppose $2^k < n < 2^{k+1}$. Then we can write $n = 2^k + m$ for some positive $m < 2^k$, and use our previous result to conclude:

$$n! \;=\; n \cdot (n-1) \cdots (2^k+1) \cdot (2^k)! \;>\; (2^k)^m \cdot (2^k)! \;>\; (2^k)^m \cdot (2^k/4)^{2^k}.$$

Our goal, then, is to prove that this quantity is greater than $(n/4)^n$. Here's one possible method: write

$$2^{km} \cdot (2^k/4)^{2^k} \;=\; (n/4)^\alpha.$$

If $\alpha > n$, then we're done. Taking logarithms, we find:

$$k \cdot m \cdot \log 2 + 2^k \cdot \log(2)(k-2) \;=\; \alpha(\log(n) - 2\log 2).$$

Solving for α gives

$$\alpha \;=\; \frac{k \cdot m \cdot \log 2 + 2^k \cdot \log(2)(k-2)}{\log(n) - 2\log 2}.$$

Remember, we want to show that $\alpha > n$. Substituting in our prior expression $n = 2^k + m$, this is equivalent to showing

$$\frac{k \cdot m \cdot \log 2 + 2^k \cdot \log(2)(k-2)}{\log(2^k+m) - 2\log 2} \;>\; 2^k + m.$$

So long as $2^k + m > 4$, the denominator is positive, so we may multiply through without altering the inequality:

$$\log(2)(k(2^k+m) - 2^{k+1}) \;>\; (2^k+m)\log(2^k+m) - \log(2)2^{k+1} - 2m\log 2$$

With a bit of algebra, we can turn this into a nicer expression:

$$\begin{aligned}
\log(2^k)(2^k+m) &\;>\; (2^k+m)(\log(2^k+m) - 2m\log 2 \\
2m\log 2 &\;>\; (2^k+m)\log(1+m/2^k) \\
2\log 2 &\;>\; (1+2^k/m)\log(1+m/2^k).
\end{aligned}$$

Let's write $t = m/2^k$. Then showing that $\alpha > n$ is equivalent to showing

$$2\log 2 \;>\; (1+1/t)\log(1+t)$$

for $t \in (0,1)$. Why $(0,1)$? Since we know $0 < m < 2^k$, then $0 < m/2^k < 1$, so t is always between 0 and 1. While we're only really interested in whether this equation holds when t is of the form $m/2^k$, if we can prove it for all t in (0,1), then it will automatically hold for the special values we care about. Letting $f(t) = (1+1/t)\log(1+t)$, we see $f'(t) = (t - \log(1+t))/t^2$, which is positive for all $t > 0$ (fun exercise: show that the limit as t approaches 0 of $f'(t)$ is $1/2$). Since $f(1) = 2\log 2$, we see that $f(t) < 2\log 2$ for all $t \in (0,1)$. Therefore $\alpha > n$, so $n! > (n/4)^n$ for all integer n.

2.5.3. *Lower bounds towards Stirling, III. Again, this subsection may safely be skipped; it's the last in our chain of seeing just how far elementary arguments can be pushed. Reading this is a great way to see how to do such arguments, and if you continue in probability and mathematics there is a good chance you'll have to argue along these lines some day.*

We've given a few proofs now showing that $n! > (n/4)^n$ for any integer n. However, we know that Stirling's formula tells us that $n! > (n/e)^n$. Why have we been messing around with 4, then, and where does e come into play? The following sketch doesn't *prove* that $n! > (n/e)^n$, but hints suggestively that e might come enter into our equations.

In our previous arguments we've taken n and then broken the number line up into the following intervals: $\{[n, n/2), [n/2, n/4), \dots\}$. The issue with this approach is that $[n, n/2)$ is a pretty big interval, so we lose a fair amount of information by approximating $n \cdot (n-1) \cdots \frac{n}{2}$ by $(n/2)^{n/2}$. It would be better if we could use a smaller interval. Therefore, let's think about using some ratio $r < 1$, and suppose $n = (1/r)^k$. We would like to divide the number line into $\{[n, rn), [rn, r^2 n), \dots\}$, although the problem we run into is that $r^\ell n$ isn't always going to be an integer for every integer $\ell < k$. Putting that issue aside for now (*this is why this isn't a proof!*), let's proceed as we typically do: having broken up the number line, we want to say that $n!$ is greater than the product of the smallest number in each interval raised to the number of integers in that interval:

$$n! \ > \ (rn)^{(1-r)n}(r^2n)^{r\cdot(1-r)n} \cdot (r^3n)^{r^2\cdot(1-r)n} \cdots (r^k \cdot n)^{r^{k-1}\cdot(1-r)n}.$$

Since $r^{k+m}n < 1$ for all $m > 1$, we can extend this product to infinity:

$$n! \ > \ (rn)^{(1-r)n}(r^2n)^{r\cdot(1-r)n} \cdot (r^3n)^{r^2\cdot(1-r)n} \cdots (r^k \cdot n)^{r^{k-1}\cdot(1-r)n} \cdots .$$

Let's simplify this a bit. Looking at the n terms, we have

$$n^{(1-r+r-r^2+r^2-\cdots)n} \ = \ n^n$$

because the sum telescopes. Looking at the r terms we see

$$\begin{aligned} r^{n(1-r)(1+2r+3r^2+\cdots)} &= r^{n(1-r)/r(r+2r^2+3r^3+\cdots)} \\ &= r^{n(1-r)/r\cdot r/(1-r)^2} \\ &= r^{n/(1-r)}, \end{aligned}$$

where in the third step we use the identity

$$\sum_{k=1}^{\infty} kr^k \ = \ \frac{r}{(1-r)^2};$$

remember we used this identity earlier as well! Combining the two terms, we have

$$n! \ > \ (r^{1/(1-r)}n)^n.$$

To make this inequality as strong as possible, we want to find the largest possible value of $r^{1/(1-r)}$ for $r \in (0,1)$. Substituting $z = 1/(1-r)$, this becomes: what is the limit as $z \to \infty$ of $(1-1/z)^z$? We've encountered this limit before in Section **REFERENCE**, and the answer is $1/e$ (of course!). Thus we see that this argument gives a heuristic proof that $n! > (n/e)^n$.

2.6. **Stationary Phase and Stirling.** Any result as important as Stirling's formula deserves multiple proofs. The proof below is a modification from Eric W. Weisstein's post "Stirling's Approximation" on *MathWorld*–A Wolfram Web Resource; see

`http://mathworld.wolfram.com/StirlingsApproximation.html`.

To prove the theorem, we'll use the identity

$$n! \ = \ \Gamma(n+1) \ = \ \int_0^\infty e^{-x}x^n\,dx. \tag{4}$$

A review of the Gamma function, including a proof of this identity, can be found **ADD REF**.

In order to get an approximation, we would like to find where the integrand is largest. Because of the exponential factor in the integrand, we will take the logarithm before differentiating. We can do this as maximizing a positive function $f(x)$ is equivalent to maximizing $\log f(x)$. There are many useful versions of this principle: in calculus it is often easier to minimize the square of the distance rather than the distance (as this avoids the square-root function).

We find

$$\frac{d}{dx}\log(e^{-x}x^n) \ = \ \frac{d}{dx}(-x+n\log x) \ = \ \frac{n}{x}-1.$$

The maximum value of the integrand is therefore seen to occur only for $x = n$. The exponential factor shrinks much more quickly than the growth of x^n, so we *assume* that $x = n+\alpha$ with $|\alpha|$ much smaller then n. We are not going to say anything further about this assumption; the purpose here is to provide another argument in support of Stirling's formula, not to provide the proof in all its glory and gory. We have

$$\log x \ = \ \log(n+\alpha) \ = \ \log n + \log\left(1+\frac{\alpha}{n}\right).$$

We now expand the second term using the Taylor series for $\log(1+x)$ to find

$$\log(n+\alpha) \;=\; \log n + \frac{\alpha}{n} - \frac{1}{2}\frac{\alpha^2}{n^2} + \cdots .$$

Therefore

$$\begin{aligned}\log(x^n e^{-x}) &= n\log x - x \;\approx\; n\left(\log n + \frac{\alpha}{n} - \frac{1}{2}\frac{\alpha^2}{n^2}\right) - (n+\alpha)\\ &= n\log n - n - \frac{\alpha^2}{2n^2}.\end{aligned}$$

It follows that

$$x^n e^{-x} \;\approx\; \exp\left(n\log n - n - \frac{\alpha^2}{2n^2}\right) \;=\; n^n e^{-n}\cdot\exp\left(-\frac{\alpha^2}{2n^2}\right).$$

Returning to the integral expression for $n!$ of (4), we have

$$\begin{aligned} n! &= \int_0^\infty e^{-x}x^n\,dx\\ &\approx \int_{-n}^\infty n^n e^{-n}\cdot\exp\left(-\frac{\alpha^2}{2n^2}\right)\,d\alpha\\ &\approx n^n e^{-n}\cdot\int_{-\infty}^\infty \exp\left(-\frac{\alpha^2}{2n^2}\right)\,d\alpha.\end{aligned}$$

In the last step, we rely on the fact that the integrand is very small for $\alpha < -n$. The integral is the the same as the one we would obtain in integrating a normal density with mean 0 and variance $\sqrt{n}$. Its value is $\sqrt{2\pi n}$. We thus have

$$n! \;\approx\; n^n e^{-n}\sqrt{2\pi n},$$

which is the statement of the theorem. □

2.7. **The Central Limit Theorem and Stirling.** We end this chapter with one more proof of Stirling's formula. We continue our trend of needing more and more input. Our first proof was very elementary, essentially just the integral test. The second was more complicated, needing the Gamma function. Our final approach involves an application of the Central Limit Theorem, one of the gems of probability. We prove the Central Limit Theorem in **ADD REF**; if you haven't seen the proof you can go to that section, or take the result on faith for now and read up on it later.

The idea of the proof is to apply the Central Limit Theorem to a sum of independent, identically distributed Poisson random variables with parameter 1. We have an explicit formula for this probability, as a sum of appropriately normalized Poissonian random variables is itself Poissonian. We also know what this probability is (or at least approximately is) by the Central Limit Theorem. Equating the two gives Stirling's formula.

Remember

(1) X has a Poisson distribution with parameter λ means

$$\text{Prob}(X=n) \;=\; \begin{cases}\frac{\lambda^n e^{-\lambda}}{n!} & \text{if } n\ge 0 \text{ is an integer}\\ 0 & \text{otherwise.}\end{cases}$$

(2) If $X_1,\dots,X_N$ are independent, identically distributed random variables with mean μ, variance σ^2 and a little more (such as the third moment is finite, or the moment generating function exists), then $X_1+\cdots+X_N$ converges to being normally distributed with mean $n\mu$ and variance $n\sigma^2$.

Let's flesh out the details. An important and useful fact about Poisson random variables is that the sum of n independent identically distributed Poisson random variables with parameter λ is a Poisson random variable with parameter $n\lambda$. As the mass function for a Poisson random variable Y with parameter λ is $\text{Prob}(Y = m) = \lambda^m e^{-\lambda}/m!$ for m a non-negative integer and 0 otherwise. Thus the probability density of $X_1 + \cdots + X_n$ is

$$f(m) \;=\; \begin{cases} n^m e^{-n}/m! & \text{if } m \text{ is a non-negative integer} \\ 0 & \text{otherwise.} \end{cases}$$

For n large, $X_1 + \cdots + X_n$ (in addition to being a Poisson random variable with parameter n) is by the Central Limit Theorem approximately normal with mean $n \cdot 1$ and variance n (as the mean and variance of a Poisson random variable with parameter λ is λ for each). We must be a bit careful due to the discreteness of the values taken on by $X_1 + \cdots + X_n$; however, a little inspection shows that the Central Limit Theorem allows us to approximate the probability $n - \frac{1}{2} \le X_1 + \cdots + X_n \le n + \frac{1}{2}$ with

$$\int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \frac{1}{\sqrt{2\pi n}} \, \exp\left(-\left(x-n\right)^2/2n\right) dx \;=\; \frac{1}{\sqrt{2\pi n}} \int_{-1/2}^{1/2} e^{-t^2/2n} dt.$$

As n is large, we may approximate $e^{-t^2/2n}$ with the zeroth term of its Taylor series expansion about $t = 0$, which is 1. Thus

$$\text{Prob}\left(n - \frac{1}{2} \le X_1 + \cdots + X_n \le n + \frac{1}{2}\right) \;\approx\; \frac{1}{\sqrt{2\pi n}} \cdot 1 \cdot 1,$$

where the second 1 is from the length of the interval; however, we can easily calculate the left hand side, as this is just the probability our Poisson random variable $X_1 + \cdots + X_n$ with parameter n takes on the value n; this is $n^n e^{-n}/n!$. We thus find

$$\frac{n^n e^{-n}}{n!} \;\approx\; \frac{1}{\sqrt{2\pi n}} \quad \Longrightarrow \quad n! \;\approx\; n^n e^{-n}\sqrt{2\pi n}.$$

One of the most common mistakes in this approach is forgetting that the Poisson is discrete and the standard normal is continuous. Thus, to approximate the Poisson's mass at n, we should integrate the continuous density of the standard normal from $n - 1/2$ to $n + 1/2$. It's just fortuitous that, for this problem, we get the same answer if we forget about the integral.