Question: What is the average number of rolls needed of a fair six-sided die to achieve every face at least twice?

Answer: 1,172,906,043 / 48,600,000 = apx. 24.13387

Solution 1

Solution 1 uses a Markov Chain. This method does not require any advanced math, but is rather tedious.

Between the number of faces that have been rolled once and the number that have been rolled twice, there are 28 possible states the roller can be in. The challenge is to find the probability of being in each state and the average number of rolls to get to one of the next possible states.

Let's start with the initial state, before any rolls, where each face needs to be rolled twice. It will take one roll only to advance to the state where five faces need to be rolled twice and one, whichever was just rolled, needs to be rolled once only.

From that state, there are two ways to go. One possibility, with probability 5/6, is the roller will roll a different side as the previous roll, leaving four faces that need to be rolled twice and two that need to be rolled once. The second possibility is the roller rolls the same side as the first roll again, with probability 1/6, leaving five faces that need to be rolled twice, and one that has satisfied the condition of needing to be rolled twice. It takes one roll to advance to one of these states.

For the following tables, think of the a scorecard with two rows and six columns. When any face is rolled the first time, it will be crossed off the first row. When it is rolled a second time, it will be crossed off the second column.

In the following table, the column headings show how many faces are left in the first row of the scorecard. The row headings show how many faces are left in the second row of the scorecard. For example, the cell for 2 in the first row and 4 in the second row is 0.595778. That means that there is a 59.58% chance that at some point in the experiment two faces will remain in the first row and four faces in the second row of the scorecard. In other words, two numbers need to be rolled twice and four need to be rolled once.

Second row	First row						
	6	5	4	3	2	1	0
6	1.000000	1.000000	0.833333	0.555556	0.277778	0.092593	0.015432
5		0.166667	0.444444	0.633333	0.565185	0.303235	0.076079
4			0.088889	0.342222	0.595778	0.540477	0.211198
3				0.085556	0.383444	0.660987	0.431527
2					0.127815	0.568473	0.715764
1						0.284236	1.000000
0							1.000000

The next table shows the average number of rolls to get out of that state to another state. The probability of getting out of that state on the next roll is always the number of faces in the second row divided by 6. The expected number of rolls it takes is the inverse of that probability. I won't get into why, but it involves a common infinite series formula.

Second Row	First row						
	6	5	4	3	2	1	0
6	1	1	1	1	1	1	1
5		1.2	1.2	1.2	1.2	1.2	1.2
4			1.5	1.5	1.5	1.5	1.5
3				2	2	2	2
2					3	3	3
1						6	6

The next table is the product of the cells in the many body of the two tables above. It represents the average number of rolls the experiment will be in that state.

Second Row	First row						
	6	5	4	3	2	1	0
6	1.000000	1.000000	0.833333	0.555556	0.277778	0.092593	0.015432
5	0.000000	0.200000	0.533333	0.760000	0.678222	0.363881	0.091295
4	0.000000	0.000000	0.133333	0.513333	0.893667	0.810715	0.316797
3	0.000000	0.000000	0.000000	0.171111	0.766889	1.321974	0.863054
2	0.000000	0.000000	0.000000	0.000000	0.383444	1.705419	2.147291
1	0.000000	0.000000	0.000000	0.000000	0.000000	1.705419	6.000000

The sum of the body of the table above is 24.1338692. This reflects the number of rolls in all states combined. In other words, the number of rolls to get to the (0,0) state.

Solution 2

Solution 2 uses integration. This method does require some skills in calculus, but is quick with the use of an integration calculator.

Instead of using a die, one roll at a time, think of a "roll event" happening randomly and suddenly, like a neutrino passing through your little toe on your left foot. Assume the time between these events has a memory-less property, where the past does not matter, with an average time between events of one unit of time. In other words, the time between events follows an exponential distribution.

Assuming an event has happened, there is a one in six chance it is any of the six die faces. So, the time between events for any given die face is six units of time.

This method works because the average time between events will be the same as if an actual die is rolled. For purposes of adjudicating the bet, it doesn't matter if the time between events is random. We can integrate over all time for the chance that the objective has not been achieved. The sum will be the average number of rolls.

Per the Poisson distribution, for any given side, after x years, the probability it has not been rolled yet is $e^{-\left(\frac{x}{6}\right)}$. Likewise, the probability any given side has been rolled exactly once is $e^{-\left(\frac{x}{6}\right)}\left(\frac{x}{6}\right)$.

Thus, the probability any given has been rolled twice or more is 1 - $e^{-\left(\frac{x}{6}\right)}\left(1+\frac{x}{6}\right)$.

The probability all six sides have been rolled at least twice is $\left(1-e^{-\left(\frac{x}{6}\right)}\left(1+\frac{x}{6}\right)\right)^6$.

So, the probability at least one side has not been rolled is 1 - $\left(1-e^{-\left(\frac{x}{6}\right)}\left(1+\frac{x}{6}\right)\right)^6$.

We not need to integrate that over all time for the average time at least one face will not have been rolled twice.

$$\int_0^{\infty} 1-\left(1-e^{-\left(\frac{x}{6}\right)}\left(1+\frac{x}{6}\right)\right)^6 dx$$

Here is where an integration calculator comes in handy, like the one at www.integral-calculator.com.

To use that calculator specifically, put in "1- (1 - exp(-x/6)*(1 + x/6))^6" in the field titled "calculate the integral of." Under options, put for the upper bound of integration ∞ and 0 for the lower bound. Then click "go."

As that calculator shows, before considering the limits of integration, the integral is (-6*x-36)*e^(-x/6)-36*e^(-x/6)-(5*(-3*x^2-18*x-54)*e^(-x/3))/12-5*(-3*x-9)*e^(-x/3)+45*e^(-x/3)+(5*(-2*x^3-12*x^2-48*x-96)*e^(-x/2))/54+(5*(-2*x^2-8*x-16)*e^(-x/2))/3+10*(-2*x-4)*e^(-x/2)-40*e^(-x/2)+(5*(6*x^4+36*x^3+162*x^2+486*x+729)*e^(-(2*x)/3))/1728+(5*(12*x^3+54*x^2+162*x+243)*e^(-(2*x)/3))/144+(5*(6*x^2+18*x+27)*e^(-(2*x)/3))/8+(5*(6*x+9)*e^(-(2*x)/3))/2+(45*e^(-(2*x)/3))/2-((3750*x^5+22500*x^4+108000*x^3+388800*x^2+933120*x+1119744)*e^(-(5*x)/6))/4050000-((3750*x^4+18000*x^3+64800*x^2+155520*x+186624)*e^(-(5*x)/6))/135000-((750*x^3+2700*x^2+6480*x+7776)*e^(-(5*x)/6))/2250-((150*x^2+360*x+432)*e^(-(5*x)/6))/75-((30*x+36)*e^(-(5*x)/6))/5-(36*e^(-(5*x)/6))/5-((-x^6-6*x^5-30*x^4-120*x^3-360*x^2-720*x-720)*e^(-x))/46656-((-x^5-5*x^4-20*x^3-60*x^2-120*x-120)*e^(-x))/1296-(5*(-x^4-4*x^3-12*x^2-24*x-24)*e^(-x))/432-(5*(-x^3-3*x^2-6*x-6)*e^(-x))/54-(5*(-x^2-2*x-2)*e^(-x))/12-(-x-1)*e^(-x)+e^(-x)

However, we don't need to concern ourselves with that. The calculator shows that integration from 0 to ∞ the answer is 390968681/16200000 = apx. 24.13386919753086.