Binomial Tests

Galvanize

Objectives

- Describe the null hypothesis testing procedure.

- Define a p-value.

- Setup a null hypothesis for a binomial test.

- Calculate a p-value for a binomial test.

Probability vs. Statistics

Probability and Statistics are closely related subjects, but there is a fundamental difference.

Probability

In probability we know the parameters of a distribution (associated with some random variable), and we would like to study properties of data generated from that distribution.

Example properties of random variables are:

The expectation of a random variable is defined by:

$$E[X] = \int_{-\infty}^{\infty} t\ f_X(t)dt$$

and the variance is defined by

$$V[X] = E\left[(X - E[X])^2\right]$$

If you know the parameters of the distribution, then you can compute the mean and variance.

Statistics

In **statistics** we have data generated from a random variable, and we would like to *infer* properties of its distribution.

A few points are evident:

- Independent and identically distributed data are important, as they allow us to pool information using data all generated from *indistinguishable* random variables.
- We can never know *exactly* the distribution that generated the data, we can only hope to approximate it.
- We *may* be able to quantify the uncertainty in our approximation (this is what much of classical statistics is about).

Fisher's Tea Experiment

Ronald Fisher's friend Muriel Bristol claims that she can tell, by actually drinking the beverage, whether milk was poured in first or second into a cup of tea. I.e. tea into the milk, or milk into the tea.

Fisher, being an upstanding skeptic, is skeptical, so devises an experiment to test her claim.

Discussion: How could we determine whether Muriel is telling the truth.

Fisher's Tea Experiment

Fisher's solution is as follows. He prepares six cups of tea, three with tea first and three with milk first. These cups are then given to Muriel arranged in a random order.

Note: For simplicity, I'll assume in my version that Muriel does *not know* that there are three of each.

He has Muriel drink each beverage, and attempt to guess if tea or milk were poured first. Her results are as follows

Cup	Result
1	Correct
2	Correct
3	Correct
4	Incorrect
5	Correct
6	Correct

Fisher's Tea Experiment

Fisher, remember, is a skeptic. He is predisposed to **not** believe Muriel. So he makes the skeptical hypothesis:

> Fisher's Hypothesis: Muriel has *no* ability to tell milk into tea from tea into milk. Therefore, each of her answers is completely random.

Fisher now asks himself an important question:

> Given that my hypothesis is correct, **how surprising is this data?**

Fisher's Tea Experiment

Let's take Fisher's hypothesis seriously. If he is correct, what should we expect the data we observe to look like?

Well, under this hypothesis, the chances of Muriel getting any single cup correctly is 0.5.

That is, if Fisher is correct, the data we observed would be generated by a simple sequence of coin flips.

Question: What is the distribution of the number of correct answers by Muriel under Fisher's hypothesis?

Fisher's Tea Experiment

If Fisher is correct, and Muriel's answers are totally random, then the distribution of the number of cups Muriel gets correct is a Binomial distribution

$$\text{\# of cups guessed correctly} \sim \text{Binomial}(n=6, p=0.5)$$

Fisher's Tea Experiment

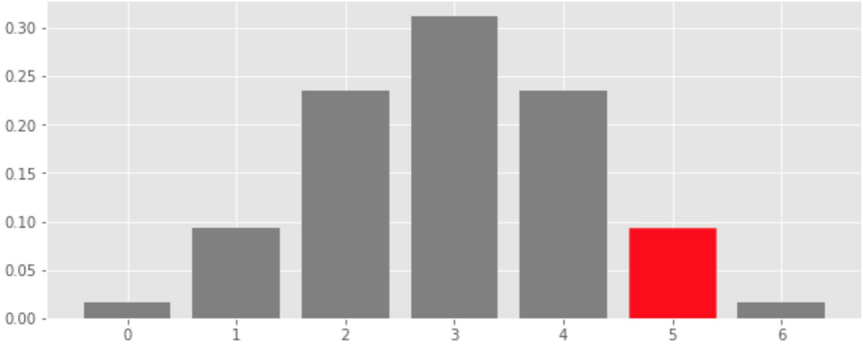
Muriel guessing all but one correctly seems like an extremely surprising event given this assumption. We can characterize our **degree of surprise in the data given our assumption** by calculating the following probability:

$$P(\text{Observing this Data} \mid \text{Fisher's Assumption})$$

In our example, we can actually calculate this probability exactly from the binomial distribution.

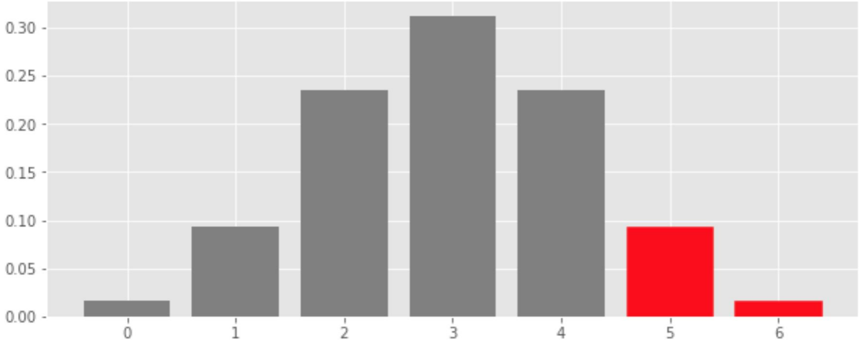
$$P(\text{Observing this Data} \mid \text{Fisher's Assumption}) = \binom{6}{5} 0.5^6$$

Fisher's Tea Experiment



Fisher observed that, in fact, he would have been surprised if he had observed this data **or any data even more extreme than this**, so a more correct measure of surprise would be

$$P(\text{Observing Data As Or More Extreme} \mid \text{Fisher's Assumption}) = \binom{6}{5}0.5^6 + \binom{6}{6}0.5^6$$



Probability of Observing Data More Equal or More Extreme than Actual: **0.11**

Do you believe Muriel?

Hypothesis Testing - Big Picture

The tea example was reported in Fisher's classic text *The Design of Experiments*, and it is a prototypical example of the logic behind Hypothesis testing.

State a scientific question - yes or no question

Take a skeptical stance, and clearly state this hypothesis - Null Hypothesis

State the opposite of your skeptical hypothesis - Alternative Hypothesis

Create a probabilistic model of the situation assuming the null hypothesis is true

Decide how surprised you need to be to reject your skeptical assumption

Collect your data

Calculate the probability of finding a result equally or more extreme than actually observed assuming the null hypothesis is true

Compare the p-value to your stated rejection threshold

Let's jump to another example with some code

See binomial-tests.ipynb