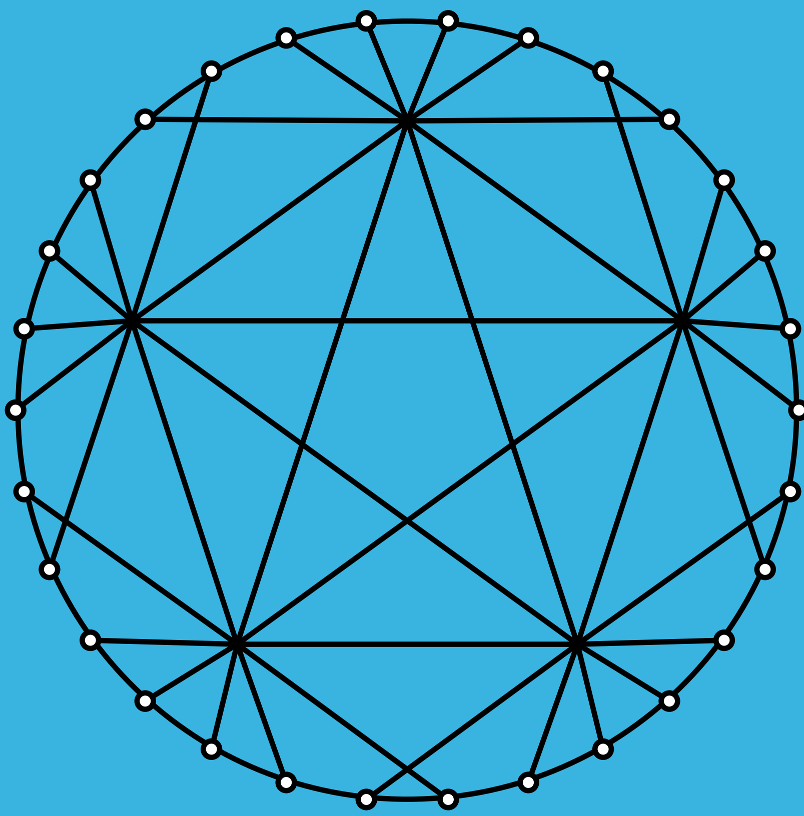
BULLETIN of The INSTITUTE of COMBINATORICS and its APPLICATIONS

Volume 92
June 2021



Boca Raton, FL, U.S.A.

ISSN: 2689-0674 (Online)
ISSN: 1183-1278 (Print)

BULLETIN OF THE ICA
Volume **92** (2021), 109–113

A die problem

Douglas R. Stinson*

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada
dstinson@uwaterloo.ca

Abstract

In this note, I discuss the following problem: What is the expected number of throws of a die until a 6 appears, given that no odd throws occur?

1 The problem and two solutions

Peter Winkler gave an entertaining talk at the 52nd Southeastern International Conference on Combinatorics, Graph Theory & Computing in March 2021. The title of his talk was "Probability in Your Head". He discussed how to find clever solutions to a variety of puzzling probability problems, including the following die problem that is attributed to Elchanan Mossel: what is the expected number of throws of a die until a 6 appears, given that no odd throws occur? Peter Winkler provides an elegant solution to this problem in his recent book [1].

To solve the problem, we want to compute a conditional expectation, where we only consider a certain subset of sequences of numbers obtained by throwing a fair die. The intuitively obvious answer to this question is 3, since it seems plausible that the problem is equivalent to throwing a 3-sided die (having faces $2, 4$ and 6) until a 6 is thrown. However, this answer is not correct; the correct answer turns out to be $3/2$.

*The author's research is supported by NSERC discovery grant RGPIN-03882.

Received: 25 March 2021
Accepted: 12 April 2021

This die problem has received a considerable amount of attention in the last few years, and it is addressed in various blogs, including the following:

- https://gilkalai.wordpress.com/ (Sept. 7, 2017)
- https://mindyourdecisions.com (Oct. 8, 2017)
- https://www.untrammeledmind.com (Dec. 7, 2017).

We encourage the reader to look at the extensive discussion about this problem in these blogs.

My first objective in this note is to write down a straightforward mathematical derivation of the correct answer. (This argument can be found, in various guises, in some of the aforementioned blogs.) This approach is not as elegant or ingenious as some of the other approaches. However, it is straightforward and relatively simple and it only requires basic notions of expectation and conditional expectation. It does require computing the sums of two infinite series, however.

My second objective is to give a more elegant proof that does not require summing infinite series.

In both approaches, I am using standard notation and formulas for random variables and their expectation. I think this is helpful in avoiding possible mathematical errors.

1.1 First Solution

Consider the set $\mathcal{S}$ of all infinite sequences of throws of a fair die. For $i \geq 1$, let $\mathcal{G}_i$ consist of all the sequences in which the first $i-1$ throws of the die are 2 or 4, and the ith throw is 6. That is, there are no odd throws of the die (among the first i throws), and the first throw of a 6 is on the ith throw. Let $\mathcal{G} = \cup_{i=1}^{\infty} \mathcal{G}_i$. The sequences in $\mathcal{G}$ are defined to be *good* and the sequences in $\mathcal{B} = \mathcal{S} \setminus \mathcal{G}$ are *bad*.

Suppose we consider a random sequence in $s \in \mathcal{S}$. Define a random variable $\mathbf{X}$ that takes on non-negative integral values as follows:

$$\mathbf{X} = \begin{cases} i & \text{if } s \in \mathcal{G}_i \\ 0 & \text{if } s \in \mathcal{S} \setminus \mathcal{G}. \end{cases}$$

Thus $\mathbf{X}$ records the position of the first 6 whenever a sequence is good.

For all $i \geq 1$, it is clear that

$$\Pr[\mathbf{X} = i] = \left(\frac{1}{3}\right)^{i-1} \left(\frac{1}{6}\right).$$

The expected value (among all possible sequences in $\mathcal{S}$, good or bad) of the first throw of a 6 in a good sequence is

$$\begin{aligned}
\mathrm{E}[\mathbf{X}] &= \sum_{i=1}^{\infty} i \Pr[\mathbf{X} = i] \\
&= \sum_{i=1}^{\infty} i \left(\frac{1}{3}\right)^{i-1} \left(\frac{1}{6}\right) \\
&= \left(\frac{1}{6}\right) \sum_{i=1}^{\infty} i \left(\frac{1}{3}\right)^{i-1} \\
&= \left(\frac{1}{6}\right) \left(\frac{1}{1-\frac{1}{3}}\right)^{2} \\
&= \frac{3}{8}.
\end{aligned}$$

Note that the above sum is an arithmetic-geometric series.

However, we want to find the conditional expectation, conditioned on the sequence being good. The probability that a sequence is good is

$$\begin{aligned}
\Pr[\mathbf{X} > 0] &= \sum_{i=1}^{\infty} \Pr[\mathbf{X} = i] \\
&= \sum_{i=1}^{\infty} \left(\frac{1}{3}\right)^{i-1} \left(\frac{1}{6}\right) \\
&= \left(\frac{1}{6}\right) \left(\frac{1}{1-\frac{1}{3}}\right) \\
&= \frac{1}{4}.
\end{aligned}$$

Note that this sum is just a geometric series.

The conditional expectation is therefore

$$\frac{\mathrm{E}[\mathbf{X}]}{\Pr[\mathbf{X} > 0]} = \frac{\frac{3}{8}}{\frac{1}{4}} = \frac{3}{2}.$$

It might be of interest to explain how the above analysis would actually relate to a sequence of tosses of a die. Suppose we consider n random sequences. We can terminate a sequence as soon as a $1, 3, 5$ or 6 is thrown. In the first three cases, the sequence is bad; while a throw of 6 yields a good sequence. We proved that the probability of a good sequence is $1/4$. Thus, on average, $3n/4$ of the n sequences will be bad and $n/4$ sequences will be good.

We also showed that $\mathrm{E}[\mathbf{X}] = 3/8$. This is saying that the total length of all the good sequences is, on average, $3n/8$. Finally, the desired conditional expectation is obtained by dividing the total length of the good sequences by the number of good sequences, obtaining $3/2$.

1.2 Second Solution

The above derivation was fairly straightforward, but it is also worthwhile to consider shorter, more elegant ways to derive the same result.

We showed above that the probability that a random sequence is good is $1/4$ by summing a geometric series. However, there is a simpler way to prove this. We observed that we can terminate a sequence of throws as soon as one of $1, 3, 5$ or 6 is thrown. Since $1, 3$ and 5 correspond to bad sequences, it is obvious that a bad sequence is three times more likely than a good sequence. Thus, the probability of a bad sequence is $3/4$ and the probability of a good sequence is $1/4$.

We can also observe that the probability of obtaining a 6 on the first throw is $1/6$; these are the sequences in $\mathcal{G}_1$. Let s denote a random sequence; then $\Pr[s \in \mathcal{G}_1] = \frac{1}{6}$. We also know that $\Pr[s \in \mathcal{B}] = \frac{3}{4}$. It therefore follows that

$$\Pr[s \in \mathcal{G} \setminus \mathcal{G}_1] = 1 - \frac{1}{6} - \frac{3}{4} = \frac{1}{12}.$$

We can now compute the conditional probabilities of s being in $\mathcal{G}_1$ or in $\mathcal{G} \setminus \mathcal{G}_1$, given that s is a good sequence:

$$\begin{aligned} \Pr[s \in \mathcal{G}_1 \mid s \in \mathcal{G}] &= \frac{2}{3} \\ \Pr[s \in \mathcal{G} \setminus \mathcal{G}_1 \mid s \in \mathcal{G}] &= \frac{1}{3}. \end{aligned}$$

Finally, we compute $E = \mathrm{E}[\mathbf{X} \mid s \in \mathcal{G}]$ by making use of this conditional probability distribution. Let A denote the event that $s \in \mathcal{G}_1$ and let B

denote the event that $s \in \mathcal{G} \setminus \mathcal{G}_1$. Clearly, $\mathrm{E}[\mathbf{X} \mid A] = 1$. It is also not hard to see that $\mathrm{E}[\mathbf{X} \mid B] = 1 + E$, because a sequence in $\mathcal{G} \setminus \mathcal{G}_1$ can be decomposed as a 2 or 4, followed by a random sequence in $\mathcal{G}$. Thus we have

$$\begin{aligned} E &= \frac{2}{3} \times 1 + \frac{1}{3} \times (1 + E) \\ &= 1 + \frac{E}{3}, \end{aligned}$$

from which it follows that $E = 3/2$.

2 Discussion

Probability puzzles can be tricky to solve, and the answers can often seem counterintuitive. In the case of problems involving expected values, I think it is helpful to define appropriate random variables to clarify exactly what it is that we want to compute. Of course this does not preclude using clever arguments to avoid the use of mathematical formulas.

Adrián Pastine suggested that various generalizations of the problem could be considered:

- What happens if, instead of throwing a 6-sided die, we throw an n-sided one?
- What happens if, instead of forbidding odd throws, we forbid throws from a set of values of size k? (And how does the result compare to throwing a k-sided die where no throws are forbidden?)
- What happens if, instead of stopping when a 6 appears, we stop when any value from a set of size r appears?

These could be fun puzzles for an interested reader to contemplate.

Acknowledgement

I would like to thank Adrián Pastine for suggesting the variants of the problem that are listed in Section 2.

References

[1] Peter Winkler. *Mathematical Puzzles.* CRC Press, 2021.