Probability of Dice Sequences

Jonathan Stokes

May 9, 2024

In rolling 6 dice what is the probability of rolling $1, 2, 3, 4$? Or more generally in rolling n dice what is the probability for rolling a length m sequence starting with 1? It is straight forward to estimate the probability of rolling these sequences using $100,000$ monte carlo simulations per dice count, see fig. 1.

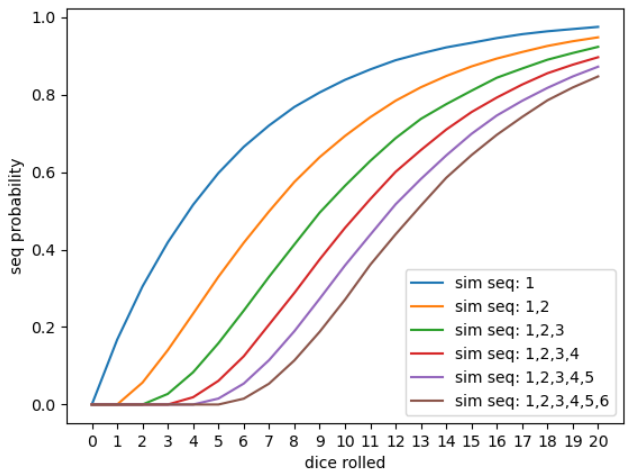

Figure 1: Monte Carlo simulation of the probability of rolling different sequences starting at 1 using $100,000$ rolls per dice count.

At first this looks similar to the probability of drawing a straight in 5 card draw poker. Lets define the event that one draws a straight in 5 card draw poker

$$A_5 : \text{In drawing 5 cards one draws a straight.}$$

If one draws a straight in 5 card draw poker including a royal flush there are

1. $\binom{10}{1}$ ways to choose the low card in the straight (recalling that the ace may be 1 or follow the king).

2. $\binom{4}{1}^5$ ways to select the card suits.

and there are $\binom{52}{5}$ ways of drawing 5 cards. So the probability of drawing the straight in 5 card draw poker is

$$P(A_5) = \frac{\binom{10}{1}\binom{4}{1}^5}{\binom{52}{5}} \approx 0.00394 \tag{1}$$

Now lets define the more general event

$$A_n : \text{In drawing } n \text{ cards one draws a straight.}$$

It follows that the probability of event A_n for $n \in \{2 \dots 13\}$ is

$$P(A_n) = \frac{\binom{15-n}{1}\binom{4}{1}^n}{\binom{52}{n}} \tag{2}$$

Now lets try applying the same reasoning to the probability of rolling a length m sequence in a roll of n dice for $n \geq m$ if we drop the requirement that the sequence starts at 1. Lets define this event as

$$A_{n,m} : \text{A length } m \text{ sequence appears in a roll of } n \text{ dice.}$$

Using the same reasoning as in drawing a straight of cards there are

1. $\binom{7-m}{1}$ ways to choose the low number in the sequence.
2. $\binom{6}{1}^{n-m}$ ways to roll the remaining dice.

and there are $\binom{6}{1}^n$ ways of rolling n dice. Therefore it would seem that the probability of rolling a length m sequence in a roll of n dice is,

$$P(A_{n,m}) = \frac{\binom{7-m}{1}\binom{6}{1}^{n-m}}{\binom{6}{1}^n} = \frac{\binom{7-m}{1}}{\binom{6}{1}^m} = \frac{7-m}{6^m} \tag{3}$$

But looking at fig. 1 this is clearly incorrect for $m = 6$ and sequence $1, 2, 3, 4, 5, 6$ as eq. (3) is constant in the number of dice rolled, n.

What did we do wrong? The term $\binom{7-m}{1}$ implicitly assumes that on any roll of the dice there are 6 different numbers to choose from. In the case of cards this type of assumption is correct since every deck has 13 distinct values (14 if the ace can be low and high). However in the case of a roll of 6 dice we may not roll a 1 leaving at most 5 numbers to choose from in which case there are not $\binom{7-m}{1}$ ways to choose the lowest number in the sequence.

So what is the solution? Lets start thinking about the case where $m = 2$ and $n = 3$. What is the probability of rolling a 1 and a 2 when rolling 3 dice? First lets define the following events.

A: 1 appears in a roll of 3 dice.

B: 2 appears in a roll of 3 dice.

To find the probability of events A and B we need to find the probability of $\neg A$ and $\neg B$ which is the probability of rolling any of the other 5 faces of the 3 independent dice. Therefore

$$P(\neg A) = P(\neg B) = \left(\frac{5}{6}\right)^3 \tag{4}$$

and by analogy the probability of not rolling a 1 and a 2 is

$$P(\neg A \cap \neg B) = \left(\frac{4}{6}\right)^3 \tag{5}$$

It follows from eq. (4) that the probability of rolling either a 1 or a 2 is

$$P(A) = P(B) = 1 - P(\neg A) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216} \tag{6}$$

To answer our initial question above we need to find $P(A \cap B)$. Now the law of total probability states

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B) \tag{7}$$

Rearranging eq. (7) gives

$$P(A|B) = \frac{P(A) - P(A|\neg B)P(\neg B)}{P(B)} \tag{8}$$

Now since $P(A|\neg B) + P(\neg A|\neg B) = 1$ we can replace $P(A|\neg B)$ with $(1 - P(\neg A|\neg B))$ giving

$$P(A|B) = \frac{P(A) - (1 - P(\neg A|\neg B))P(\neg B)}{P(B)} \tag{9}$$

Notice it follows that

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) & (10)\\ &= P(A) - (1 - P(\neg A|\neg B))P(\neg B) & (11)\\ &= P(A) - P(\neg B) + P(\neg A \cap \neg B) & (12) \end{aligned}$$

Plugging in the values above into eq. (12) gives $P(A \cap B) \approx 0.139$.

Now given eq. (12) does this generalize for $n \geq 2$ dice rolled? Lets first define the following events

$$A_n\text{: 1 appears in a roll of } n \text{ dice.} \tag{13}$$

$$B_n\text{: 2 appears in a roll of } n \text{ dice.} \tag{14}$$

$$\begin{aligned} P(A_n \cap B_n) &= P(A_n) - P(\neg B_n) + P(\neg A_n \cap \neg B_n) & (15)\\ &= 1 - \left(\frac{5}{6}\right)^n - \left(\frac{5}{6}\right)^n + \left(\frac{4}{6}\right)^n & (16)\\ &= 1 - 2\left(\frac{5}{6}\right)^n + \left(\frac{2}{3}\right)^n & (17) \end{aligned}$$

and as $n \to \infty$, $P(A_n \cap B_n) \to 1$ as expected.

Can the solution above be generalized to the case of a length m sequence? Lets first consider the case where $m = 3$. In this case, by induction from eq. (12) we write,

$$\begin{aligned} P(A_n \cap B_n \cap C_n) &= P(A_n) - P(\neg B_n) - P(\neg C_n) & (18)\\ &+ P(\neg A_n \cap \neg B_n) + P(\neg A_n \cap \neg C_n) + P(\neg B_n \cap \neg C_n) & (19)\\ &- P(\neg A_n \cap \neg B_n \cap \neg C_n) & (20)\\ &= 1 - 3\left(\frac{5}{6}\right)^n + 3\left(\frac{4}{6}\right)^n - \left(\frac{3}{6}\right)^n & (21) \end{aligned}$$

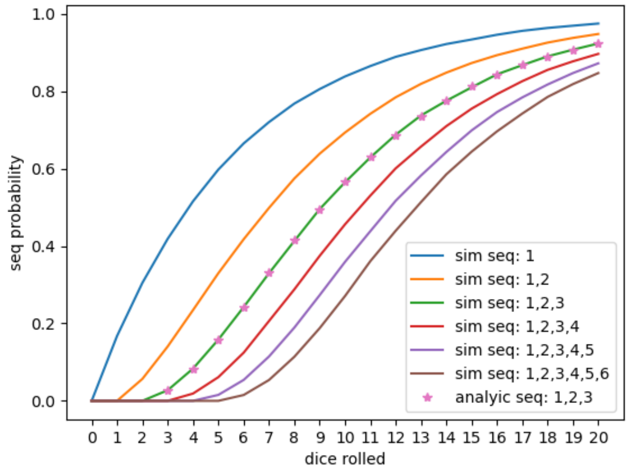

Figure 2: Plot of eq. (21)

Figure 2 compares the analytic solution given by eq. (21) with the simulated results. As expected they are similar.

Now we want an equation expressing the probability of rolling the sequence of dice 1 to m where $m \in \{1, 2, 3, 4, 5, 6\}$. Given $n \geq m$ lets define the events

$$A_{n,i}\text{: } i \text{ appears in a roll of } n \text{ dice.}$$

for $i \in \{1 \dots m\}$.

Now given eq. (17) and eq. (21) again by induction we can derive a closed form solution for the probability of a sequence of length m starting at 1 appearing in a given roll of dice as,

$$P\left(\bigcap_{i=0}^{m} A_{n,i}\right) = \sum_{i=0}^{m} (-1)^i \binom{m}{i} \left(\frac{6-i}{6}\right)^n \tag{22}$$

Using this equation to compute the analytic probability of rolling a dice sequence starting at 1 of length m and comparing these to the simulated results gives.

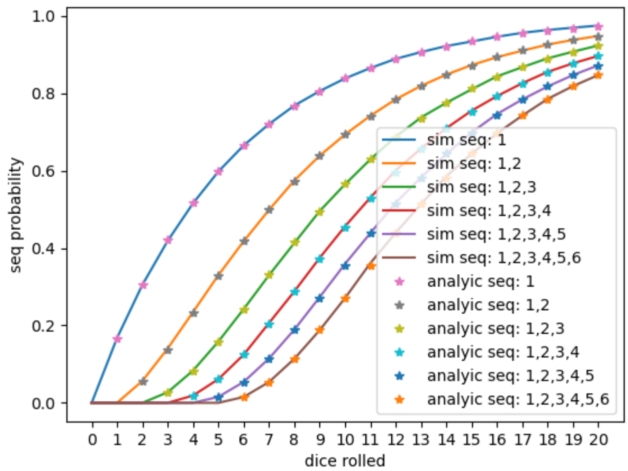

Figure 3: Plot of eq. (22)