Notes: Hypothesis Testing, Fisher's Exact Test

Foundations of Data Analysis

February 12, 2019

These notes are an introduction to the **frequentist** approach to hypothesis testing, namely, the **null hypothesis statistical test**. We will cover what is known as the Fisher exact test, the first example of a null hypothesis statistical test.

The Lady Tasting Tea

Many of the modern principles used today for designing experiments and testing hypotheses were introduced by Ronald A. Fisher in his 1935 book *The Design of Experiments*. As the story goes, he came up with these ideas at a party where a woman claimed to be able to tell if a tea was prepared with milk added to the cup first or with milk added after the tea was poured. Fisher designed an experiment where the lady was presented with 8 cups of tea, 4 with milk first, 4 with tea first, in random order. She then tasted each cup and reported which four she thought had milk added first. Now the question Fisher asked is, "how do we test whether she really is skilled at this or if she's just guessing?"

To do this, Fisher introduced the idea of a **null hypothesis**, which can be thought of as a "default position" or "the status quo" where nothing very interesting is happening. In the lady tasting tea experiment, the null hypothesis was that the lady could not really tell the difference between teas, and she is just guessing. Now, the idea of hypothesis testing is to attempt to *disprove* or *reject* the null hypothesis, or more accurately, to see how much the data collected in the experiment provides evidence that the null hypothesis is false.
The idea is to *assume* the null hypothesis is true, i.e., that the lady is just guessing. Under this assumption and given the outcome of the experiment, we can now compute the probability of her performing as well as she did or better. Let's see how this works with an example outcome. Let's assume the lady gets all 8 cups correct. We can build a table of this outcome (this is called a **contigency table**):

		Lady's Answer: Milk First	Lady's Answer: Tea First
Truth	Milk First	4	0
	Tea First	0	4

Under the null hypothesis assumption (that she is guessing), what is the probability of this outcome? The lady knows there are exactly 4 cups of each, so she is essentially choosing 4 cups at random out of 8. There are "8 choose 4" ways to do this, so her probability is

$$P(\text{"all correct"}) = \frac{1}{\text{"number of ways to guess"}} = \frac{1}{\binom{8}{4}} = \frac{1}{70} \approx 0.014.$$

So, if she is guessing, there is only a 1.4% chance that she will get all cups correct. Now, let's look at the general situation. Notice that if we set how many cups with milk first that she gets correct, this determines the entire table. This is because she knows to choose 4 cups in each category, and thus each row and each column must sum to 4. The table for the general outcome looks like this:

		Lady's Answer	
		Milk First	Tea First
Truth	Milk First	k	4 - k
	Tea First	4 - k	k

First, notice that correct answers are on the diagonal. So, a value of k means that the lady actually has $2k$ answers correct. Here the counting problem to compute the probability of a general outcome is more difficult, but it follows what is called a **hypergeometric distribution**.

$$p(k) = \frac{\binom{4}{k}\binom{4}{4-k}}{\binom{8}{4}} = \frac{1}{70}\binom{4}{k}^2.$$

Now, this is the probability of the lady getting *exactly* $2k$ answers correct. What we originally wanted to ask is "what is the probability of her getting this outcome *or better*?" To get this, we need to sum over all values k or greater (up to the max of 4). Letting X be the total number of correct answers, this is:

$$P(\text{"}2k\text{ correct or better"}) = P(X \geq 2k) = \sum_{i=k}^{4} p(i).$$

Here are the probabilities of the 5 possible outcomes for the experiment:

$$p(0) = \frac{1}{70}, \quad p(1) = \frac{16}{70}, \quad p(2) = \frac{36}{70}, \quad p(3) = \frac{16}{70}, \quad p(4) = \frac{1}{70}.$$

Notice the symmetric in the probabilities. It is just as hard to get all wrong as it is to get all correct! Finally, the probabilities for getting $2k$ correct answers or better are

$$P(X \geq 0) = 1, \quad P(X \geq 2) = \frac{69}{70}, \quad P(X \geq 4) = \frac{53}{70}, \quad P(X \geq 6) = \frac{17}{70}, \quad P(X \geq 8) = \frac{1}{70}.$$

Now let's think about the more general case. Let's say we had N cups of tea, K cups with milk first. The lady selects n cups as ones she thinks are with milk first, getting k correct. The chances of getting these k "successes" by just choosing at random with uniform probability is

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

This is the general formula for the hypergeometric distribution. Of course, this distribution works for problems beyond tea tasting. It is applicable anytime you have a fixed number of binary values, and you want to know the probability of correctly selecting some of them by chance.

By the way, according to the legend, the lady got all 8 cups correct!

Summary of General Hypothesis Test Procedure:

1. Define the **null hypothesis**, which is the uninteresting or default explanation.
2. Assume that the null hypothesis is true, and determine the probability rules for the possible outcomes of the experiment.
3. After collecting data, compute the probability of the final outcome or even more extreme outcomes.

Further reading:

Ronald Fisher: `http://en.wikipedia.org/wiki/Ronald_A._Fisher`
Fisher's Exact Test: `http://en.wikipedia.org/wiki/Fisher's_exact_test`
Hypergeometric Distribution: `http://en.wikipedia.org/wiki/Hypergeometric_distribution`