Hypothesis Testing with the Binomial Distribution

This is a revised version of the assignment sheet distributed on October 30, 2002. Problems 1 and 2 are due on November 4.

Introduction

In these exercises, we are going to apply a form of statistical reasoning called "hypothesis testing" to some data that the class generated. The basic form of reasoning goes as follows:

1) We make a hypothesis about the random and non-random factors that influence the data that we gather under specific conditions. This hypothesis will contain enough information for us to be able to compute the likelihoods of selected decisive properties in a typical data set.
2) We collect data, and examine what decisive properties it displays. If they are sufficiently unlikely under the hypothesis made in the first step, then we reject the hypothesis. (The meaning of "sufficiently unlikely" mean is determined by purposes for which you are gathering the data, or the context in which you report it.

Rather than going into detail on this point, let us examine an example of this kind of reasoning. The important thing to remember is that there are two steps: 1) formulating a hypothesis and making a probabilistic prediction and 2) examining data, and deciding how strongly it supports or contradicts the hypothesis.

The Question to be Answered

In a recent assignment, students were asked to submit a sequence of 300 't's and 'h's that was as close to random as they could produce by hand. Thirty-six students submitted work. We are going to look at only one feature of the work: the symbol that students chose to start with. **Did students choose this randomly, or was there a bias?**

To address this question, we will follow the pattern of reasoning outlined above. We will make the hypothesis that students were uninfluenced in choosing between 't' and 'h', so that each student had an equal likelihood of choosing one or the other. Based on this hypothesis, we can state the probabilities of various imbalances that may occur. For example, if every student was equally likely to choose 't' or 'h', then it would be truly extraordinary if all but one or two started with 't'. This would lead us to question the hypothesis.

In essence, "hypothesis testing" is not very much more complicated than this, *except* that we use probability theory to make precise statements about the likelihoods. Rather than having to say, "truly extraordinary," we get a precise measure of the improbability that anyone can compute and that everyone will understand the same way. The advantage of this is that we will not need to deal with different peoples estimations of what "truly extraordinary" means.

Looking at a Related Question

Before jumping in to the data, there is a small complication. Twenty-three students followed directions and submitted sequences of 't's and 'h's. Thirteen students submitted sequences of '0's and '1's instead. Students were not penalized for choosing numbers rather than letters, since the directions were given in a hurry at a time when students might not have been able to pay full attention. It's possible that the directions were not heard, or were heard by only a few people. (In fact, it's even possible that I said, "Either/or," as some students recall.)

The complication is lucky, in a way, because it gives us the opportunity to display an example illustrating hypothesis testing. We will examine the question of whether the choice between letters and numbers was influenced by any factors at all—the directions that I tried to give included. This illustrates the way the reasoning works.

We shall imagine a scenario that might have produced the data, and then we shall ask whether or not the data that was actually obtained could easily have arisen under the scenario. If the data is unlikely to have been created under the hypothetical scenario, then this is a good reason for rejecting the hypothesis and supposing that some other set of circumstances might have been responsible for creating the data.

As background, note that I have previously asked students to write sequences of the numbers 0 and 1. Therefore, it would not have been unreasonable for students to assume that '0's and '1's were also required on this assignment. On the other hand, we frequently used 't's and 'h's in examples discussed in class. Moreover, I did indicate that I preferred 't's and 'h's when I gave the assignment. We will assume that because of their prior experience in the class, students picked between letters and numbers, and did not consider using other symbols. This is consistent with the work submitted, since no one chose to use any other kind of symbol.

- **Hypothesis.** On the whole, all of the various factors that might have influenced the choice between letters and numbers cancelled one another out, so a randomly chosen student would be equally likely to choose letters or to choose numbers.

Now we ask how likely it is that that the results that we actually saw would have come about if the hypothesis were actually true. What catches our attention about the data is that there are far fewer submissions with numbers than there are with letters: 23 sequences with letters out of 36 responses. How likely would it be for an imbalance this great or greater to occur under the hypothesis? The question can be rephrased as follows. Assume that 36 people with no preference choose randomly between the alternatives, letters or numbers. *What is the probability that 13 or fewer choose numbers?*

To answer this we use the binomial coefficients. The number of ways for the 36 people to make a choice of either letters or numbers is $2^{36} = 68,719,476,736$. If people have no preference, then all these ways are equally likely. The number of ways for 13 or fewer people to choose numbers is

$$\binom{36}{0} + \binom{36}{1} + \binom{36}{2} + \cdots \; etc \; \cdots + \binom{36}{12} + \binom{36}{13}$$

This sum works out to $4,552,602,248$. Thus, under the hypothesis the probability of no more than 13 people choosing numbers is

$$\frac{4,552,602,248}{68,719,476,736} \cong 0.066 = 6.6\%.$$

This is a small probability. It suggests that people did not act randomly. Of course, it does not *prove* that they did not, but it tilts the balance toward that conclusion.

To understand more about what the 6.6% means, consider what we would have thought if 16 people chose numbers. Under the hypothesis, the probability of no more than 16 people choosing numbers rather than letters is about 31%. (We can also work this out with binomial coefficients, but I will not burden you with the calculation.) So, if 16 had chosen numbers rather than letters, we would *not* have a good reason for questioning our hypothesis. Of course, we would not conclude that people definitely did act without preference, but we would not be uncomfortable thinking this.

Problem 1.

Among the 23 who submitted letters, 17 began their sequence with a 't 'and 6 began with an 'h.' Is there a bias toward beginning with a 't'? Answering this question means testing the following hypothesis:

- **Hypothesis.** Students who wrote letters were equally likely to begin with a 't' or an 'h.'

Under the hypothesis, how likely would it be for us to see an imbalance as great as the one in our data? That is, how likely would it be for 6 or fewer people to have chosen to begin with h, if people picked between 't' and 'h' randomly?

Problem 2.

Among the 13 who submitted numbers, 9 began their sequence with a '0'and 4 began with an '1.' Is there a bias toward beginning with a '0'? Is the evidence for a bias in this case stronger or weaker than the evidence for a bias in the case of letters?

Problem 3.

At the beginning of this class, students were asked to begin a new sequence of letters. We gathered data onn how they began. Based on the new data, how strong is the evidence for a preference for beginning a sequence with 't' rather that 'h'? Record the data we collected below.

Total # students _____; # beginning with 't' _____; #beginning with 'h' _____.