Summer 2022

Dice Testing with the Running Chi-Square Distribution

Warren Campbell
Civil Engineering, warren.campbell@wku.edu

Hunter Wimsatt
Western Kentucky University, hunter.wimsatt565@topper.wku.edu

Summer 2022

DICE TESTING WITH THE RUNNING CHI-SQUARE DISTRIBUTION

C. Warren Campbell
School of Engineering and Applied Sciences
Western Kentucky University, *warren.campbell@wku.edu*

Hunter A. Wimsatt
Western Kentucky University and Carol Martin Gatton Academy of Mathematics and Science,
hunter.wimsatt565@topper.wku.edu

Dice Testing with the Running Chi-Square Statistic

C. Warren Campbell
Kenneth E. and Irene S. Hall Professor of Civil Engineering
School of Engineering and Applied Sciences
Western Kentucky University
warren.campbell@wku.edu

Hunter A. Wimsatt
Carol Martin Gatton Academy of Mathematics and Science
Western Kentucky University

WKU® SCHOOL OF ENGINEERING & APPLIED SCIENCES

THE GATTON ACADEMY of Mathematics and Science

Dice Testing with the Running Chi-Square Statistic

Abstract:

Dice are not fair. Producing a geometrically precise, uniform die is not possible. Casino dice come as close as possible to perfectly random dice because they are machined to an accuracy of a few ten-thousandths of an inch and the putty used for the pips (dots) is the same density as the plastic die body. Polyhedral dice used in games like Dungeons and Dragons are far more difficult to manufacture to high tolerances. Some dice are fairer than others. Since 20-sided dice (D20s) are very difficult to manufacture precisely, they were the focus of this study. The running chi-square distribution (Campbell and Dolan 2019, 2020) was the tool of choice for these tests. The chi-square statistic has several interesting mathematical characteristics that are described. An unfair die will have a constant term and a linear trend added to the running chi-square statistic. The steeper the slope of the trend, the more unfair the die. 30 dice were rolled at least 3000 times each with an automated dice-rolling device, and some of the same dice were rolled 3000 times by hand in a dice tower. Results indicate that the method of rolling heavily influences the outcome.

Background:

Dice cannot be manufactured with absolute precision, so dice are not fair. Campbell and Dolan (2019) rolled several dice 3000 times and compared the results with float tests of some of the same dice. They concluded that while float tests may indicate that dice are not balanced, the float test sides that floated up are uncorrelated with the sides of dice that come up most often when physically rolled. Consequently, they believe the usefulness of float tests is very limited. However, there are many Youtube videos on float testing dice. One has nearly one million views (Fisher, 2015a). Some studies testing the fairness of dice with the chi-square test have applied the test in a way that has insufficient power to statistically determine the fairness or not. One such example rolled five 20-sided dice (D20s) 100 times by hand and another 100 times with a dice tower and drew conclusions regarding the fairness of the dice and the usefulness of the dice tower to cause unfair dice to be fair (Fisher 2015b). Campbell and Dolan (2019) again showed the weakness of these statistical arguments.

Campbell and Dolan (2020) also showed that there is some value to measuring the spread of diameters for dice since these show statistically significant positive correlations between the range of diameters for a given die and the χ^2 value obtained after 3000 rolls. The two papers by Campbell and Dolan also show the futility of testing dice fairness using only a few hundred rolls. They give several examples of dice they tested that rolled fair with several hundred rlls, but tested unfair after 3000 rolls. They also provide examples of dice tested that appear highly unfair after a hundred rolls, but ultimately tested fair after 3000 rolls. The running chi-square statistic was used in these papers to show the large statistical oscillations in χ^2 during the early stages of dice tests that can lead to erroneous conclusions concerning dice fairness.

The chi-square test has been used in several studies to test the fairness of dice. Typically, only one value of χ^2 is calculated. The running chi-square test calculates χ^2 for one roll, 2 rolls, 3 rolls, and so on and plots the curve. The resulting plot provides clear visual evidence of the lack of power the chi-square goodness of fit test for only a few hundred rolls of the die.

The chi-square statistic for dice testing is given by the following equation. The running chi-square statistic is simply the chi-square statistic calculated after every roll in a die test.

$$\chi^2 = \sum_{i=1}^{s} \frac{\left(O_i - E_i\right)^2}{E_i} \qquad (1)$$

$s =$ number of sides of the die

$O_i =$ observed occurences of the ith side of the die

$E_i =$ expected occurences of the ith side $= \dfrac{n}{s}$

$n =$ number of rolls of the die

χ^2 is asymptotically chi-square distributed as *n* approaches ∞. A precise distribution can be derived, but it is mathematically much more complex, and the chi-square distribution is good after 100 rolls.

Example:
A Mini Planet D20 was rolled 3000 times with the automated roller with the following results. A one came up 130 times in 3000 rolls, a two 149 times, and so on. All of the observed frequencies for each side of the die are given below with the calculation of the χ^2 value.

Figures 1 and 2 give examples of the running chi-square statistic for 3000 hand rolls of one die that tested fair, and one that tested unfair. Figure 2 shows that before 200 rolls, the unfair die would have tested fair. The red line on the chart is the 95 percent confidence line for a fair die. That is, only once in 20 tests would a fair die test unfair. Using only 100 rolls, someone might conclude that the die is fair. However, after 3 thousand rolls the χ^2 value is almost 80 which is characteristic of a very unfair die.

Also notice that of the O_i, only 6 have an odd number of observations (O_2, O_5, O_6, O_{10}, O_{11}, and O_{17}). For an even number of sides to the die, in this case 20, there must also be an even number of odd number observations. If not, the sum of all observations could not add up to 3000. This will be important in a proof to follow.

$O_1 = 130$

$O_2 = 149$

$O_3 = 138$

$O_4 = 132$

$O_5 = 151$

$O_6 = 145$

$O_7 = 156$

$O_8 = 146$

$O_9 = 152$

$O_{10} = 143$

$O_{11} = 149$

$O_{12} = 170$

$O_{13} = 150$

$O_{14} = 177$

$O_{15} = 164$

$O_{16} = 134$

$O_{17} = 147$

$O_{18} = 138$

$O_{19} = 165$

$O_{20} = 164$

$$\chi^2 = \frac{(130-150)^2}{150} + \frac{(149-150)^2}{150} + \frac{(138-150)^2}{150} + \ldots + \frac{(164-150)^2}{150} = 21.04$$

This is the final answer. Figure 1 shows the running chi square statistic as a function of the roll number. This curve begins at 19 (always s – 1) and ends at 21.04. This figure is typical of a die that rolls fair after 3000 rolls. The red line is the 95 percent confidence line, that is the value of the statistic that will only be exceeded 5 percent of the time for a fair die.

An unfair die has a linear trend, and this is exemplified in Figure 2. In the next section we show that for the null hypothesis that the die is fair, and unfair die running chi square statistic will have an added constant term and a liner trend like the one shown in Figure 2.

Figure 1. Typical running chi square statistic for a fair die.

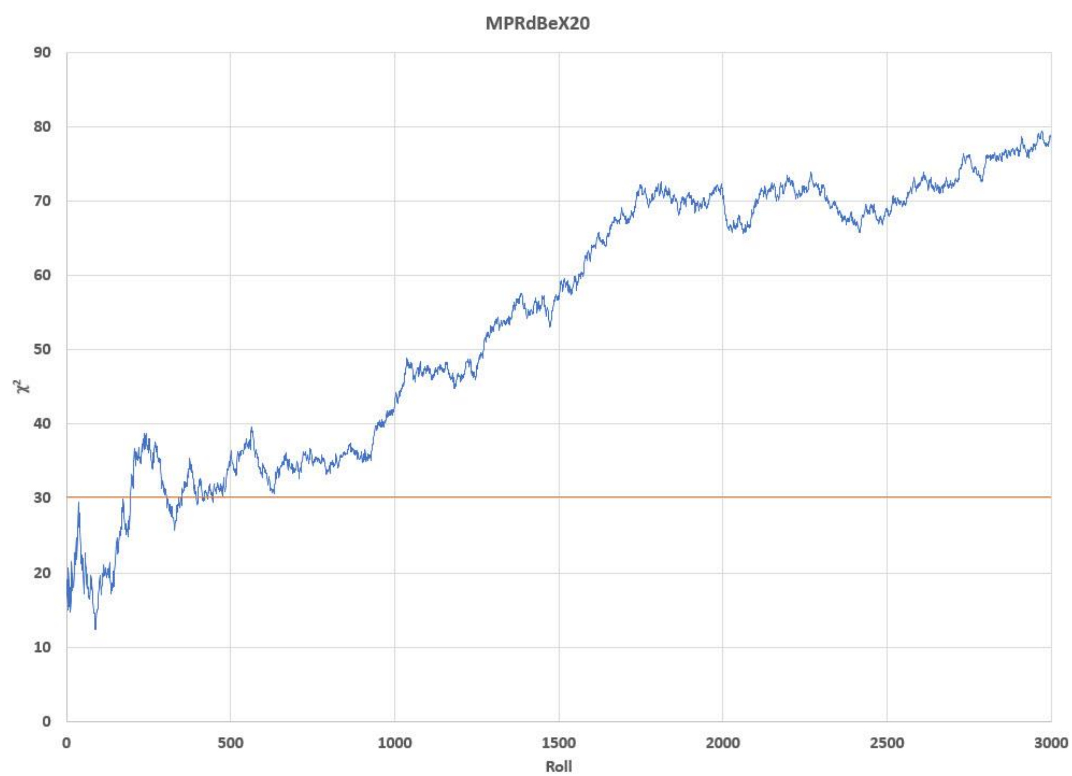

Figure 2. Typical running chi square statistic for an unfair die.

Mathematical Characteristics of the Running Chi-Square Statistic:

After 1 roll, the chi-square statistic will have a value of the number of sides *s* of the die minus 1. The running chi-square statistic for a D20 always begins at 19.

$$\chi^2 = \frac{\left(1-\frac{1}{s}\right)^2}{\frac{1}{s}} + (s-1)\frac{\left(0-\frac{1}{s}\right)^2}{\frac{1}{s}} = \frac{1-\frac{2}{s}+\frac{1}{s^2}}{\frac{1}{s}} + (s-1)\frac{1}{s} = s-2+\frac{1}{s}+1-\frac{1}{s} = s-1 \qquad (2)$$

For the 2nd roll, there are two possibilities. Say for example the first roll is a 1. Then either the 2nd roll is a 1 or one of the 19 other numbers for a D20. If the first number is not repeated, then the result is $s-2$. If the 1st number is repeated, the statistic is 2(*s* - 1).

$$\chi^2 = 2\frac{\left(1-\frac{2}{s}\right)^2}{\frac{2}{s}} + (s-2)\frac{2}{s} = \frac{(s-2)^2}{s} + 2 - \frac{4}{s} = \frac{s^2-4s+4+2s-4}{s} = s-2 \qquad (3)$$

$$\chi^2 = \frac{\left(2-\frac{2}{s}\right)^2}{\frac{2}{s}} + (s-1)\frac{2}{s} = \frac{\left(\frac{2s-2}{s}\right)^2}{\frac{2}{s}} + 2 - \frac{2}{s} = \frac{4s^2-8s+4+4s-4}{2s} = 2(s-1) \qquad (4)$$

It is also easy to show that the statistic will be an integer for 1, 2, *s*/2, s, and 2s rolls as long as *s* is an even number.

Based on proof by counter-example by one of us (Wimsatt) we can derive a general equation that allows us to deduce the largest roll that will always produce an integer value.

E is the expected value where *E* = *n/s* where *n* is the number of rolls and *s* is the number of sides of the die. Then let:

$O_1 = 1$
$O_2 = E - 1$
$O_3 = 2E$
O_4 through $O_s = E$. Then

$$\chi^2 = \frac{(1-E)^2}{E} + \frac{(-1)^2}{E} + \frac{(2E-E)^2}{E} + (s-3)\cdot 0 = \frac{1-2E+E^2+1+E^2}{E} = \frac{2}{E} - 2 + 2E \qquad (6)$$

From the result of Equation 6, since the last 2 terms will be integers if *E* is an integer then the largest integer value of *E* that will always result in an integer value of χ^2 is 2. So for a D20, the largest number of rolls that will always produce an integer is 40 which gives *E* = 2. This does not prove that χ^2 will be an integer for 40 rolls. It proves that for no number of rolls greater than 40 can the statistic always be an integer. For a D8, the largest number of rolls will be 16, for a D10, 20, and so on.

We can prove that for *E* = 2, the value of χ^2 will always be an integer.

$$\chi^2 = \frac{(O_1-2)^2}{2} + \frac{(O_2-2)^2}{2} + \ldots + \frac{(O_s-2)^2}{2} = \frac{O_1^2+O_2^2+\ldots+O_s^2}{2} - 2(O_1+O_2+\ldots+O_s) + s\cdot 2 \qquad (7)$$

If the sum of squares is always an even value, then χ^2 must be an integer. For all the O_i that are even, the squares will all be even as will their sum. The number of odd observations must be even for the sum of all the observations to add up to an even number. Therefore, the sum of squares of all the observations must be even. Then for a D4, 8 rolls will give an integer value of χ^2, 12 rolls for a D6, 20 for a D10, 24 for a D12, and 40 for a D20 will always give an integer. This information is useful for testing codes and spreadsheets evaluating χ^2.

It is also easy to show that if a roll result does not repeat an earlier roll, χ^2 will decrement by 1 for each roll. This is possible only for *s* rolls. So, for a D20, as long as a roll is not repeated, sequential values of χ^2 will be 19, 18, 17, etc. The value of χ^2 goes all the way to zero after 20 rolls if no number is repeated. After that, the number must repeat.

The maximum value of the χ^2 statistic is $n\,(s\text{-}1)$ where n is the number of rolls. This number is delivered by a perfectly loaded die that always rolls one number.

$$\chi^2 = \frac{\left(n-\frac{n}{s}\right)^2}{\frac{n}{s}} + (s-1)\frac{\left(0-\frac{n}{s}\right)^2}{\frac{n}{s}} = \frac{n^2-2\frac{n^2}{s}}{\frac{n}{s}} + s\frac{n}{s} = sn-2n+n = n(s-1) \qquad (8)$$

We rolled a loaded D6 3000 times that rolled 1 about 80 percent of the time. The resulting χ^2 value was 8744 out of a possible 15000. Figure 3 shows the running chi-square statistic. The linear trend is seen clearly because normal random oscillations are small relative to the climbing statistic. The least-squares fit line for this has a slope of 2.899 while the theoretical slope is 2.915.

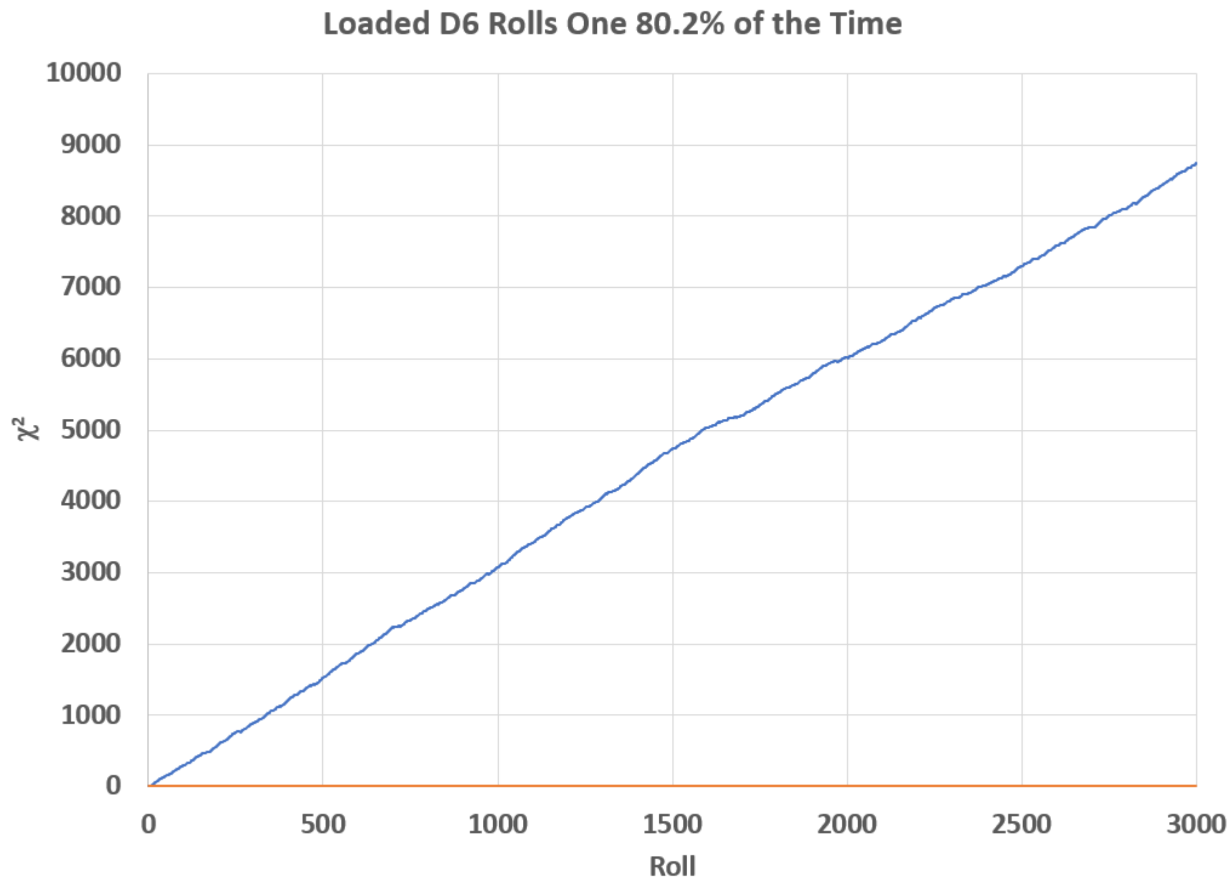

Figure 3. Running chi-square statistic for a loaded die.

For an unfair die, if the null hypothesis is that the die is fair, then the χ^2 statistic will have an added constant term and a linear trend. The more unfair the die the steeper the trend.

$$\chi^2 = \sum_{i=1}^{s} \frac{[(p + \Delta_i)n + \delta n_i - pn]^2}{pn} = \sum_{i=1}^{s} \frac{\Delta_i^2 n^2 + 2\Delta_i \delta n_i n + \delta n_i^2}{pn} = \sum_{i=1}^{s} \frac{\Delta_i^2}{p} n + 2\sum_{i=1}^{s} \frac{\Delta_i \delta n_i}{p} + \sum_{i=1}^{s} \frac{\delta n_i^2}{pn}$$

$$= s\left(\sum_{i=1}^{s} \Delta_i^2\right) n + 2s \sum_{i=1}^{s} \Delta_i \delta n_i + \chi_{fair}^2 \qquad (9)$$

$p = \frac{1}{s} =$ the probability for each side of the die based on the null hypothesis
$\Delta_i =$ the deviation of the actual probability of side i from the expected fair probability p.
$\delta n_i =$ the random deviation of the observed value of side i from the real probability which is $p + \Delta_i$ multiplied times the number of rolls n
$n =$ the total number of rolls
$\chi_{fair}^2 =$ the value of χ^2 if the die is fair, i.e. $\Delta_i = 0$.

Notice that the first term in the RHS of Eqn. 9 is a linear function of n and the slope is the number of sides of the die multiplied times the sum of squares of the probability deviations from fair. The second term is a constant term, that is, not a function of n. The final term is the value of the statistic if the die were fair.

There are some other interesting characteristics of the χ^2 statistic. For the chi-square distribution, there is a finite probability of occurrence of the χ^2 statistic for any positive, finite value of the statistic. However, the maximum value has been shown to be *n (s-1)*. Also, for 1, 2, *s*/2, *s*, and 2*s* rolls the value can only be an integer. Certainly, the chi-square distribution has a finite, non-zero value for any interval of positive real numbers. The value of χ^2 only asymptotically approaches a chi-square distribution. Also, from the definition of the statistic, it can only assume rational values for any finite number of rolls. There are infinitely more irrational numbers than rational ones.

Since a 20-sided die is the most difficult to manufacture precisely, this is the focus here. We chose 3000 rolls partly because of our feeling that this is the typical life of a die and partly because 3000 rolls give the test enough power to detect the unfairness of most dice.

Dice Testing Results

Gamers want to know who makes the best dice. A Google search quickly confirms the interest in the fairness of dice which begs the question, “who makes the best dice?” Unfortunately, we cannot answer that question without much more dice testing. If we say that one manufacturer is better than another, we would like to be 95 percent sure that our conclusion is correct. We have tested no more than 5 dice from any manufacturer. We need to test many more dice to achieve that confidence level. Suppose that Manufacturer1 produces dice so that 1 out of 5 dice on average will be fair. Manufacturer 2 can produce dice so on average 2 of 5 are fair. For random samples of dice drawn from these two manufacturers, the number of fair dice in these samples follows the binomial distribution.

Using this distribution we find that if we tested 5 dice from each manufacturer, on average we would conclude that manufacturer 1 has the fairest dice 14 percent of the time. If we test 10 dice each, the probability of a wrong conclusion would occur 10.5 percent of the time. If we test 20 dice, the probability of an error is still 5.6 percent. We need to test more than 20 dice from each manufacturer to achieve the level of confidence we desire.

We had 2 methods of testing dice. Method 1 was an automated rolling (High Roller) device (Figure 4) developed by senior engineering students at WKU. Method 2 was hand rolling using a dice tower (Figure 5). At the center of the device is a high roller cup connected to a servo motor. A cap with a circular hole in it allows access for photography. Under the inside of the cap is an LED strip providing light for photography. The servo motor is controlled by an Arduino controller which causes the cup to rotate approximately 90 degrees and back to an upright position. Initially, there was a single roll and a photograph. The rotation and photography were repeated more than 3000 times with the hope that 3000 useful photographs would be obtained. At times, the die would roll and rest on the edge of the cup wall with the point of a vertex facing the camera. In these cases, it was impossible to determine the roll. The single roll and photography mode did not appear to be giving independent rolls, so we changed the mode to 2 rolls, then a photograph and the results were much better.

In one case with a heavy metal die we designate as DHBkGdM20 (D20 Die Hard die with a black body, gold numbers, made of metal), we observed 9 consecutive 5s. That kind of statistical run we would only expect to occur on average once in 20^8 = 25.6 billion rolls. Ten times in 3000 rolls, we observed runs of 6 or more. To check the independence from one roll to the next, we correlated 2999 rolls with 2999 rolls offset by one roll fully expecting a high correlation coefficient. What we got was 0.0339. The correlation was clearly not sensitive to dependence from one roll to the next. We needed a better indicator.

We settled on looking at the probability distribution of rolls following the current one. If the rolls were actually independent, we would expect the next roll to be uniformly distributed for each die side. That is, the probability of rolling a 1 next = the probability of a 2 = … = the probability of a 20. This was something we could check with the chi-square test. The result for this particular die is shown in Figure 6. The figure compares next roll probability distribution for the High Roller (HR) device, to the same die rolled in a dice tower, to a simulation of fair rolls, and to the 95 percent probability value of χ^2. For every single side, the HR χ^2 value exceeded the 95 percent confidence level. Rolling in the dice tower appears to be a reasonably fair way to roll since only 3 sides out of the 20 sides of the die did it exceed the 95 percent confidence level.

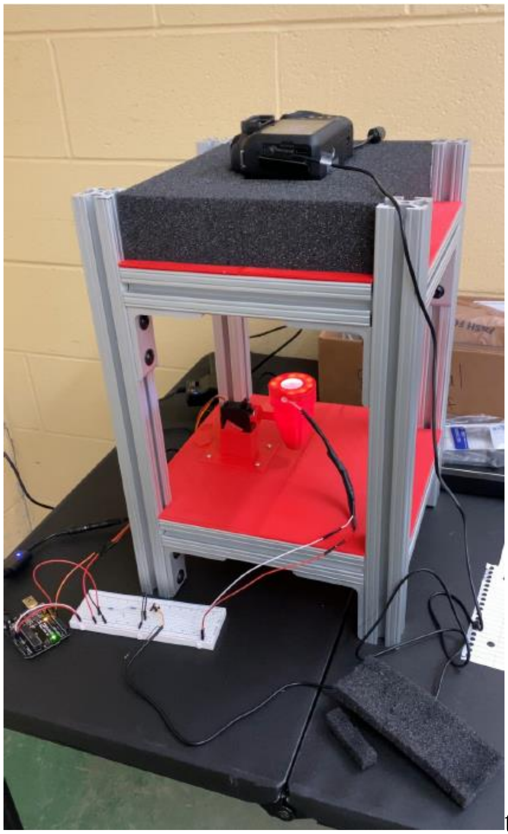
the

Figure 4. The High Roller device.

Figure 5. Dice tower.

One thing illustrated very clearly by this study is that the method of rolling is very important. A few dice were rolled by the High Roller device and by hand with a dice tower. Results were very different. In one case, the χ^2 value obtained in the High Roller device was low compared to the dice tower result (Figure 8). In two other cases the opposite occurred (Figure 9 and 10). For the dice used to produce Figures 9 and 10, the observed values give the best estimate of the die side probabilities. We assumed that these were correct and simulated 5 additional sets of rolls based on those probabilities. The results are shown in Figures 11 and 12. There is a clear separation in the dice tower and High Roller results, but only after thousands of rolls. Based on this limited data, our preliminary conclusion is that the method of rolling is important.

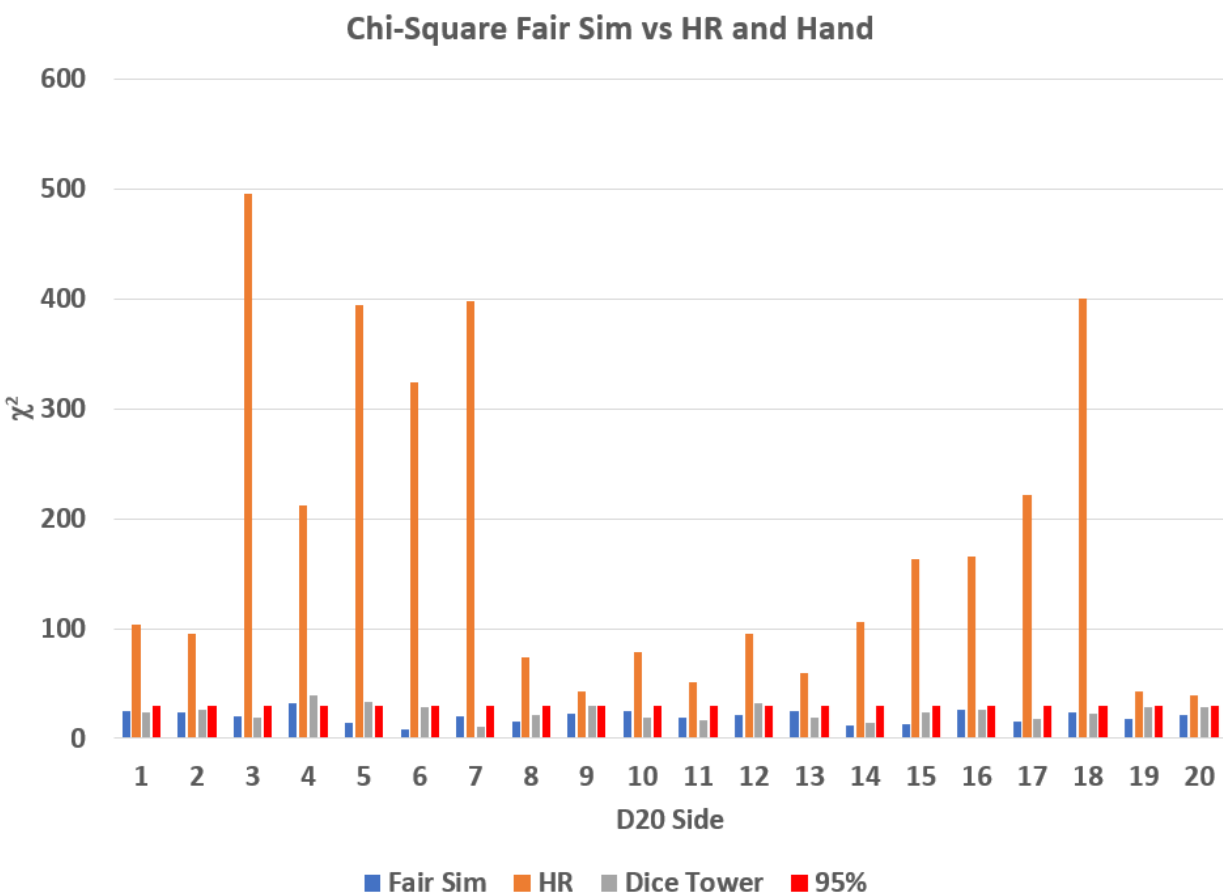

Figure 6. Next roll probability test for DHBkGdM20.

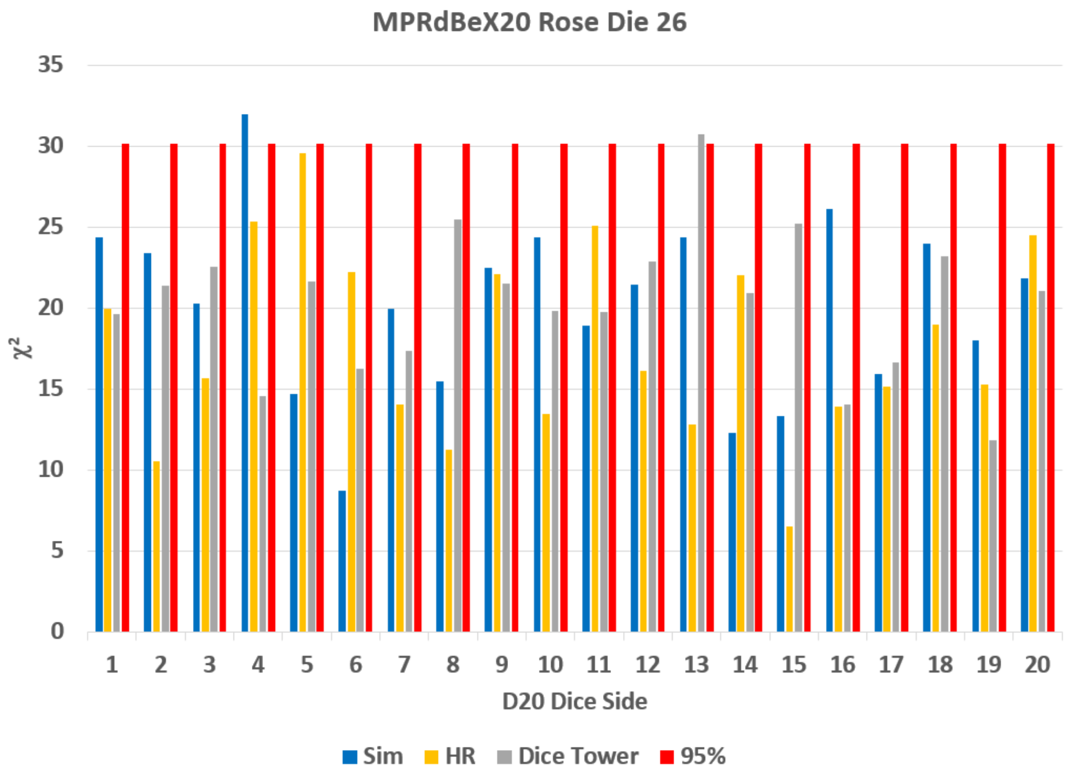

Figure 7. Next roll probability test for MPRdBeX20.

Figure 8. High Roller with χ^2 less than for the dice tower.

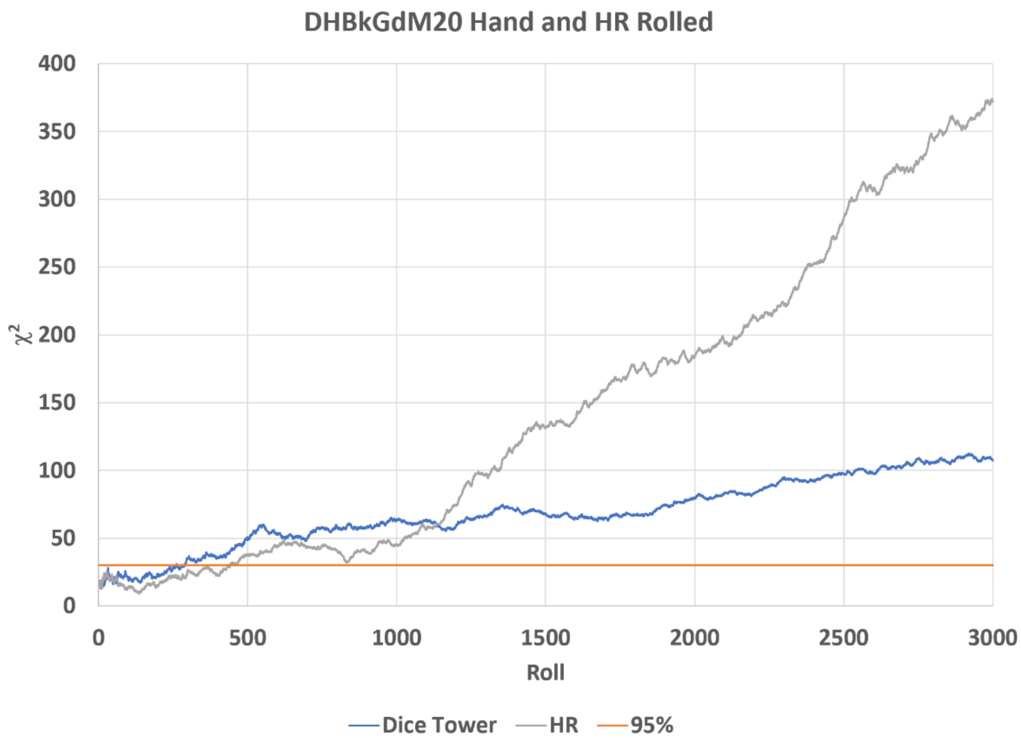

Figure 9. High Roller with χ^2 greater than for the dice tower.

Figure 10. High Roller with χ^2 greater than for the dice tower.

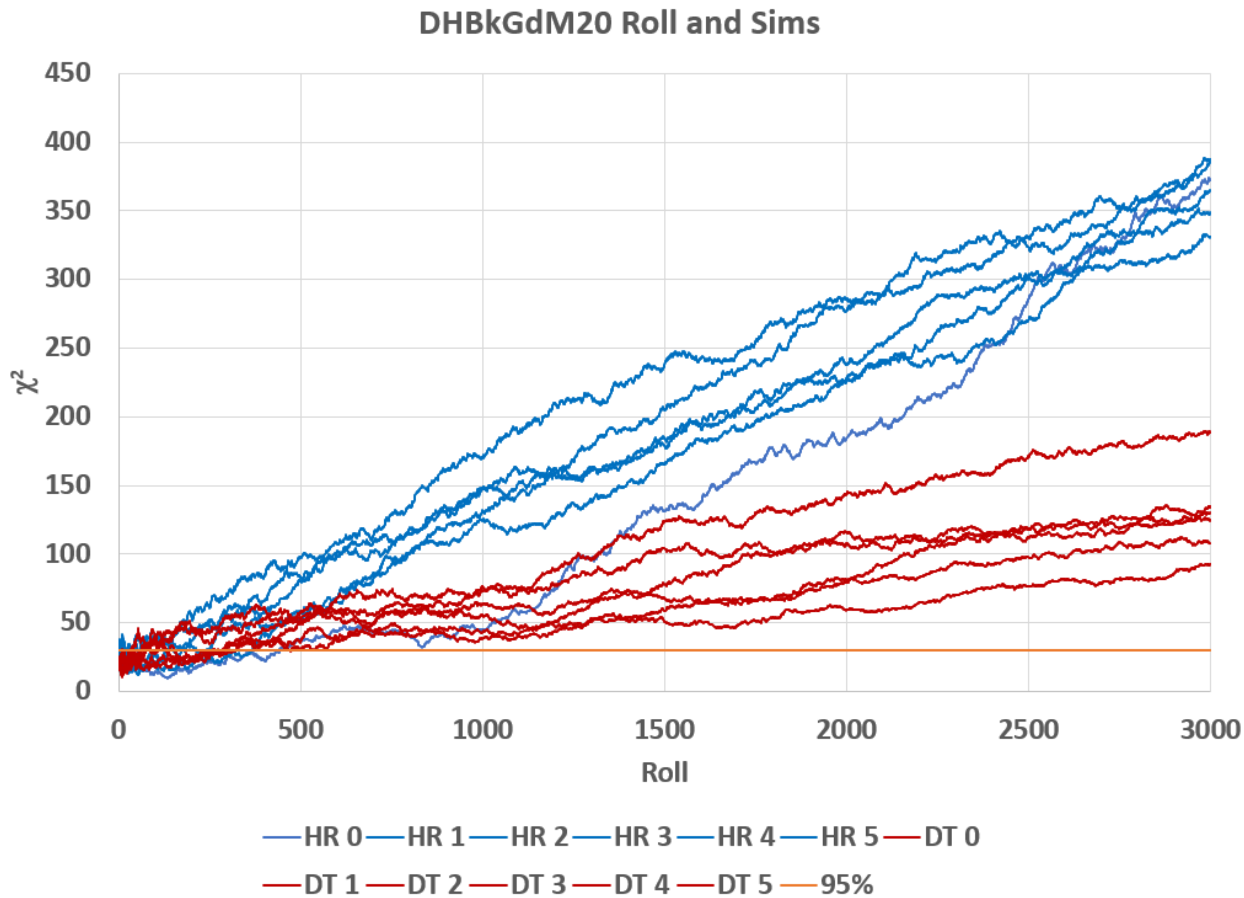

Figure 10. Dice roll and simulations for the Die Hard black and gold D20 metal die.

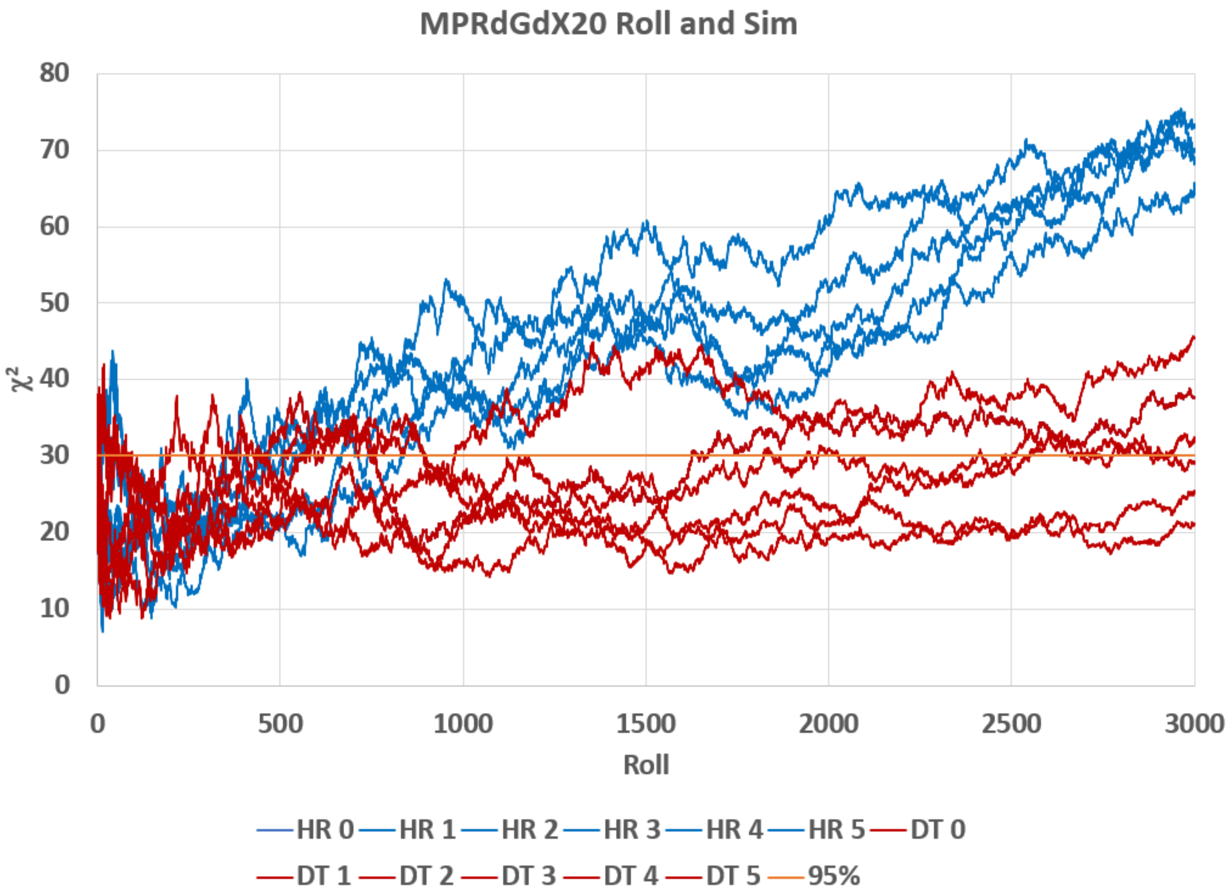

Figure 11. Dice roll and simulations for the Mini Planet red and gold plastic D20.

Summary and Conclusions

All dice are unfair; some are more unfair than others. The running chi-square statistic is a useful method of testing the unfairness of dice but must be applied over many rolls to obtain enough power in the statistical test. Plots of the running chi-square statistic clearly show that a few hundred rolls are not sufficient to determine the fairness of dice. The running chi-square test has a linear trend that is steeper the more unfair the die.

Dice roll outcomes follow the Dirichlet distribution. This distribution asymptotically approaches the chi-square distribution. The chi-square distribution is a continuous distribution though the χ^2 statistic only produces rational numbers. The value of χ^2 after one roll is $s - 1$ where s is the number of sides of the die. After 2 rolls, it is either s - 2 or $2(s - 1)$. χ^2 is an integer after 1, 2, $s/2$, s, and after $2s$ rolls. The maximum value of χ^2 is $n\ (s - 1)$. This information in this paragraph is useful for testing code or spreadsheets.

The outcome of dice rolls is strongly dependent on the method of rolling. Dice can test unfair by one method and fair by another. The best rolling method for testing dice is the one closest to the way the dice will be rolled in practice.

Acknowledgments

We are indebted to the following WKU engineering students who developed and operated the High Roller Device. Without their contribution, this study would not have been possible.

Jordon Charlton
Dylan Harris
Faruk Hasanovic
Landen Kent
Chet Massaro
Jack Poole
Briana Robertson
Travis White

References

Campbell, C. Warren and Dolan, William P. "Dice Questions Answered," (2020), https://digitalcommons.wku.edu/seas_faculty_pubs/2/, 32 pp.

Campbell, C. Warren and Dolan, William P. "Dice Mythbusters" (2019). *Standard Research Conference Select Presentations. Paper 48. https://digitalcommons.wku.edu/sel_pres/48,* 49pp.

Fisher, Daniel (2015a). "How to Check the Balance of Your D20," https://www.youtube.com/watch?v=VI3N4Qg-JZM.

Fisher, Daniel (2015b). "Does a Dice Tower Make Your Dice Roll More Random?", https://www.youtube.com/watch?v=5HXYOgjp3go.