Sampling Distributions

The Long Run Distribution of a Proportion

What happens when you repeat an experiment over and over again? A good way to understand this is to think again about flipping an ordinary coin a large number of times and counting the number of heads you obtain. Intuitively you would expect the fraction of heads you obtain to get closer and closer to 0.5 as the coin is flipped more times. What happens to the distribution of the proportion of heads?

Example 1

Let's begin with an experiment in which we flip a coin twice. Figure 1 is a graph of the probabilities of the fractions of heads you may get. For a quarter of the experiments in which we flip two coins we will not obtain any heads, half of the time we will observe a single heads, and a quarter of the time we will get two heads.

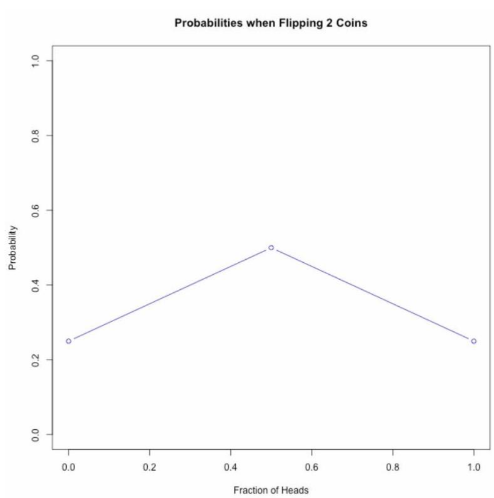

Figure 1: Probabilities when Flipping 2 Coins

Now, let's imagine flipping three coins. For 1/8th of the time we will not get heads, 3/8th of the time there will be one head, 3/8th of the time there will be two heads, and 1/8th of the time there will be three heads. (See Figure 2.)

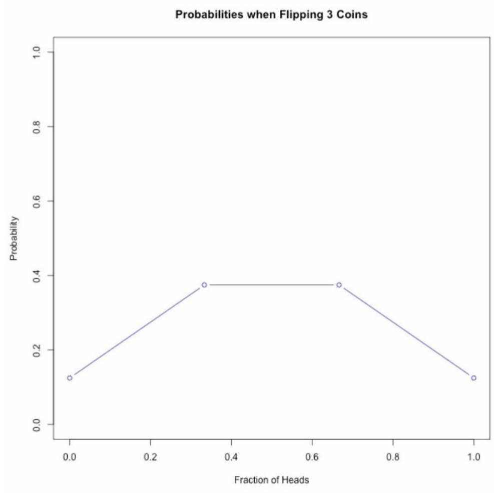

Figure 2: Probabilities when Flipping 3 Coins

What if we flip a coin 10 times? From the graph in Figure 3 we can see that the fractions of heads that are most likely are all close to 0.5 or 5 heads out of 10 flips. Getting 4 or 6 heads of 10 tries are also quite likely. Getting no heads at all or only one head, or getting 9 heads or 10 heads out of 10 are all very unlikely outcomes.

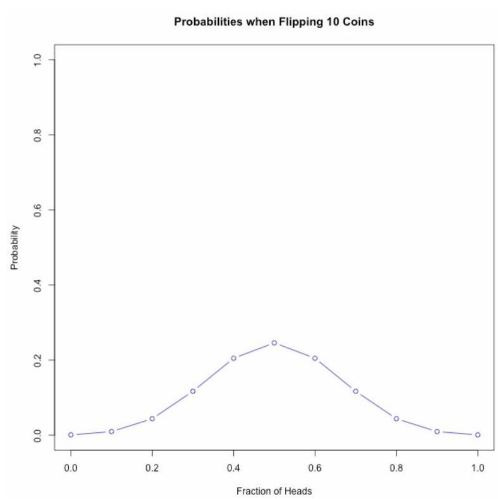

Figure 3: Probabilities when Flipping 10 Coins

If we flip even more coins, we can see this pattern more clearly. Figure 4 is a graph of the probabilities when we flip 30 coins.

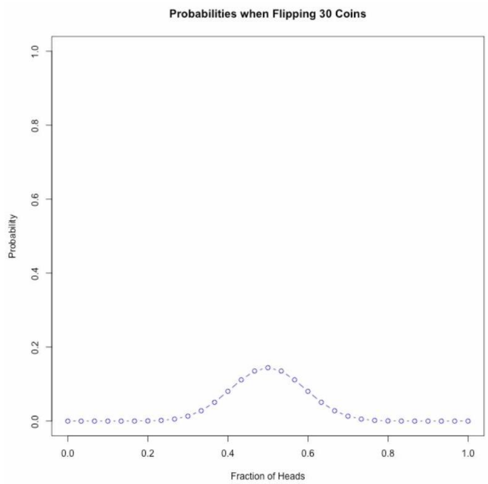

Figure 4: Probabilities when Flipping 30 Coins

Again, we see that the most likely outcomes are those in which about half of the flips are heads – those with probabilities of 0.5. We can also start to notice something about the shape of the graph of these probabilities. Does this shape look familiar?

Figure 5 is a graph of the probabilities when we flip 100 coins. The first thing we see is that all the individual probabilities are getting quite small. This is because there are a large number of possible results you can get for the fraction of heads when you flip 100 coins.

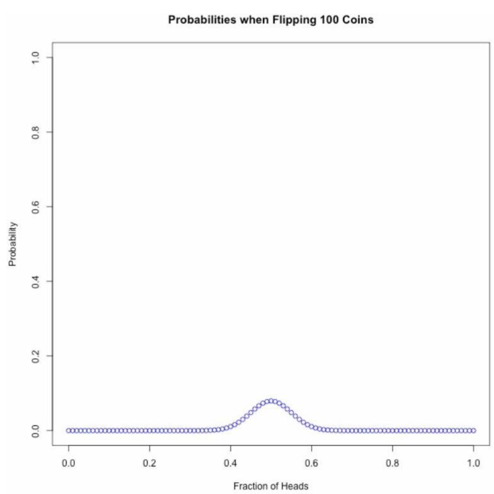

Figure 5: Probabilities when Flipping 100 Coins

To see the shape more clearly, Figure 6 is another version of the same graph stretched out so we can see the probabilities better.

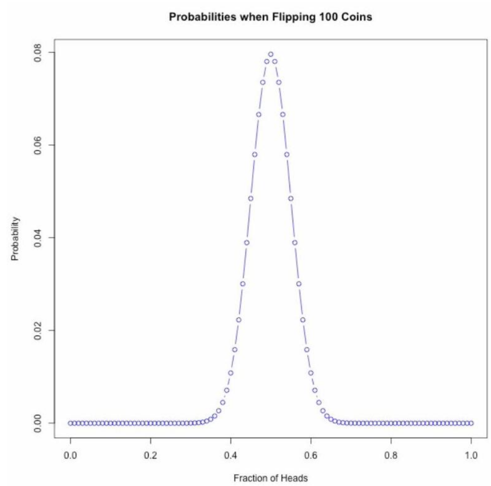

Figure 6: Probabilities when Flipping 100 Coins close up

The curved shape of this distribution should remind you of the probability density function for the Normal distribution that we saw previously. So, what can we say so far? We have seen that as we flip more and more coins, the fraction of heads seems to get closer and closer to being 1/2, which is what we expect intuitively. And, when we look at the distribution of the fraction of heads that we can get, it looks more and more bell-shaped like the density of the Normal distribution.

General conclusion:

Central Limit Theorem (CLT): If an experiment is repeated over and over, then the probabilities for the average result, or the proportion of successes, will converge to a Normal distribution.

The Central Limit Theorem is usually stated in terms of averages. Why does it also hold for proportions? A proportion is the count of the number of successes in n trials divided by n. If the outcome of each trial is thought of as an observation of a Bernoulli random variable which is 1 if the outcome of the trial is a success and 0 otherwise, then the sum of the outcomes of the n Bernoulli random variables is the number of successes, and the average of the n Bernoulli random variables is the proportion of successes.