Elchanan Mossel's dice problem

Jimmy Jin

May 29, 2018

I found the following truly wonderful problem on Gil Kalai's blog. It's a fairly simple dice problem given as an example by Elchanan Mossel during a lecture at UPenn (supposedly).

> You throw a fair six-sided die until you get 6. What is the expected number of throws (including the throw giving 6) conditioned on the event that all throws gave even numbers?

Give it a shot first, and then I will explain why I love it so much, especially for teaching.

How to get the wrong answer

Whenever a statistician sees something like "number of trials until X happens," they think of the geometric distribution. That's the distribution which models the number of trials (with binary outcomes, so success or failure) it takes until you get one success (including the last, successful roll)[1]. In this case, the trials are the successive rolls of the die. A "success" is the event of rolling a 6. The probability of rolling a 6 on a fair die is 1/6.

Without further conditioning, the number of times we would need to roll until a 6 would be a geometric random variable with parameter $p = 1/6$. Therefore the expected value would be $1/p = 1/(1/6) = 6$ times. So how does this change when we condition on the event that all rolls before the 6 were even?

Well, conditioning on that event is *sorta* like restricting the possible die outcomes to only 2, 4, and 6. In a three-sided die with outcomes 2, 4, and 6, the probability of rolling a 6 is now 1/3. So now the number of rolls until 6 is a geometric(1/3) random variable, and the expected number of rolls is $1/p = 1/(1/3) = 3$.

There are lots of other wrong solutions which more or less boil down to this same reasoning.

How to get the right answer

Since we stop when we hit a 6, here's another way to phrase the question:

> What is the expected number of times you can roll only 2's or 4's until you roll any other number, **conditional on that number being 6**?

Now here's the trick. It turns out that this quantity is the same as the unconditional quantity:

> What is the expected number of times you can roll only 2's or 4's until you roll any other number?

Explanation: Let N be the number of times you roll only 2's or 4's until you roll any other number. Let X be the die number on that last roll (i.e. the roll which came up something other than a 2 or 4). Obviously, X and N are independent, which means that

$$\mathbb{E}(N) = \mathbb{E}(N|X)$$

Now N is a plain vanilla geometric random variable if we define a "success" as rolling any number except a 2 or a 4. The probability of success is therefore $p = 2/3$, so

$$\mathbb{E}(N) = 1/(2/3) = 3/2$$

[1] for brevity, I will always count the final roll when counting rolls "until"

The paradox of the 3-sided die

So when we pretend the situation boils down to rolling a 3-sided die, we got an answer of 3. When we condition on a 6-sided die roll though, we got 1.5.

In other words, it takes **twice as long** to get a 6 when rolling a 3-sided die with faces 2, 4, and 6 than it does rolling a 6-sided die and conditioning on all even numbers before the 6.

When you think about it, this seems completely impossible. How can it be that, when compared to rolling a 3-sided die, conditioning on the **same set of numbers** on a die with **more faces** has now made it easier to roll consecutive 2's and 4's before rolling a 6?

Let me put a different question to you:

> You throw a fair **hundred**-sided die until you get 6. What is the expected number of throws (including the throw giving 6) conditioned on the event that all throws **came up 2, 4, or 6**?

Repeating the analysis we did in the "right answer" section, we see that the number of trials needed is a geometric(98/100) random variable, so its expectation is $\mathbb{E}(N) = 1/(98/100) = 100/98$.

So surprisingly, when we increase the number of faces we see that it takes an even **shorter** amount of time to roll a 6 relative to a 3-sided die.

The subtraction fallacy

EDIT: *this explanation has been updated since my original wording was a bit misleading (see Prof. Meyer's comment below).*

The source of this paradox is something you might call the *subtraction fallacy*—the belief that conditioning is equivalent to simply taking away some part of the sample space and then being done.

Well actually, conditioning **does** amount to taking away part of the sample space. But the key is you either have to take away the events on the **right** sample space, or at least be clever about updating probabilities. Let me explain.

In my view, the fallacy arises in the urge to use a simplified sample space that doesn't reflect the problem. In the context of this die problem, the wrong way to think about this is:

> OK, any outcome with odd throws doesn't count. Therefore, I can ignore the odd rolls. But wait, then I'm only left with outcomes of 2, 4, or 6, so let me just consider a 3-sided die...

What is happening here is that we are (wrongly) couching the problem in terms of a single die roll, with sample space $\{1, 2, 3, 4, 5, 6\}$. The basic urge is to ignore the odd outcomes and re-examine the problem as repeated rolls of a die with only even faces $\{2, 4, 6\}$. This gets you into trouble because then you need to be careful how to assign probabilities to these events, which are **not** simply 1/3, 1/3, and 1/3.

A more natural sample space for this problem is the space of *sequences* of die rolls. Suppose I roll a 1, then a 5, then a 6. I might represent this sequence of rolls as a single "event" in the sequential space as:

$$(1, 5, 6)$$

Some other examples of events on this space might be:

$$(1, 1, 4, 2, 5, 6), (2, 2, 4, 2, 6), (2, 4, 3, 6), \ldots$$

Conditioning on the event of only rolling 2's and 4's before a 6 would result in "deleting" the events $(1, 1, 4, 2, 5, 6)$ and $(2, 4, 3, 6)$, but keeping the event $(2, 2, 4, 2, 6)$. Then calculating the expected number of rolls on this reduced space of die roll sequences results in the right answer.

The advantage of this approach is that the probabilities assigned to these sequences can be obtained from the original probabilities on the space of sequences by a simple scaling to account for the portion of the sample space taken away. No other adjusting required.

I thank Professor Albert Meyer for spurring this clarification. His original explanation was somewhat different, in that he did reduce the sample space in the context of a single die, but updated the probabilities to reflect the conditioning:

> You dispute the "subtraction fallacy" that conditioning on all evens is the same as restricting to a die that only throws evens, but that is exactly what conditioning does: the restriction to the even-sided die is correct; the blunder is in assuming that the even-sided die is fair. Rather it has probabilities 1/6, 1/6, and 2/3 of respectively throwing two, four, and six...

I prefer the view on the space of sequences of die rolls, but both lenses get you to the same conclusion.

The right perspective in English

Here's the "right" subtraction view in plain English:

> If I know that I've rolled a run of only even outcomes on a 6-sided die, how does that change the likelihood that the run was 1 roll long, or 2 rolls long, etc...

The key observation is that if you require all previous throws to be even, then it makes long runs very unlikely because of the probability of rolling odd numbers in between.

On a 3-sided die (with faces 2, 4, and 6), you have a 1/3 chance of rolling a number that is not a 2 or 4. On a 6-sided die, you have a 2/3 chance of rolling a number that is not a 2 or 4. The chance of having long runs of 2's or 4's is therefore much less with the 6-sided die. Therefore, the conditional probability of short runs becomes greater and the expected number of rolls until hitting a 6 goes **down**.

The main takeaway is that we cannot just ignore the odd outcomes in the sample space here—the **relative** likelihood of rolling say 2 versus 100 even numbers before rolling a 6 involves factoring in the possibility of odd rolls. The other half of the sample space still matters.