THE PROBABILITY OF “HEADS” WHEN FLIPPING A COIN

BY JAMES BERNHARD

In this paper, we describe our test of whether flipping a particular coin in a certain way is fair. We begin by explaining the testing process, including how we collected the data. We then analyze the data, present the results of our test, and close by commenting on the strengths and weaknesses of our methods.

THE PROCESS

The coin whose flip we tested was a 1994 penny from our own personal collection in Tacoma, Washington. The flip was generated by a vertical flick of the thumb, after which the coin was projected upwards about 2 feet, completing at least four full rotations before being caught one-handed in mid-air. After the coin was caught in one hand, it was placed covered on the back of the catcher’s other hand and then revealed.

To address the question of the fairness of this flip, we first modeled the flip as a Bernoulli trial in which “heads” was success, and we let $X$ be the associated Bernoulli random variable. We denoted the probability that $X = 1$ by $\lambda$. This allowed us to translate the question of the flip’s fairness into the statistical question of whether $\lambda = 1/2$. To test for fairness, we conducted a hypothesis test with null and alternative hypotheses given by:

$$H_{null} : \lambda = 1/2 \quad \text{and} \quad H_{alt} : \lambda \neq 1/2.$$  

At the outset, we chose a significance level of 0.05, since this is traditional and we had no particular reason to vary it. This completed the Preliminaries step in the hypothesis test.

We then carried out the Data step in the hypothesis test. We decided to collect a sample of 16 coin flips. We did not carry out any formal computations to see how much of a deviation from fairness this would detect, but we hoped that it would at least be sufficient to detect large deviations from fairness. We collected our sample on February 8, 2012 in a private home in Tacoma, Washington. Our data is given in the first two columns of the accompanying Excel spreadsheet, Column A giving the
trial number and Column B the outcome. For the outcome, we denoted “heads” by “h” and “tails” by “h”.

**ANALYSIS OF THE DATA**

We next conducted the *Estimate* step of the desired hypothesis test. For this, we decided to use the usual 16-estimator $L$, which is the proportion of successes in a 16-sample of $X$. We have included our computations in the accompanying Excel spreadsheet.

In Cell E4, we computed the 16-estimate $\ell$, which is the value of $L$ on our particular 16-sample. We found this by dividing the number of successes in our sample by the sample size. We counted the number of successes in Cell E3 with the `COUNTIF()` function, and we used that the sample size was 16, recorded in Cell E2. This gave us that $\ell = 10/16 = 0.625$.

Proceeding to the *Distribution* step of the hypothesis test, we determined that since the 16-estimator was $L$, then its distribution under the null hypothesis was $B_{prop}[16, 1/2]$. In this expression, 16 is the sample size and $1/2$ is the value of $\lambda$ under the null hypothesis.

Because we would need the probabilities in this distribution to compute the $p$-value, we computed them in Column I of the Excel spreadsheet. We needed the number of successes in Column G because Excel doesn’t compute proportionalized binomial distributions, only binomial distributions. We set up Column I to use the `BINOM.DIST()` function applied to Column G to compute $B_{prop}[16, 1/2]$.

For the $p$-value step of the hypothesis test, we computed the probability that the 16-estimator $L$ would be at least as extreme as its value $\ell$ on our sample, which was $10/16$. Since the alternative hypothesis included both directions, the meaning of “at least as extreme” in this context was “at least as far from $1/2$”, the random variable mean of $L$ under the null hypothesis. Continuing to translate, we found that “at least as far as $10/16$ from $1/2$” meant “at least as far as $10/16$ from $8/16$”, meaning either less than or equal to $6/16$ or greater than or equal to $10/16$. Therefore the $p$-value was given by:

$$ p = \text{Prob}(L \leq 6/16 \text{ or } L \geq 10/16). $$
In Cell E5, we computed this probability by adding up all of the probabilities of the success proportions satisfying the specified condition, and we found that:

\[ p = 0.454. \]

Continuing to the *Interpretation* step of the hypothesis test, we found that the \( p \)-value 0.454 was not less than our chosen significance level 0.05. Therefore we did not have statistically significant evidence to reject the null hypothesis that \( \lambda = 1/2 \).

**CONCLUSIONS**

Translating the results of our significance test from statistical language back to what we were modeling, we did not find statistically significant evidence that this method of flipping this coin was unfair. In other words, as far as we could tell, this method of flipping this coin may well be fair. We did not find evidence to the contrary.

**ASSESSMENT**

One strength of our analysis was that we were able to control the process of flipping the coin fairly well. We were able to repeat this process reasonably consistently in a nearly constant environment. Differences in the flips due to variation in external conditions were probably fairly small.

However, a weakness of our method was that because we did not calculate the sample size needed to detect various amounts of deviation from fairness, we don’t know how large the deviation would have had to have been for us to detect it. Perhaps a larger sample size would have been helpful. We didn’t find evidence of unfairness, but because our sample size was so small, that might not be saying much.

Another weakness was that although the environment was fairly constant and we tried to flip the coin about the same way each time, we still might have varied our method of flipping inadvertently. Because of this, a Bernoulli model assuming a constant probability might not have been appropriate.

Both of these weaknesses could detract from or even invalidate our results, and in future tests we should try to eliminate or at least address them.