Statistical Thinking for the 21st Century

Copyright 2020 Russell A. Poldrack

Chapter 6

Probability

6.2.2 Empirical probability

Another way to determine the probability of an event is to do an experiment and determine the probability. **From the relative frequency of the different outcomes we observe in our experiment, we can compute the empirical probability.**

Empirical means "based on observation or experience." Therefore, an empirical probability is the probability based on observation or experience.

For example, let's say that we are interested in knowing the probability of rain in San Francisco. We can compute an empirical probability by observing the number of days that it rains.

According to the National Weather Service, in 2017 there were 73 rainy days. To compute the probability of rain in San Francisco, we simply divide the number of rainy days by the number of days counted (365), giving P(rain in SF in 2017) = 0.200. Our computation of probability is similar to our computation of relative frequency.

The rain in San Francisco example used a large sample of measurements; we measured the weather every day for a year. Therefore, our measurement of the probability of rain in San Francisco respected the *law of large numbers*,

The law of large numbers states that **the empirical probability will approach the base-rate probability as the sample size increases.**

We can demonstrate the law of large numbers by simulating a large number of coin flips, and looking at the probability of heads after each flip. The left panel of Figure 6.1 shows that **when the number of samples** (i.e., the number of coin flip trials) **increases, the empirical probability approaches the true or base-rate probability** of 0.500. However, **when the number of samples** (i.e., the number of coin flip trials) **is small, the empirical probability can be very far off from the base-rate probability**.

A real-world example of this was seen in the 2017 special election for the US Senate in Georgia, which pitted the Republican Roy Moore against Democrat Doug Jones. The right panel of Figure 6.1 shows the votes reported for each of the candidates over the course of the evening, as an increasing number of ballots were counted.

Early in the evening the vote counts were especially unreliable, swinging from a large initial lead for Jones to a long period where Moore had the lead, until finally Jones took the lead to win the race.

These two examples show that while large samples will ultimately converge on the base-rate probability, the results with small samples can be far off. Unfortunately, many people forget this and over interpret results from small samples.

The tendency of over-interpreting results from small samples was referred to as the *law of small numbers* by the psychologists Danny Kahneman and Amos Tversky, who showed that people (even trained researchers) often behave as if the law of large numbers applies even to small samples, giving too much credence to results from small datasets. We will see examples throughout the course of just how unstable statistical results can be when they are generated on the basis of small samples.

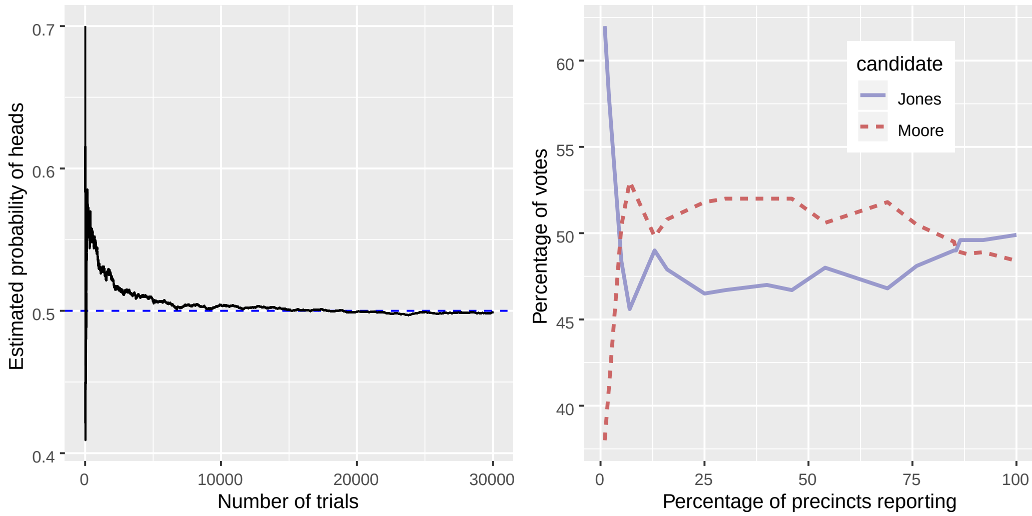

Figure 6.1:

Left: A demonstration of the law of large numbers. A coin was flipped 30,000 times, and after each flip the probability of heads was computed based on the number of heads and tail collected up to that point. **It takes about 15,000 flips for the probability to settle at the true probability of 0.500.**

Right: Relative proportion of the vote in the Dec 12, 2017 special election for the US Senate seat in Georgia, as a function of the percentage of precincts reporting. **It takes numerous precincts to report their votes before the probability settles to the final result.**
These data were transcribed from https://www.ajc.com/news/national/alabama-senate-race-live-updates-roy-moore-doug-jones/KPRfkdaweoiXICW3FHjXqI/