The probability of Yahtzee

Camden Mac Leod

December 26, 2021

Contents

1 Next-roll probability vs. next-few-roll probability

1.1 Next-roll probability

This is simple to calculate. If you are rolling five dice and need five of them to match, you have a simple problem that can be calculated with binomial probability. Binomial probability predicts yes/no (binary) results from a set of *independent* experiments. This is useful in predicting the next roll and would therefore be useful on the final dice roll because it does not predict the outcome of a later roll. To calculate the binomial probability, first you need to calculate the number of combinations from your provided numbers.

- nCr is the number of combinations in a given set
- r is the number of dice you need to match the first number
- n is the number of dice you are rolling

$$nCr = \frac{n!}{r!(n-r)!}$$

You then take this result and plug it into a binomial probability formula using the same values for the variables. p is the probability of an individual occurrence, so in this case it's $1/6$.

$$P(x = r) = nCr \times p^r \times (1-p)^{n-r}$$

If we wanted to calculate the chances for a Yahtzee then, you see that after one roll your chances are roughly 0.08%, or $\frac{8}{10{,}000}$. Notice below that we use 4 instead of 5 because the first dice rolled doesn't matter. The probability of rolling all five dice to be a certain value (like all 1's) is roughly 0.012%, but that multiplied by 6 gives us roughly 0.08%.

$$P(x) = \frac{4!}{(4-4)!4!} \times (1/6)^4 \times (1 - 1/6)^{4-4} \approx 0.08\%$$

1.2 Next-few-roll probability

Next-roll probability is simple. You plug the numbers into the formula and get your expected result. The task becomes infinitely more complicated when you have more rolls involved. A smart player might roll a Yahtzee over the course of three turns like this:

1. 5 5 3 1 2
2. 5 5 5 4 6
3. 5 5 5 5 5

Their chances of getting a Yahtzee on the first attempt was 0.08% as we calculated above. But as they set aside other dice and got closer to a Yahtzee, their chances went up. How can you draw a conclusion on the *overall* probability of getting a Yahtzee over the course of all three rolls, rather than just one? What if the player was allowed five rolls? At how many rolls does your chance of obtaining a Yahtzee go above 50%?

These all require a *dependent* formula that takes into account the previous rolls and calculates a possible outcome.

2 Markov chains and transition matrices

A Markov chain describes the likelihood of moving from one state to the next solely based on the previous state. Markov chains are useful in many real-world situations like cruise control, typing prediction, and page ranking used at Google. The process is simple in theory, but a bit tricky when you have so many plugs and slots. We should be able to calculate our answer given the transition matrix below in a 5x5 grid (five being the amount of dice available to roll).

$$\begin{pmatrix} ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \end{pmatrix}$$

We can fill in each question mark with a fraction that describes our likelihood of staying in that state or progressing to the next. It's important to start small, so we'll come back to this later.

3 Probability with less dice

It's tricky to calculate the probability of achieving different goals when you have five dice. So we'll start small and it should make the whole process much easier to understand.

3.1 One dice

This is as simple as it gets. When you roll a single die your chances of getting any result are $1/6$ because the die is six-sided. If we want a specific result, our chance then is $1/6$ and our chance of *not getting that result* is $5/6$.

$$\begin{pmatrix} A & 1/6 \\ B & 5/6 \end{pmatrix}$$

This is the last time we'll assume we want a specific result, like getting exactly a 6. That's a whole other ballgame. In a Yahtzee it doesn't matter what numbers are present on the dice as long as they match.

3.2 Two dice

This is still very simple. In a given two-dice roll you have $6 \times 6 = 36$ possible outcomes. When you roll two dice one of two things will happen.

- Both dice match (AA)
- Neither dice match (AB)

We can calculate our chances of both dice matching to be $^{6}/_{36}$. This makes sense because there are six ways you can score a pair from two dice: $\{1,1\}, \{2,2\}, \{3,3\}, \{4,4\}, \{5,5\}, \{6,6\}$. So now we can fill in our dice probability chart.

$$\begin{pmatrix} A & A & ^{6}/_{36} \\ A & B & ^{30}/_{36} \end{pmatrix}$$

Remember, we can only get one of the two outcomes so they have to add up to 1. $^{6}/_{36} + ^{30}/_{36} = 1$ so we know we're good.

3.3 Three dice

Getting a little more complicated but not by much. In a given three-dice roll you have $6 \times 6 \times 6 = 216$ possibilities. When you roll three dice one of the following will happen.

- They all match (AAA)
- Two of them match (AAB)
- None of them match (ABC)

There are six ways all the dice can match: $\{1,1,1\}, \{2,2,2\}, \{3,3,3\}, \{4,4,4\}, \{5,5,5\}, \{6,6,6\}$. In our probability chart we see the following values:

$$\begin{pmatrix} A & A & A & ^{6}/_{216} \\ A & A & B & ^{90}/_{216} \\ A & B & C & ^{120}/_{216} \end{pmatrix}$$

To calculate the second row we realize that there are six possibilities for A, five for B, and three places where B could be present. $6 \times 5 \times 3 = ^{90}/_{216}$.

To calculate the final row we could take $^{216}/_{216} - ^{6}/_{216} - ^{90}/_{216}$, but it's better to check our math. The chances of rolling the first die one particular number is 6. The chances of rolling another die that is different from the first is $^{5}/_{6}$. The chances of rolling a third dice different from both the first dice is $^{4}/_{6}$. This gives us $^{5}/_{6} \times ^{4}/_{6} = ^{120}/_{216}$.

3.4 Four dice

There are $6^4 = 1296$ possibilities here.

- They all match (AAAA) - $^6/_{1296}$
- Three match (AAAB) - $6 \times 5 \times 4 = {^{120}/_{1296}}$
- Two pairs (AABB) - $6 \times 5 \times 3 = {^{90}/_{1296}}$
- One pair with two others (AABC) - $6 \times 5 \times 4 \times 3 \times 2 = {^{720}/_{1296}}$
- None of them match (ABCD) - $^6/_6 \times {^5/_6} \times {^4/_6} \times {^3/_6} = {^{360}/_{1296}}$

Our new concepts here are the two pairs and one pair with two others.

Two pairs is $6 \times 5 \times 3$ because this is the same probability as seen in the three dice stage except another dice is added. It's the same equation, but because we added a new dice it's $^1/_6$ less, e.g. $^{90}/_{216}$ vs. $^{90}/_{1296}$.

One pair with two others is the most likely option here coming in at $^{720}/_{1296}$ (55%). The calculation is $6 \times 5 \times 4 \times 3 \times 2$ because there is a probability of 5 for B in 3 positions and 4 for C in 2 positions. So written in different terms, the calculation is $(6 \times 1) \times (5 \times 3) \times (4 \times 2)$.

Now we can draw our probability chart for four dice.

$$\begin{pmatrix} A & A & A & A & ^6/_{1296} \\ A & A & A & B & ^{120}/_{1296} \\ A & A & B & B & ^{90}/_{1296} \\ A & A & B & C & ^{720}/_{1296} \\ A & B & C & D & ^{360}/_{1296} \end{pmatrix}$$

3.5 Five dice

This is as hard as it gets, fortunately. There are $6^5 = 7776$ possibilities here. We'll use the official Yahtzee names here.

- Yahtzee (AAAAA) - $^6/_{7776}$
- Four-of-a-kind (AAAAB) - $6 \times 5 \times 5 = {^{150}/_{7776}}$
- Full house (AAABB) - $6 \times 5 \times 5 \times 2 = {^{300}/_{7776}}$
- Three-of-a-kind (AAABC) - $6 \times 5 \times 5 \times 4 \times 2 = {^{1200}/_{7776}}$
- Small straight (AABCD) - $6 \times 5 \times 5 \times 4 \times 3 \times 2 = {^{3600}/_{7776}}$
- Two pairs (AABBC) - $6 \times 5 \times 3 \times 4 \times 5 = {^{1800}/_{7776}}$
- Long straight (ABCDE) - $^6/_6 \times {^5/_6} \times {^4/_6} \times {^3/_6} \times {^2/_6} = {^{720}/_{7776}}$

Below is our probability chart for all five dice.

$$\begin{pmatrix} A & A & A & A & A & {}^{6}\!/_{7776} \\ A & A & A & A & B & {}^{150}\!/_{7776} \\ A & A & A & B & B & {}^{300}\!/_{7776} \\ A & A & A & B & C & {}^{1200}\!/_{7776} \\ A & A & B & C & D & {}^{3600}\!/_{7776} \\ A & A & B & B & C & {}^{1800}\!/_{7776} \\ A & B & C & D & E & {}^{720}\!/_{7776} \end{pmatrix}$$

Looking at the probability here, we see that you are more than twice as likely to roll a small straight than a three-of-a-kind, and twenty-four times more likely to roll a small straight than a four-of-a-kind.

Both a three-of-a-kind and four-of-a-kind can score a maximum of 30 points (the value of a small straight). You have a $^{1}/_{7776}$ chance of scoring *the same amount of points* in a three-of-a-kind or four-of-a-kind (excluding a Yahtzee), because the roll would have to consist of five 6's. That's 0.012%. Not great odds.

Strategically, this means a player should always cancel out a four-of-a-kind and three-of-a-kind before a small straight if it comes down to it. It also means that you should wait until later in the game to attempt a small straight and potentially re-roll one that you might get from chance, because you are so likely to get one again (55.56%).

4 Creating our transition matrix

Now we get to the exciting part. A transition matrix shows us the fractional chances of getting from one state to the next. This will go hand in hand with our Markov chain later.

$$\begin{pmatrix} ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \end{pmatrix}$$

We can fill in our unavailable spaces with 0's since a smart player would never re-roll dice they already had matching. That is, if you rolled five dice and three matched, you would always re-roll exactly two.

$$\begin{pmatrix} ? & ? & ? & ? & ? \\ 0 & ? & ? & ? & ? \\ 0 & 0 & ? & ? & ? \\ 0 & 0 & 0 & ? & ? \\ 0 & 0 & 0 & 0 & ? \end{pmatrix}$$

We also know that once you roll five matching dice you would always keep it. This means our final position $(5, 5)$ has a 1 in it.

$$\begin{pmatrix} ? & ? & ? & ? & ? \\ 0 & ? & ? & ? & ? \\ 0 & 0 & ? & ? & ? \\ 0 & 0 & 0 & ? & ? \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We can fill in the numbers in the column above the 1 because they represent our chances of moving from the state below to the state above. In this case, it's always $\frac{1}{6}^n$ where n is the distance from 1.

$$\begin{pmatrix} ? & ? & ? & ? & 1/1296 \\ 0 & ? & ? & ? & 1/216 \\ 0 & 0 & ? & ? & 1/36 \\ 0 & 0 & 0 & ? & 1/6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In English this means if we roll five dice our chances of getting all five matching is in the position $(5,1)$ or $1/1296$, which is 0.08%. You should start to become familiar with these numbers and as we've found in several other places, our chances of rolling a Yahtzee *is* 0.08% so we know we're on the right track.

An important note is that each row here must equal 1 because you have to have at least one of the outcomes occur. This makes it easy to figure out some of our next rows and double check our math. Let's fill in the fourth row.

$$\begin{pmatrix} ? & ? & ? & ? & 1/1296 \\ 0 & ? & ? & ? & 1/216 \\ 0 & 0 & ? & ? & 1/36 \\ 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Thus, our fourth row (which represents when a person is rolling one die) tells us we have a $1/6$ chance of getting the number we need and a $5/6$ chance of staying at that level.

For our third row we have two dice we are re-rolling. We have a $5/6 \times 5/6 = 25/36$ chance of not getting what we need. This means we have $10/36$ chance of getting one matching dice but not the other ($1/6 \times 5/6$).

$$\begin{pmatrix} ? & ? & ? & ? & 1/1296 \\ 0 & ? & ? & ? & 1/216 \\ 0 & 0 & 25/36 & 10/36 & 1/36 \\ 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Our fourth row is the hardest to calculate. In our fourth row we are re-rolling three dice. We have a $1/6 \times 1/6 \times 5/6 \times 3 = 15/216$ chance of getting a pair from three. Now you might be catching on now and thinking that the chance of missing all three would be $5/6 \times 5/6 \times 5/6 = 125/216$. This is incorrect because a player might roll a pair on the first roll and then roll three matching dice (that don't match the pair) on their second roll. They would then keep the three and re-roll the pair.

In order for us to calculate this we simply need to subtract the chance of getting all three matching from that number. If we go back to our three dice section we see a $^6/_{216}$ chance of getting three matching. We also have to subtract one from that number because one of those matches would create a Yahtzee. This gives us $^{125}/_{216} - ^5/_{216} = ^{120}/_{216}$.

Lastly, we have the chance we get three matching and re-roll the other two. This would be $^5/_6 \times ^5/_6 \times ^1/_6 \times 3$. This is simply the reverse of our above equation which got us $^{15}/_{216}$. That leaves us $^{75}/_{216}$, which we also add that magic $^5/_{216}$ chance that we jumped ship and are re-rolling the pair instead of the three-of-a-kind. Thankfully, adding up all four of these equals $^{216}/_{216}$.

$$\begin{pmatrix} ? & ? & ? & ? & ^1/_{1296} \\ 0 & ^{120}/_{216} & ^{80}/_{216} & ^{15}/_{216} & ^1/_{216} \\ 0 & 0 & ^{25}/_{36} & ^{10}/_{36} & ^1/_{36} \\ 0 & 0 & 0 & ^5/_6 & ^1/_6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We only have one row left, which is much easier than the last one because we don't need to assume they switch/don't switch their current path.

- Cell $(1,1)$ represents your chance of rolling and nothing matching. That's $^5/_6 \times ^4/_6 \times ^3/_6 \times ^2/_6 = ^{120}/_{1296}$. If you did roll with nothing matching you would essentially have a large straight. You'll notice in the five dice section that 120 is six times less than your chance of getting a large straight so we know we're on the right path.
- Cell $(1,2)$ is your chance of rolling a pair. Again we can reference our five dice section. There's two possibilities of a pair; $^{1800}/_{7776} + ^{3600}/_{7776} \div 6 = ^{900}/_{1296}$.
- Cell $(1,3)$ is your chance of rolling three-of-a-kind from five. That's $^{1200}/_{7776} + ^{300}/_{7776} \div 6 = ^{250}/_{1296}$.
- Cell $(1,4)$ is your chance of rolling a four-of-a-kind from five. That's $^{25}/_{1296}$.

$$\begin{pmatrix} ^{120}/_{1296} & ^{900}/_{1296} & ^{250}/_{1296} & ^{25}/_{1296} & ^1/_{1296} \\ 0 & ^{120}/_{216} & ^{80}/_{216} & ^{15}/_{216} & ^1/_{216} \\ 0 & 0 & ^{25}/_{36} & ^{10}/_{36} & ^1/_{36} \\ 0 & 0 & 0 & ^5/_6 & ^1/_6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Yay! Our whole matrix is filled out.

5 Calculating probability over several rounds

This is simple now that we have our base matrix finished. We simply need to send it to the third power to analyze its probability over three rolls.

$$P_3 = \begin{pmatrix} ^{120}/_{1296} & ^{900}/_{1296} & ^{250}/_{1296} & ^{25}/_{1296} & ^{1}/_{1296} \\ 0 & ^{120}/_{216} & ^{80}/_{216} & ^{15}/_{216} & ^{1}/_{216} \\ 0 & 0 & ^{25}/_{36} & ^{10}/_{36} & ^{1}/_{36} \\ 0 & 0 & 0 & ^{5}/_{6} & ^{1}/_{6} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^3$$

This is a slightly intensive process with such a big matrix. It's better to just leave it to a computer because a task like this could take quite a long time by hand. Here are our results:

$$P_3 = \begin{pmatrix} ^{6000}/_{7558272} & ^{1935000}/_{7558272} & ^{3419375}/_{7558272} & ^{1850000}/_{7558272} & ^{347897}/_{7558272} \\ 0 & ^{1296000}/_{7558272} & ^{3294000}/_{7558272} & ^{2389500}/_{7558272} & ^{584772}/_{7558272} \\ 0 & 0 & ^{2531250}/_{7558272} & ^{3685500}/_{7558272} & ^{1341522}/_{7558272} \\ 0 & 0 & 0 & ^{4374000}/_{7558272} & ^{3184272}/_{7558272} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

If we convert those to percentages it's much more readable:

$$P_3 = \begin{pmatrix} 0.1\% & 25.6\% & 45.2\% & 24.5\% & 4.6\% \\ 0 & 17.1\% & 43.6\% & 31.6\% & 7.7\% \\ 0 & 0 & 33.5\% & 48.8\% & 17.7\% \\ 0 & 0 & 0 & 57.9\% & 42.1\% \\ 0 & 0 & 0 & 0 & 100\% \end{pmatrix}$$

You'll notice each row adds up to 100% exactly how our first transition matrix had fractions that added up to 1.

6 What do our results tell us?

Our most important number is in cell $(1, 5)$. This tells us the chance of a smart player getting a Yahtzee over three turns is roughly 4.6%. Much higher than over one round (0.08%). Below are some more conclusions we can draw.

- If the player was allowed five rolls their chance of getting a Yahtzee would rise from 4.6% to 17%.
- After 10 rolls your chance of obtaining a Yahtzee is higher than 50%.
- Your chance of getting a four-of-a-kind over three rolls is 24.5%.
- Your chance of getting a three-of-a-kind is 45.2%.
- You're more likely to get a three-of-a-kind (45.2%) than a pair (25.6%). This makes sense because you would have to be pretty unlucky to never get more than a pair from rolling three times.

7 Rolls-to-probability table

x rolls	Pair	Three-of-a-kind	Four-of-a-kind	Yahtzee
1	69.4%	19.3%	1.9%	0.08%
2	45.0%	40.9%	12.0%	1.3%
3	25.6%	45.2%	24.5%	4.6%
4	14.3%	40.9%	34.7%	10.1%
5	7.9%	33.7%	41.3%	17.1%
6	4.4%	26.3%	44.3%	24.9%
7	2.5%	19.9%	44.5%	33.1%
8	1.4%	14.7%	42.8%	41.0%
9	0.8%	10.7%	39.9%	48.6%
10	0.4%	7.7%	36.3%	55.6%
11	0.2%	5.5%	32.4%	61.8%
12	0.1%	3.9%	28.6%	67.4%
13	0.07%	2.7%	24.9%	72.2%
14	0.04%	2.0%	21.5%	76.5%
15	0.02%	1.3%	18.5%	80.1%
16	0.01%	1.0%	15.8%	83.2%
17	0.007%	0.7%	13.4%	85.9%
18	0.004%	0.5%	11.4%	88.2%
19	0.002%	0.3%	9.6%	90.0%
20	0.001%	0.2%	8.0%	91.6%
21	0.001%	0.2%	6.8%	93.0%
22	0%	0.1%	5.7%	94.2%
23	0%	0.08%	4.8%	95.1%
24	0%	0.05%	4.0%	95.9%
25	0%	0.04%	3.4%	96.6%

References

[1] Beast, Altered, and amd. "How to Cube a Matrix, Specifically a Transition (Probability) Matrix." Mathematics Stack Exchange, 4 Sept. 2019, https://math.stackexchange.com/questions/3343932/how-to-cube-a-matrix-specifically-a-transition-probability-matrix.

[2] Berry, Nick. *Data Genetics*, https://datagenetics.com/index.html

[3] Moehring, Brian, and CyanCoding. "Binomial Probability Doesn't Give the Correct Number?" Mathematics Stack Exchange, 27 Dec. 2021, https://math.stackexchange.com/a/4343342/1008984.

[4] NagwaEd. "Squaring and Cubing a Matrix." Nagwa, https://www.nagwa.com/en/videos/689158201380/.

[5] Pilling, Matthew, and CyanCoding. "How to Predict the Likelihood of Multiple Future Occurrences Based on Probability." Mathematics Stack Exchange, 26 Dec. 2021, https://math.stackexchange.com/a/4342534/1008984.