Experiment #5

Dylan Robertson
PHSX 815

March 27, 2023

Abstract

Struggling graduate students often look for many different forms of relief from the constant pressure of doing homework, studying for exams, and writing a thesis. Many of them turn to gambling, among other vices. However, not every gambler is an honest person, and somebody may be trying to swindle you out of your money so that they can go to Hyvee and buy some more premium coffee that they will need to motivate themselves to go into the lab. This project simulated rolling dice and compared two simple hypotheses in order to help you discern which of your friends is a lying S.O.B. trying to fund their way through grad school by taking everyone else's money, and who is an honest person that plays by the rules. The die was rolled 25 times per experiment, and 1000 experiments were simulated. The Log Likelyhood Ratio was calculated for each experiment and plotted. The statistical power of the test to discern between the two hypotheses was found to be 78%

1 Introduction

The null hypothesis, H_0, assumes that our friends are nice, honest people and states that the die is fairly weighted and equal for all sides: $p_0 = \frac{1}{6}$.

The alternative hypothesis, H_1, states that the dice is unfairly weighted to favor landing on a "1", "3", or a "6": $p_1 = 0.3$, $p_2 = 0.1$, $p_3 = 0.2$, $p_4 = 0.1$, $p_5 = 0.1$, $p_6 = 0.2$.

Each dice roll was done by pulling a random number from a uniform distribution in the interval $[0, 1)$ and then pulling from a categorical distribution.

This was repeated $N_{roll} = 25$ times in each experiment and for $N_{exp} = 1000$ experiments. This was repeated for each hypothesis.

2 Theory

The Categorical Distribution describes the distribution of outcomes for a single dice roll and has the form of a "pick me" function.

$$P(x|\vec{p}) = p_1^{x=1} p_2^{x=2} ... p_k^{x=k}, \tag{1}$$

where $\vec{p}$ is a vector of the different probabilities of $k-1$ outcomes, and the probability of $x = i$ is p_i. Hence you "pick out" the probability of the face that you are interested in rolling. The probability of the k^{th} outcome comes from the normalization requirement: $\sum_{i=1}^{k} p_i = 1$.

In general, the algorithm for a Categorical Distribution function works as follows. First, sample a random number R from a uniform distribution between 0 and 1. Then, assign intervals in this range to correspond to a "face" on the die based upon the probability of rolling each side.

The code that implements the Categorical Distribution for a 6-sided die can be seen in the file **Random.py** in the Github for Project 1.

```python
#create function that rolls a dice
    def Categorical(self, p1, p2, p3, p4, p5):
      R = self.rand(); #samples a random number R from a uniform
          distribution between 0 and 1.

      if R < p1:
        return 1
      if R < p1 + p2:
        return 2
      if R < p1 + p2 + p3:
        return 3
      if R < p1 + p2 + p3 + p4:
        return 4
      if R < p1 + p2 + p3 + p4 + p5:
        return 5
      else:
        return 6
```

For this simulation of a dice game (Yahtzee!, Monopoly, Beer die, etc.), a dice was rolled and the output of each roll was recorded. Each experiment consisted of rolling the dice N_{roll} number of times, and for N_{exp} number of experiments. In the following simulation, $N_{roll} = 25$, and $N_{exp} = 1000$.

When a dice is rolled multiple times in a single experiment, the distribution of outcomes follows a multinomial distribution.

$$P(\vec{N}|\vec{p}) = \frac{N_{tot}!}{N_1!...N_k!} p_1^{N_1}...p_k^{N_k}, \tag{2}$$

where $\vec{N}$ is a vector $(N_1, N_2, ..., N_k)$ defining how many times each face was rolled in the experiment. Note that $\vec{p}$ is unique to each hypothesis, and simply describes the probability of rolling each face under the given hypothesis.

The likelyhood ratio (LR) is how to compare which of the two hypotheses is favored by the data for a given experiment:

$$LR = \frac{P(x|H_0)}{P(x|H_1)}, \tag{3}$$

where "x" is the data obtained, namely $\vec{N}$ under a Hypothesis determined by $\vec{p}$.

Plugging in the multinomial distribution for the numerator and denominator yields,

$$LR = \frac{p_0^{N_{tot}}}{p_1^{N_1} p_2^{N_2} p_3^{N_3} p_4^{N_4} p_5^{N_5} p_6^{N_6}}, \tag{4}$$

since the normalization factors cancels, and all of the probabilities in the Null hypothesis are equal.

A more useful metric is the Log Likelyhood Ratio (LLR) which has the advantage of turning the above multiplication into addition:

$$\lambda_j = N_{tot} \log(p_0) - \sum_{i=1}^{6} N_i \log(p_i), \tag{5}$$

where λ_j is the Log Likelyhood Ratio of the j^{th} experiment. The value of p_i is determined by each hypothesis.

The above LLR is calculated for each experiment under the Null Hypothesis and added to an array, $(\lambda_1, \lambda_2, ..., \lambda_{N_{exp}})$, which is then sorted in ascending order.

The critical LLR value is defined by picking a confidence level for our simulation. Picking a confidence level is done by choosing a value $\alpha = \frac{1-CL}{100}$. For this simulation we will choose $\alpha = 0.05$ so that our Confidence Level is 95%.

We now search through our LLR array and find the critical value, λ_α, that corresponds to our confidence level. For a 95% confidence interval this would mean that 95% of LLR values are beneath the critical value.

The simulation is then repeated under the Alternative Hypothesis, and the Log Likelyhood Ratio is calculated for each experiment under the new hypothesis. This list is also sorted in ascending order. We then find λ_α in this second array and the percent of LLR values above it defines our β value. β is related to the power of test, which tells us how "good" our test is at distinguishing the two hypotheses for a given set of data.

$$Power = 1 - \beta. \tag{6}$$

3 Results

Using **DiceHist.py**, the results of the simulation for each Hypothesis can be seen in Figs. 1 and 2. Fig. 1 shows that each side of the dice is approximately equally likely to be rolled under the Null Hypothesis, while Fig. 2 shows that certain sides of the dice are heavily favored over others.

Normalized versions of these graphs can be seen in Figs. 3 and 4 and show the probability of rolling each side of the face under a given hypothesis.

Finally, using **DiceAnalysis.py**, the Log Likelyhood Ratio was calculated for each experiment under both hypothesis and the distribution can be seen in Fig. 5. The critical value was found to be $\lambda_\alpha = -1.1579700515665117$, which corresponded to a value of $\beta = 0.22$.

4 Summary

The above simulation now lets struggling graduate students hit the Vegas strip or play a game of die with other students and see if anyone is trying to cheat them our of their money (which they don't have a lot of). A given value of λ would tell you to either reject the Null Hypothesis or not by looking at

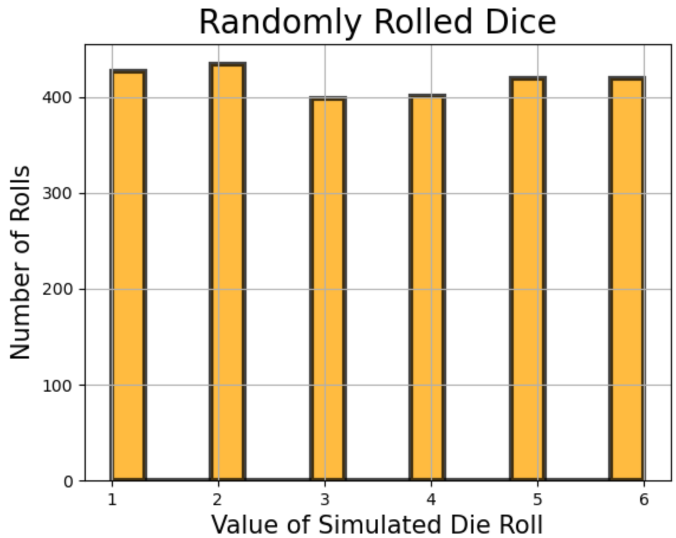

Figure 1: Counting the outcome of each dice roll under the Null Hypothesis. The dice was rolled 25 times per experiment, and was repeated for 100 experiments. $N_1 = 427$, $N_2 = 434$, $N_3 = 399$, $N_4 = 401$, $N_5 = 419$, $N_6 = 420$.

Fig. 5. Hopefully you don't lose any friends by applying your knowledge of statistics.

In order to improve the power of the test, you would want to simulate the experiment more times, as increasing the number of rolls per experiment is not practical.

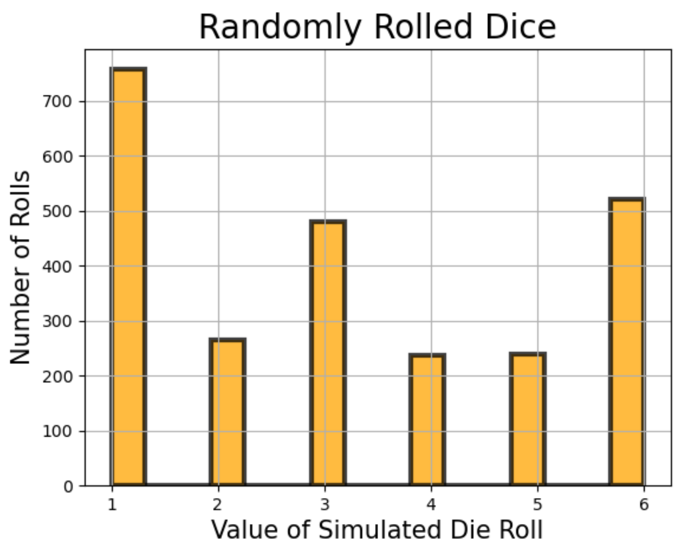

Figure 2: Counting the outcome of each dice roll under the Alternative Hypothesis. The dice was rolled 25 times per experiment, and was repeated for 100 experiments. $N_1 = 757$, $N_2 = 265$, $N_3 = 481$, $N_4 = 237$, $N_5 = 240$, $N_6 = 520$.

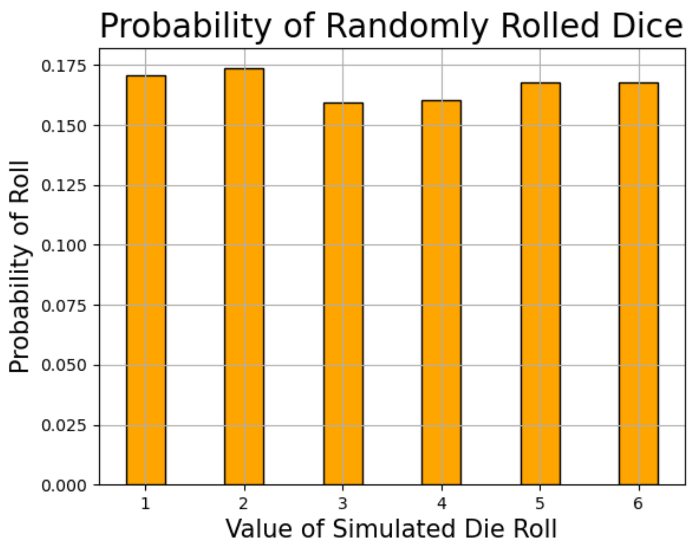

Figure 3: The probability of rolling each side of the under the Null Hypothesis. The dice was rolled 25 times per experiment, and was repeated for 100 experiments. $p_1 = 0.1708$, $p_2 = 0.1736$, $p_3 = 0.1596$, $p_4 = 0.1604$, $p_5 = 0.1676$, and $p_6 = 0.1680$.

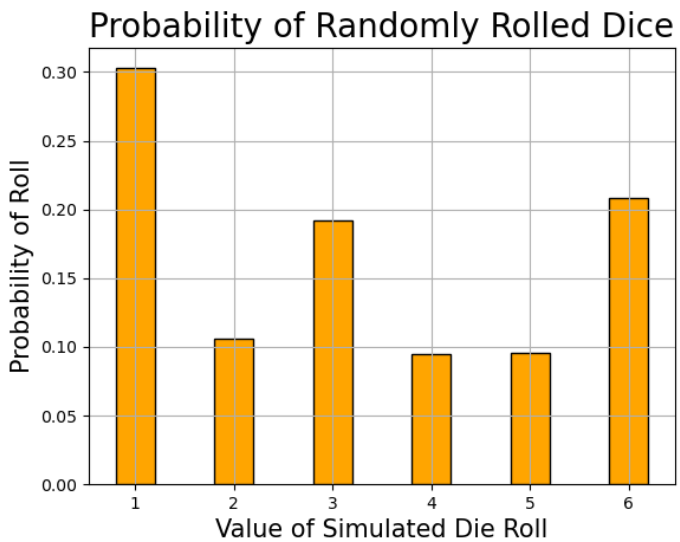

Figure 4: The probability of rolling each side of the under the Alternative Hypothesis. The dice was rolled 25 times per experiment, and was repeated for 100 experiments. $p_1 = 0.3028$, $p_2 = 0.1060$, $p_3 = 0.1924$, $p_4 = 0.0948$, $p_5 = 0.0960$, and $p_6 = 0.208$.

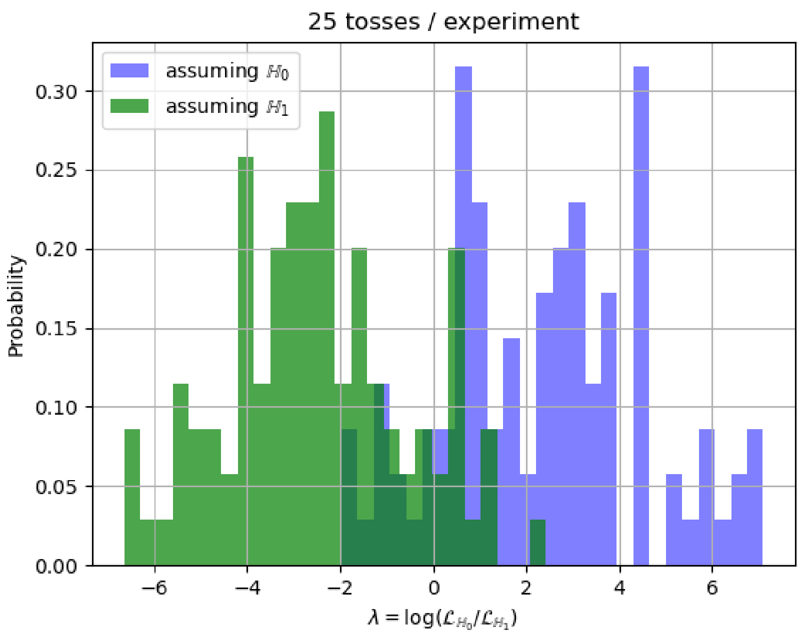

Figure 5: A plot of the log likelyhood ratios calculated for each experiment under both hypotheses. $\alpha = 0.05$ and $\beta = 0.22$. The power of the test was found to be 0.78.