An Exposition on the Expected Maximum of Two Dice and the Nature of Formulas for Finite Sets

Reese Hatfield, Zachariah Pence

July 7, 2022

1 Introduction

In many games or other simulations, it is a common task to roll two dice, with the same number of sides, and choose the largest of the two numbers, (or either one if they are equal). Six sided dice are the most common, and via experimentation, or exhaustion of all possibilities, it is possible to determine the expected outcome. The expected outcome being the average of the maximum of the two dice, for which this can be expressed mathematically.

2 Setup & Simulation

Let X and X' be discrete random variables such that $X \sim U(1, N)$ and $X' \sim U(1, N)$ respectively. The expected maximum of two dice with sides can be calculated using Equation (1).

$$\bar{x} = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} \max\{X_i, X_i'\}}{N} \tag{1}$$

To get an idea of what this might look like, a simulation was ran where for a given value N two random integers from 1 to N were chosen and the maximum of those value were taken; repeating this process 10 000 times will yield an approximation for $\bar{x}$ in Equation (1). The results of the simulation are represented as the data points in Figure (1).

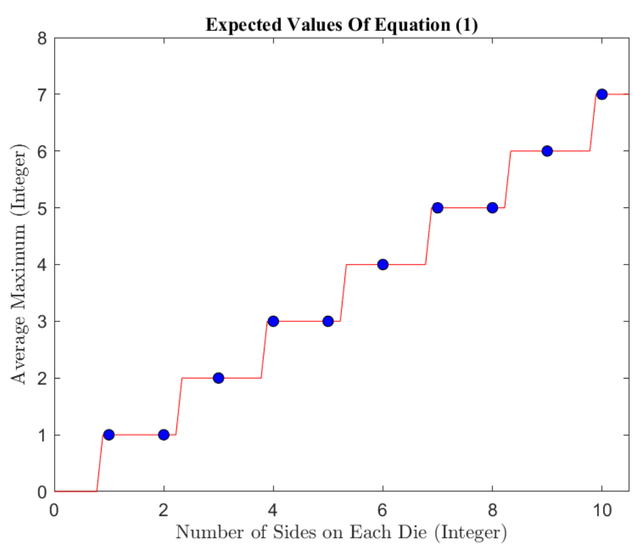

Figure 1: Results of Simulation with Best Fit Function

For discrete distributions, in essence the positive integers, Equation (2) can be used as a best-fit function (as shown in Figure (1)). It can be assumed, although in this paper it is not proven, that this formula continues for all $N \in \mathbb{Z}^+$.

$$f(N) = \left\lfloor \frac{2}{3}(N - \frac{2}{3}) \right\rfloor \tag{2}$$

3 Further Discussion

While Equation (2) works to determine the expected maximum of two dice rolls, there are likely many other formulas that could determine the correct output as well. For example, while not particularly practical, it is possible to extend N to the set of real numbers $\mathbb{R}$. There are many common interpolation/extrapolation formulas that this could be done with, but included in Figure (2) is the Lagrange Interpolating Polynomial for the data set of the first 10 outputs of the function. When just considering the data set for natural

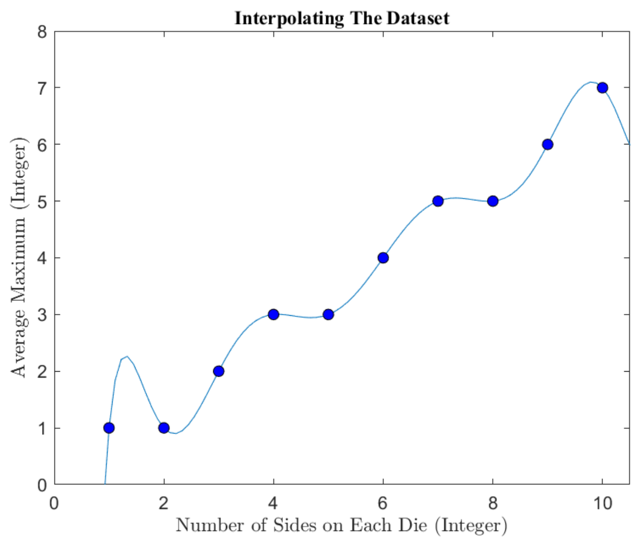

Figure 2: Lagrange Interpolating The Head of The Data set

numbers given, both of these formulas are "correct" in the sense that they return the correct output. So which of them is actually correct?

Well, the first function outputs the correct for all values of $N \in \mathbb{Z}$. But in practice, one may not be able to guarantee that the simpler formula holds up for higher values of N. In this specific case, where the problem is "Find the formula for the expected value of the two dice rolls with N sides", the proof that it works is quite trivial and is left as an exercise for the reader. However, for more complex examples, the proof may be difficult, or not obtainable at all.

This example is a toy problem, meant to show that in many scenarios, especially in that which involve computer science, it may be impossible to find an end-all-be-all solution to your problem. The more outputs one can test, the stronger belief one can have that their solution is the correct one, however, testing cannot prove a hypothesis in more rigorous, infinite mathematics.

It is similar to the idea that π is a normal number. That is to say, that in the decimal expansion of π, its infinite sequence of digits is distributed uniformly in the sense that they have an equal, natural density. We believe this to be true, having tested it to over twenty-two trillion digits, but a proof avoids the larger mathematical community, regardless of how many digits we test.

In conclusion, formulas for finite data sets do not always have a concrete "correct" or best solution. This may occur when there is an infinite, yet unproven consistency in the set given. If this is the case, there could be many possible formulas for the given output, despite intuition saying there is a single best answer. Intuition must always be back with rigor.

References

[1] Callum Ilkiw. *Slaying the dragon. Chalkdust.* 2022. URL: https://chalkdustmagazine.com/features/slaying-the-dragon/ (visited on 07/05/2022).

[2] Peter Trueb. *Digit Statistics of the First 22.4 Trillion Decimal Digits of Pi.* 2016. DOI: 10.48550/ARXIV.1612.00489. URL: https://arxiv.org/abs/1612.00489.

[3] Edward Warring. *Problems concerning Interpolations.* 1779. DOI: 10.1098/rstl.1779.0008. URL: https://archive.org/details/paper-doi-10_1098_rstl_1779_0008/page/n7/mode/2up.