Lecture 7

Statistics

If you pull any stray coin from your pocket, it's a safe bet that it is a fair coin—that is, a coin for which the probability of landing heads or tails up are both .5 (also expressible as 1/2 or 50%). It becomes painfully obvious after flipping the coin a few times that the coin's fairness *does not* guarantee that exactly half of the flips will land heads up and half will land tails up. As a somewhat silly illustration of this, note that if you flip the coin an odd number of times, the closest you can get to having equal numbers of each outcome will still have them off by 1 (e.g., 7 flips yields 4 heads and 3 tails). More significantly, there is nothing about a fair coin that prevents it from going on runs—repeated outcomes—of heads or tails. If you flip the coin ten times, you might get ten heads. The probability is small ($.5^{10} \approx .001$), but not *that* small. We can expect to see a run of ten heads about once in every 1000 sets of ten flips.

So, if saying that the probabilities of getting heads or tails when you flip a fair coin are both .5 doesn't mean that you'll get half heads and half tails whenever you flip a fair coin, what does it mean? One answer to that question appeals to what is called the "Law of Large Numbers", which says, roughly:

Law of Large Numbers (LLN). If $p(E) = n$, then, as the number of trials goes to infinity, the relative frequency of E can be expected to approach n.

Relative frequency here means the fraction of trials on which the event occurs to the total number of trials.

Going back to our coin flipping, we've already observed that for a given set of flips (aka. *trials*) there is no guarantee that exactly half of the outcomes will be heads even if the coin is fair. However, what the LLN tells us is that we can expect, *as our number of flips goes to infinity*, to see the fraction of heads-up flips approach 1/2.

You can kind of see why this should be the case by looking at the possible outcomes of progressively more flips. If you flip a coin once, it is certain that the relative frequency of heads to is either 1 or 0—far from the 1/2 suggested by the coin's fairness. If we flip the coin twice, we have three possible relative frequencies for heads: 1, 0, and 1/2. But note their respective probabilities. There are four possible outcomes for two coin flips and two of those (namely HT and TH) yield a relative frequency of heads equal to 1/2. This means that the probability that you get a sequence of outcomes with a relative frequency of heads that matches the chance of heads is 1/2 (since it happens in 2 of the 4 possible sequences), which is a lot better than that the certainty when flipping once that the frequency and chance won't match. If you flip a coin four times there are 16 possible sequences, and we can group them according to their frequencies of heads—indicated as $f(\text{H})$ to distinguish it from the probability/chance of heads—as in the figure below.

		TTHH		
		THTH		
	THHH	THHT	TTTH	
	HTHH	HTTH	TTHT	
	HHTH	HTHT	THTT	
HHHH	HHHT	HHTT	HTTT	TTTT
$f(\text{H}) = 1$	$f(\text{H}) = \frac{3}{4}$	$f(\text{H}) = \frac{1}{2}$	$f(\text{H}) = \frac{1}{4}$	$f(\text{H}) = 0$

In six of the sixteen possible sequences we get that a frequency that matches the chance, so the probability of that matching—$p(f(\text{H}) = p(\text{H}))$—is 3/8. This is less than the 1/2 when we were flipping the coin twice, so have things gotten worse? Not quite. At the very least, the chance matching frequency is the most probable since $p(f(\text{H}) = 1) = p(f(\text{H}) = 0) = 1/16$ and $p(f(\text{H}) = 3/4) = p(f(\text{H}) = 1/4) = 1/4$. Furthermore, the probability that you get the extremal frequencies of 0 or 1 is *much* smaller. When flipping twice, the probability of getting an extremal frequency was 1/2. After flipping four times, that has dropped to only 1/8. The more likely frequencies of 3/4 and 1/4 (which, together, are even more likely than the chance matching frequency of 1/2) are at least *close* to 1/2 (compared to the extremal frequencies of 0 and 1).

As we flip the coin more times, the probability of getting $f(\text{H}) = 1/2$ will go down. But it will always be the most likely frequency. And, more importantly, the *total probability* of getting frequencies close to 1/2 will go up and the total probability of getting frequencies far from 1/2 will go down. We can see this evolution in the figure below that groups outcomes of five flips by their relative frequency of heads.

		HHHTT	HHTTT		
		HHTHT	HTHTT		
		HTHHT	THHTT		
		THHHT	HTTHT		
		HHTTH	THTHT		
	HHHHT	HTHTH	TTHHT	HTTTT	
	THHHH	TTHHH	HTTTH	THTTT	
	HTHHH	THTHH	THTTH	TTHTT	
	HHTHH	HTTHH	TTHTH	TTTHT	
HHHHH	HHHTH	THHTH	TTTHH	TTTTH	TTTTT

In this case, of course, the probability of getting a relative frequency of heads equal to 1/2 is 0. But the probability of getting the nearest possible frequencies (3/5 and 2/5) is a whopping 5/8. And the probability of getting the extremal frequencies has fallen farther to 1/16.

In the discussion so far, much has been made of the idea of *closeness*. We said things like "[the frequencies] are at least *close*", "the nearest possible frequencies", and "extremal frequencies", all with an eye towards describing how close or far the values of the discussed frequencies of heads are to the underlying probability/chance of heads (namely, 1/2). We're being vague, and that should be corrected.

When we talk about closeness in this context, what we specifically mean is *closeness to the truth*. How close to the *true* probability is the *actual* frequency? More generally, how close to the *expected* outcome is the *observed* outcome?

One way of measuring this (and there are many) is the by looking at the weighted average of the square of the difference between the observed and expected values of an experiment. We call this a χ^2 statistic (pronounced "ki" (as in kite) "squared"):

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Where i indicates possible outcomes of each trial in the experiment, O_i indicates the number of times the i outcome was observed to have occurred, E_i indicates the number of times the i outcome was expected to occur, and Σ_i indicates that you should sum over all the possible values of i.

To be very clear with an example: We can think of all our our coin flipping as an experiment investigating the fairness of the coin. Suppose that we flip a coin 1000 times and get 502 heads and 498 tails. The possible outcomes of each trial in our experiment are "heads" or "tails", so those are the two possible values of i. Since we flipped a coin 1000 times and we expect it to be fair, the expected outcomes of heads and tails are both 500—that is, $E_{\text{heads}} = E_{\text{tails}} = 500$. This initial assumption, the one that gives us our expected values, is often called the *null hypothesis*. Our observed outcomes, as already noted, are $O_{\text{heads}} = 502$ and $O_{\text{tails}} = 498$. Our χ^2 statistic, then, is

$$\begin{aligned} \chi^2 &= \frac{(O_{\text{heads}} - E_{\text{heads}})^2}{E_{\text{heads}}} + \frac{(O_{\text{tails}} - E_{\text{tails}})^2}{E_{\text{tails}}} \\ &= \frac{(502 - 500)^2}{500} + \frac{(498 - 500)^2}{500} = \frac{2^2}{500} + \frac{(-2)^2}{500} = 2 \times \frac{4}{500} = .016 \end{aligned}$$

So, is the coin fair? To answer that question we have to check a table:

d.f. \ p	.995	.99	.975	.95	.9	.1	.05	.025	.01	.005
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75

To use this table we need to know the *degrees of freedom* (d.f.) in our experiment. The degrees of freedom in an experiment is, for our purposes, the number of possible outcomes of a trial minus 1. So, in our current example, there is 1 degree of freedom because there are two possible outcomes of a trial (heads or tails) and $2-1=1$. The more proper way to think about degrees of freedom in this case—though don't worry about internalizing this—is to recognize that that the probabilities of the outcomes must sum to 1, $1 = p(\text{heads}) + p(\text{tails})$, and so if we imagine one of the probabilities is *free* to vary, then the other is constrained by the need to satisfy the requirement that they sum to 1.

So our degrees of freedom are 1. This means we want to look at the first row (one down from the top) of the table. We then trace over to the right until we find the first number bigger than our χ^2. In this case, that's .02 in the column under $p = .9$. The χ^2 *test* (as opposed to statistic) is the process of finding that p-value and interpreting it. What the p-value is is the probability that you are wrong if you reject the null hypothesis. In our case, if we reject our null hypothesis—our initial assumption that the coin is fair—there is about a 90% chance that we are wrong. (Because our chart only goes out to two decimal places, there are minor errors. Using a p-value calculator like the one at `graphpad.com/quickcalcs/pvalue1.cfm` gives us the more precise $p = 0.8993$.)

So is the coin fair? We can be pretty confident that it is. Or, at least, we can be pretty confident we'd be wrong if we assumed it wasn't. It is often assumed that a *significant difference* between the observed data and the expected/null hypothesis is had when $p \leq .005$. Only if you have such a significant difference should the null hypothesis be rejected. But note two very important things: (1) The threshold of $p \leq .005$ isn't special. You could easily insist that the difference has to be greater (i.e. p has to be smaller) or permit a lesser (i.e. p can be larger) to count as significant. You would need to revise how confident you are about any claims you made based on your conclusion, but you should probably be keeping track of how confident you are of those claims anyway. (2) Rejecting the null hypothesis is *just* that. *It does not provide you with a new hypothesis to adopt.* If it turned out that we should reject our null hypothesis that the coin is fair, the χ^2 test doesn't tell us anything about what we should now believe about the probabilities of the coin landing heads or tails.