The Laws of Probability[1]

Part 1: What Makes a Coin Fair?
by Elizabeth S. Meckes

Everyone knows what it means to toss a fair coin, right? It means that it's equally likely to land on heads or tails. But what does that really mean? You toss it once, it lands on heads, so what? Is it fair? Is it unfair? How do you know?

At this point, you're probably going to tell me that I should toss it a bunch of times. If it lands on heads every time, we're pretty sure it's not a fair coin. We know what should happen, and it pushes us a little closer to knowing what fair means: if we toss the coin a lot of times, we should get about equal numbers of heads and tails. And that's perfectly fine for a Saturday afternoon, but not very satisfying to a mathematician.

There's a big difference between what we mean when we talk about "laws" in physics and when we talk about "laws" in mathematics. In physics, we're trying to describe the reality that we see, and to do it accurately enough to be able to make valid predictions. But in math, even though we often start with real, physical observations like coin tosses, our mindset is different. We want to come up with some axioms (statements we will assume) which seem reasonable based on our observations and are as simple as possible; then we want to see how much we can prove. Our goal is to start from these very simple assumptions, things we feel comfortable assuming, and prove that the more complicated things we think we've observed follow just from those axioms.

Understanding what a fair coin is is a great way to see the difference between mathematical and physical laws at work. The idea that I can't predict whether the coin lands on heads or tails is very hard to turn into a mathematical axiom; it's not even clear how to test it by experiment. The suggestion I imagined you making before, that I should check fairness by tossing the coin a lot, led us to the general idea that a coin is fair if when you toss it a lot of times, it lands on heads about half the time. But that's still awfully fuzzy. We could make it sound a bit math-ier by saying that if H_n is the number of times out of n tosses that the coin lands on heads, then we should have $\lim_{n\to\infty} H_n / n = 1/2$. But really I'm just conning you with fancy language and notation. If I toss the coin n times, I get a certain number for H_n. And then if I do it again, I get a different number: H_n is random! Even if I could toss a coin an infinite number of times in order to take the limit, how do I know I'd get the same thing if I did the whole process again?

The answer that probabilists have settled on is that going through limits is a bad way to define fair. Instead, we assume that we can assign numbers called "probabilities" to events in a way that satisfy a small set of axioms which are so simple and so intuitively reasonable that we don't mind taking them as a starting point. Then, we prove the limiting statement above: that if you toss the coin a lot of times, the limiting proportion of times it lands on heads tends to 1/2.

So, what are these axioms? The first one is that I can assign a numerical probability, which I'll call $\mathbf{P}(E)$ to any **event** E. Sticking just with coin tossing, an event is anything I can describe in terms of the outcomes of a series of coin tosses. So E could be the event that the first three tosses are heads, heads, tails. Or it could be that the seventh toss is tails. Or it could be that every other toss is a heads (forever – this is math, so I can have an infinite sequence of tosses). I moreover assume that for any event E, $\mathbf{P}(E)$ is between 0 and 1 (including possibly 0 or 1). For example, if E is the event that the first toss is heads, and I'm trying to talk about a fair coin, then $\mathbf{P}(E)$ should be 1/2.

[1] This content was supported in part by a grant from MathWorks.

My second axiom is very simple: if E is just the event that something, anything, happens, then $\mathbf{P}(E) = 1$. And here's the third and final one, which is as complicated as it gets: if I have a bunch of different events $E_1, E_2, \ldots$ with no overlap, then I can figure out the probability that one of them happens by adding up the individual probabilities. This has to work even if there are infinitely many E_k.

And that's it. Those are the properties that something I call probability has to have. Now, back to our fair coin. Like we said above, if E is the event that the first toss is heads, then $\mathbf{P}(E)$ should be 1/2. And if E_2 is the event that the second toss is heads, then $\mathbf{P}(E_2)$ should be 1/2. And so on; each individual toss should be equally likely to be heads or tails. But there's one other important feature of a fair coin: **independence**. How the toss came out on the first try shouldn't tell you anything about what's going to happen next, and vice versa. For our coin tossing, this means that all of the possible strings of outcomes of a given length should be equally likely: e.g., the first three trials have eight total possible outcomes, as shown at right, and each has probability 1/8.

Phew. Okay, now we really know what a fair coin is. So what about tossing it a lot of times? We can start from just the three axioms above and prove what's called the strong law of large numbers. In symbols, if H_n is the number of heads in the first n tosses of a fair coin, then the strong law of large numbers says that

$$\mathbf{P}\left[\lim_{n\to\infty}\frac{H_n}{n}=\frac{1}{2}\right]=1.$$

What this means is that it's essentially certain that in an infinite sequence of independent tosses of a fair coin, the limiting proportion of heads would be 1/2. I really have to have that cheater word "essentially" there: it's of course possible that the limit might be something else (or even not exist). In principle, I could toss a fair coin forever and get heads every single time. But what the strong law of large numbers says is that the probability that that will happen is zero. It's not that it can't happen, but it won't.

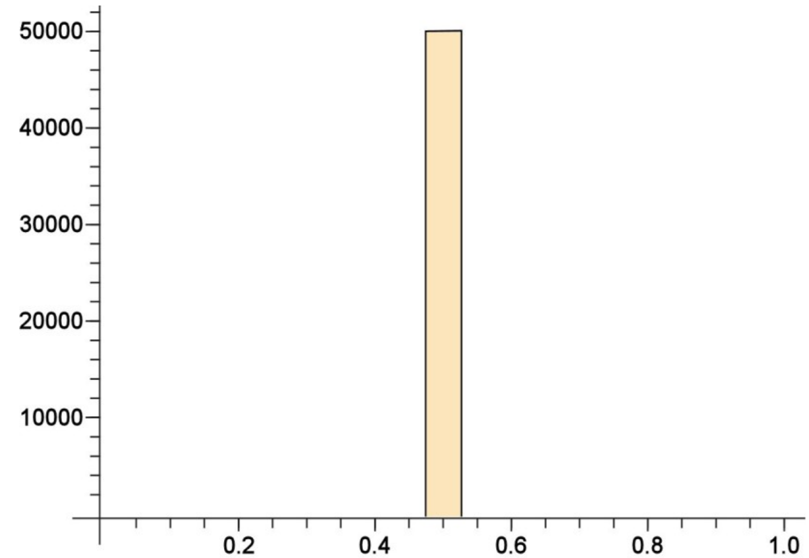
So caveats and technicalities aside, modern mathematics has triumphed: we can start with very simple, very reasonable assumptions about how anything called probability should work, and our intuition about what fairness should mean becomes a theorem we can prove.

The Laws of Probability[2]

Part 2: Zooming In.

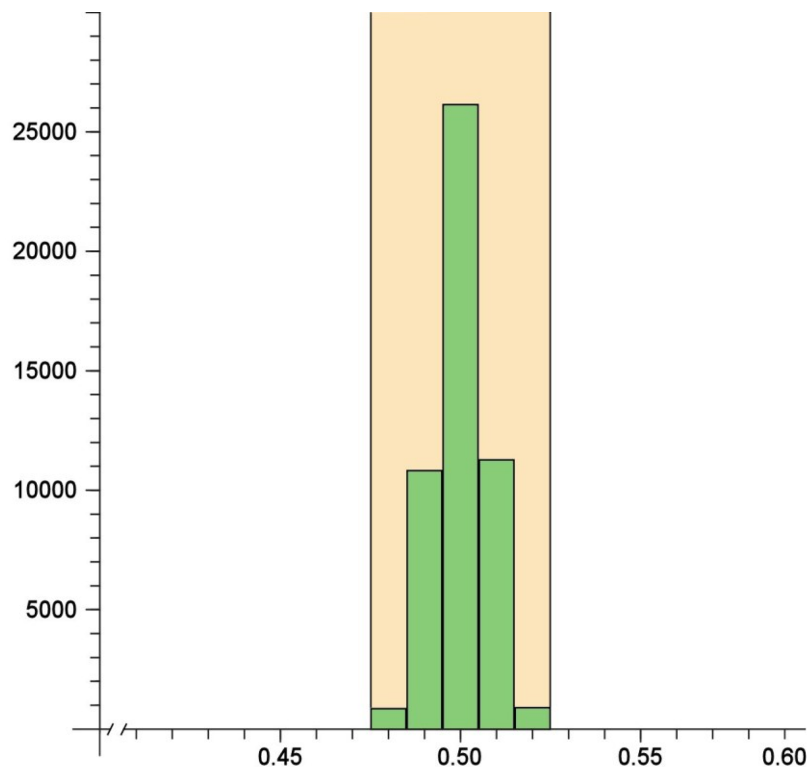
by Elizabeth S. Meckes

Last time we talked about how mathematicians define a fair coin, and the theorem that the limiting proportion of heads when a fair coin is tossed a lot of times is 1/2. Here's a computer simulation of that effect: in the histogram below, the computer did the experiment of tossing a coin 5000 times and counting the proportion of heads. It did this experiment 50,000 times (so the computer tossed 250,000,000 coins), and the histogram below shows the proportions of heads in the 50,000 experiments.



At this resolution, you see exactly what anyone would expect: a big bar at 1/2, meaning that the proportion of heads was between .475 and .525 every time.

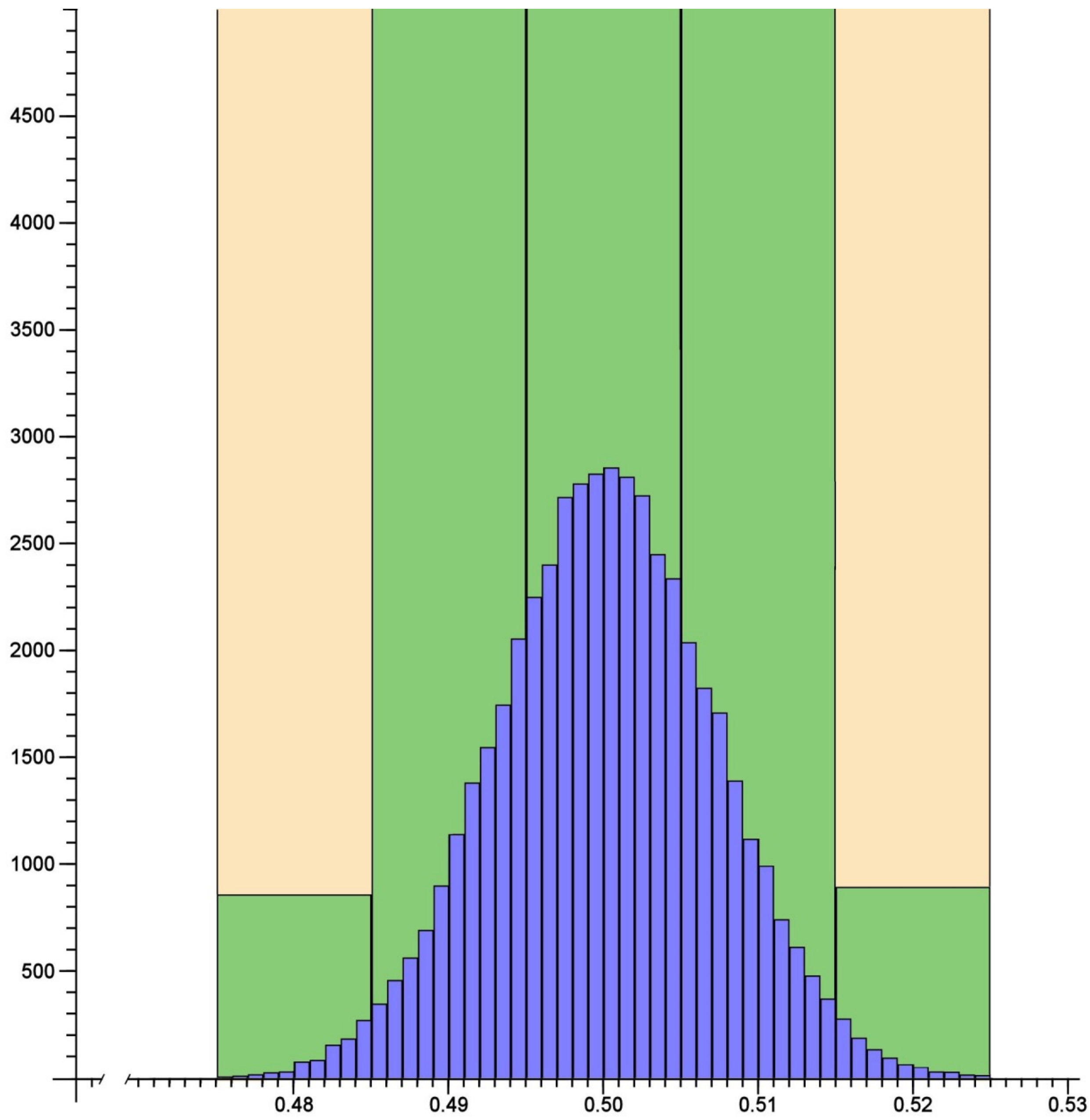
Let's zoom in; we'll make a histogram (bottom right) of what's going on inside that one big bar. This represents exactly the same set of coin tosses, but now the bars correspond to counting points within intervals of length .01, whereas before the bars covered intervals of length .05.



As you can see, it's still reassuringly concentrated around 1/2: there still don't seem to be any outcomes with fewer than 47.5% or more than 52.5% heads. But, because we've zoomed in enough, we're starting to be able to observe the fact that we're not going to get *exactly* half heads and half tails. We can see that the picture looks reasonably symmetric, which seems natural enough: it's just as likely to get a few more heads than tails as vice versa. We can also see that the bars drop off from the one that covers 1/2; the uneven outcomes got more unlikely as they got more uneven.

[2] This content was supported in part by a grant from MathWorks.

Okay, now let's really zoom in:



Whoa! Amazing! That beautiful curve was hidden in that first picture, but because we were looking at a coarse scale, we didn't see it.

The curve you see there is what people often call the "bell-shaped curve" (because it looks like the outline of a bell), or sometimes, the Gaussian curve, for the mathematician Carl Friedrich Gauss. He wasn't the first to see it, though – this curve has shown up as scientists observed and recorded findings about a huge number of features of the natural world. And lurking behind that ubiquitous curve is the **central limit theorem**.

Remember the law of large numbers from last time: if H_n is the number of heads in n tosses of a fair coin, then

$$\mathbf{P}\left[\lim_{n\to\infty}\frac{H_n}{n}=\frac{1}{2}\right]=1.$$

That theorem basically corresponds to the first picture. To zoom in, the first thing we have to do is to center our attention at 1/2. Just like when you zoom in on an online map, you need to move the spot you're interested in to the center; here we move the big bar in the first picture to the center of the number line (namely, 0), and consider

$$\frac{H_n}{n} - \frac{1}{2}$$

Now that we've centered things, we have to figure out how to turn the idea of zooming in into a mathematical operation we can do. Each time we zoomed in, we looked at a shorter interval on the *x*-axis, but we stretched it out so we could see it, making it have the same physical length as the original histogram. That means that we *multiplied* by larger and larger numbers, but only looked at the interval around our point of interest that fit the physical width of our original histogram. It turns out that for H_n, the right zoom factor is $\sqrt{n}$, and so finally the central limit theorem is a theorem about the random variables

$$X_n = \sqrt{n}\left(\frac{H_n}{n} - \frac{1}{2}\right)$$

Specifically, the central limit theorem says that X_n *converges weakly to the standard Gaussian distribution*, as n tends to infinity. There's a specific technical meaning to that, of course, but what it really means is that as n and the number of experiments we do grow, when we plot histograms at the kind of scale that we did in the third picture, they smooth out into the bell-shaped curve.

So why does all this mean that the bell-shaped curve occurs so often in nature? Well, one of the things about the central limit theorem (and many other theorems in probability) is that it is quite robust, in the sense that if you change the hypotheses a little, the same basic result is still true. In fact, you can push pretty much every aspect of the situation I've described above. The coin tosses don't have to be fair, and they don't even have to be unfair in the same way. Each coin can land on heads with a different probability, as long as those probabilities aren't too different (and they shouldn't be 1 or 0 – no two-headed or two-tailed coins!). The coin tosses don't need to be genuinely independent, either; as long as they don't depend too much on each other, we're okay. And actually, the coin tosses don't even have to be coin tosses! The random variable H_n can be a sum of weakly dependent random variables with pretty much any description you like – they can even be random vectors living in infinite-dimensional spaces! But back down to earth – why do we see the bell-shaped curve in observational data?

Think of it this way: any time you have a feature of something, say the weight of an adult female rabbit, it depends on a lot of little things. For (ridiculously over-simplified) instance, set the hypothetical rabbit's weight at the average weight of an adult female rabbit, but then add on an ounce if its mother was particularly large and take one away if she was particularly small. Now do the same for the father. Now either add or subtract an ounce depending on whether that rabbit's neighborhood is particularly full of lettuce or rather lacking in it. And so on and so forth; the weight of the rabbit is now seen as a random variable which is a sum of a lot of (basically independent??) things, and so the distribution of rabbit weights should look like a bell-shaped curve around the average.