`PHSX815_Project1`: Comparing the biased dice with fair ones and studying the distribution of their outcome

Kushal Rijal

February 2021

1 Introduction

When it comes to the game of dice (Ludo being played online), it is not surprising if one suspects the outcome of the dice roll. Considering the fact that it is always in favour of the developers if they can tweak the randomness of the outcome to make the game interesting. There are a lot of ways that can be achieved and one of the simplest way to do that is by designing the dice roll to favour the outcome of '6', since, this outcome has potential to bring the player lagging behind to come ahead and lead unexpectedly . In this project we are trying to find a method and design a test to identify a biased die at a certain confidence level ($\alpha = 5\,\%$).

This paper is organized as follows: Sec. 2 explains the hypotheses we are testing to see if our dice is fair or loaded. A description of the computer simulation developed to simulate these possibilities is provided in Sec. 3, with an analysis of the outputs included in Sec. 4. Finally, conclusions are presented in Sec. 5 .

2 Hypotheses to test identify biased dice

In order to identify if the Dice is fair or not, we will test the following two hypotheses:

1. The $P(outcome - 6 = 16.67\,\%)$ i.e H_0
2. The $P(outcome - 6 = 20\,\%)$ i.e H_1

3 Code and Experimental Simulation

First, we will deliberately design a die that will favour the outcome of 6 (assigned probability $20\,\%$). Then we want to design a test that will make sure that the higher frequency of outcome - 6 is not just by chance. In order to do this, we will simulate a fair die thousands of times and see, by directly comparing, how likely was our particular outcome that favoured 6. In this simulation we will use random number generator to generate a categorical distribution. We will find the p-value for our hypothesis and calculate the power of the test. Since log likelihood test is the most powerful test [1] that we can perform to test our hypothesis, in a different simulation, we want to use it to see if we can get better value for power of the test.

Most of the algorithms we are using in this project are similar to the ones we did in our class. But, for the first simulation, we will be using a different codes/algorithm. After noticing that we got 58 sixes in 250 trials, we will generate a distribution of different number of sixes obtained from fair dice. In order to do that, we will simulate thousands of experiment with 250 dice and see how many sixes we can get from each experiment. The result of the experiment is saved in a file. Then we will count the number of sixes in each experiment from the output file and plot the distribution. We will then fit this distribution with Gaussian curve fit to find the fitted value of mean and check what is the probability of getting 58 sixes from the 250 Dice.

4 Analysis

The bar diagram of the outcome of 250 trials is shown in Fig. 1. We can immediately see that the outcome of six is higher than other outcomes.

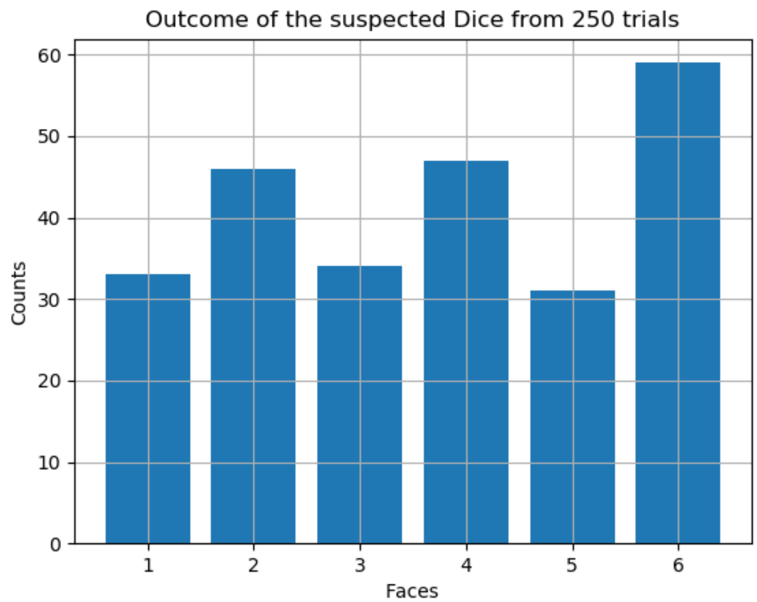

Figure 1: Outcome of 250 trials from the suspected dice. It can be seen that outcome 6 has poped up more than any other outcome.

If the die was equally weighted, we would expect that the number for each of these rolls would be about 42 as we rolled the die 250 times and each number should have a probability of $1/6$ of showing up. We notice that the number of 6's is especially large and would like to test if this could be due to chance. As we already know the probabilities of each number occurring in a fair die, we can simulate thousands of dice rolls and see how likely it is to observe 59 ones in a set of 250 rolls.

To get our statistical results we will use 10,000 trials. This is a distribution of the number of ones that we observe under our null hypothesis that each number is equally likely. The plots for the distribution and their respective gaussian fit are shown in the fig 2 , below.

In the case of 10,000 trials we can calculate the p-value for our test by counting the number of sixes that are greater than or equal to 59 and dividing that by the total number of trials. This represents the area to the left of the yellow dotted line in the plot. After calculating the z-score for X = 59, mean =

41.49 and standard deviation = 5.85 (obtained from fit), we find that the p-value is 0.0014 for the right tailed test which is much smaller than our α which is equal to 0.05.

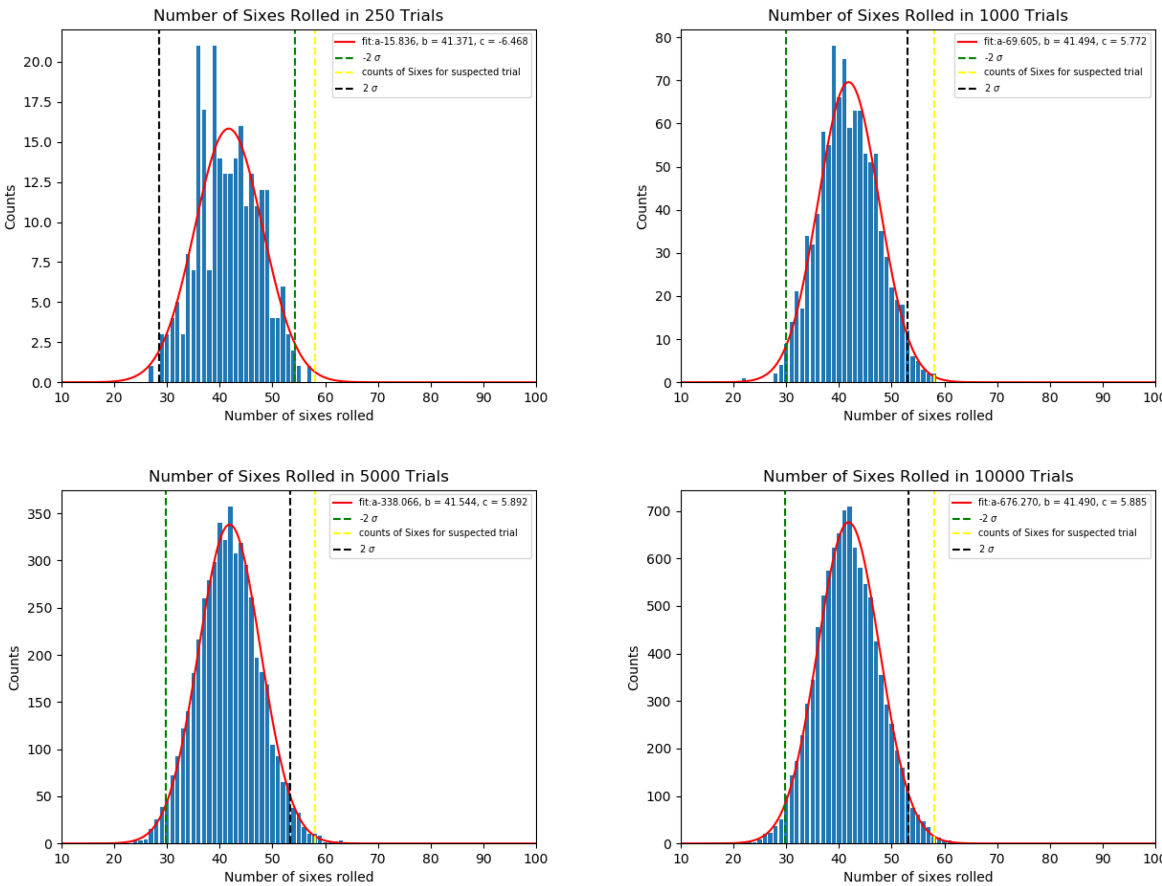

Figure 2: Simulations of the number of sixes Rolled for different number of experiments. We can immediately notice one can get rid of the noise in the data by doing more number of experiments. The parameters of the gaussian fit are shown in the legends where 'a', 'b' and 'c' represent amplitude, mean and standard deviation, respectively.

But p-value of the test is not always a good strategy to test the hypothesis. We should also calculate the power of the test. In order to do that we studied the Log-Likelihood ratios of the two hypotheses. The plot is shown in Fig. 3. From the figure we see that the power of our test for 10,000 experiments is 1 - β = 0.394 . Usually the power of test, if greater than 0.8, is considered a good one. However, in our case, it is fairly low as compared to 0.8. This means our probability of rejecting the null hypothesis when, in fact, it is false is low. To put that simply, our chance of committing type II error is high.

We simulated fair and unfair dice for different number of tosses and experiments under the null hypothesis that the dice is fair against the alternative hypothesis that the dice is biased (by assigning their respective probability to the fair and unfair dice). We can see from the plots below that our chance of committing type II error did not change much when we increased the number of experiments, but there was a notable change when we increased the number of dice roll for each experiment.

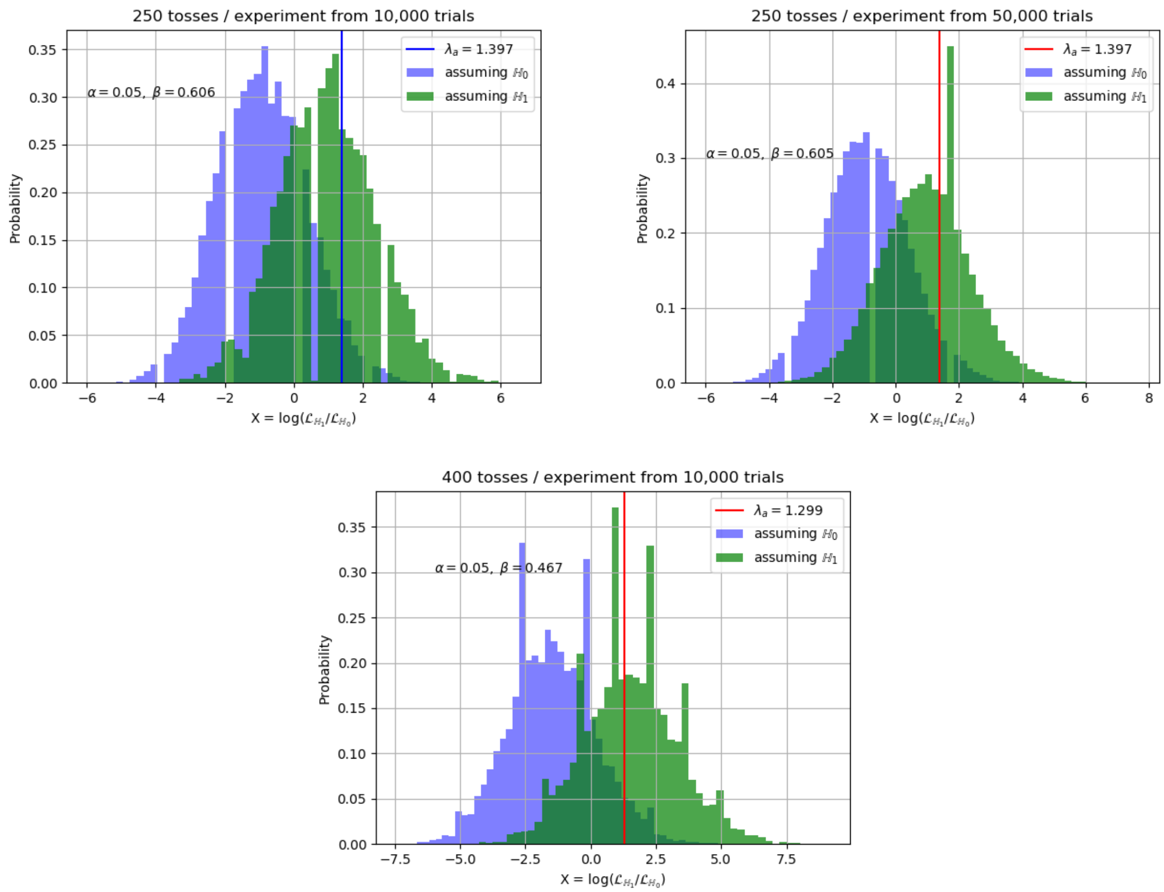

Figure 3: Power of test for different number of dice rolls and different number of experiment. λ_a is the critical value, α represents the significance level for our test and β is the type II error.

5 Conclusion

As a conclusion, since our p-value was quite small, we can assert that we have enough evidence that the dice used in the online game favours the outcome 6. So, we reject the null hypothesis and claim that the dice was indeed biased. However, from the log-likelihood test we noticed that our chance of committing the Type II error is significantly high with the simulated no. of events and dice rolls. Overall, one should conduct more number of trials and dice rolls in order to increase the power of our test.

References

[1] J. Neyman and E. S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231** no. 694-706, (1933) 289–337.