Optiver Prove It: Episode 3

Gabriel Romon[1]

[1] *e-mail:* gabriel.s.romon@gmail.com

The video containing the problem statement is available here: https://www.youtube.com/watch?v=YKE5SRXK_nI.

1. Expected number of rolls with an n-sided die

Let $n \geq 1$ denote the number of sides of the die, and let $X_1, X_2 \ldots$ be i.i.d. random variables uniformly distributed over $\{1, \ldots, n\}$. For each integer $p \geq 1$, let $S_p = \sum_{k=1}^{p} X_p$ and define the random variable N as

$$N = \inf\{p \geq 1 : S_p \geq n\}.$$

1.1. A first-step analysis

For each $i \in \{1, \ldots, n\}$, let e_i denote the expected number of rolls needed to reach a sum $\geq i$. Clearly, $e_1 = 1$ and we are looking for the value of e_n.

If on the first throw we obtain a value $j < i$, then the game starts over on the next turn, except that the target to be surpassed is now $i - j$. For convenience, let $N_i = \inf\{p \geq 1 : S_p \geq i\}$. Conditioning on the first throw, we obtain therefore

$$\begin{aligned} e_i = \mathbb{E}[N_i] &= \sum_{j=1}^{i-1} \mathbb{E}[N_i|X_1 = j]\mathbb{P}(X_1 = j) + \sum_{j=i}^{n} \mathbb{E}[N_i|X_1 = j]\mathbb{P}(X_1 = j) \\ &= \sum_{j=1}^{i-1} (1 + e_{i-j})\frac{1}{n} + \sum_{j=i}^{n} 1 \cdot \frac{1}{n} \\ &= 1 + \frac{1}{n}\sum_{j=1}^{i-1} e_{i-j}. \end{aligned}$$

This recursion makes it possible, for a given value of n, to numerically compute $e_2, e_3, \ldots$ up to e_n. However, obtaining a closed form for e_n seems tedious and we will not proceed further with this approach.

1.2. A combinatorial approach

Note that the random variable N takes values in $\{1, \ldots, n\}$ and

$$\mathbb{E}[N] = \sum_{p=0}^{\infty} \mathbb{P}(N > p) = 1 + \sum_{p=1}^{n-1} \mathbb{P}(N > p) = 1 + \sum_{p=1}^{n-1} \mathbb{P}(S_p < n).$$

Next,

$$\begin{aligned}
\mathbb{P}(S_p < n) &= \mathbb{P}\left((X_1,\dots,X_p) \in \left\{(x_1,\dots,x_p) \in \{1,\dots,n\}^p : \sum_{k=1}^{p} x_k < n\right\}\right)\\
&\overset{(i)}{=} \frac{\operatorname{card}\{(x_1,\dots,x_p) \in \{1,\dots,n\}^p : \sum_{k=1}^{p} x_k < n\}}{\operatorname{card}(\{1,\dots,n\}^p)}\\
&= \frac{1}{n^p}\sum_{s=p}^{n-1} \operatorname{card}\left\{(x_1,\dots,x_p) \in \{1,\dots,n\}^p : \sum_{k=1}^{p} x_k = s\right\}\\
&\overset{(ii)}{=} \frac{1}{n^p}\sum_{s=p}^{n-1}\binom{s-1}{p-1}\\
&\overset{(iii)}{=} \frac{1}{n^p}\binom{n-1}{p},
\end{aligned} \tag{1}$$

where (i) holds because the random vector $(X_1,\dots,X_p)$ is uniformly distributed over the set $\{1,\dots,n\}^p$, (ii) follows from stars and bars, and (iii) from the hockeystick identity.

Thus

$$\mathbb{E}[N] = 1 + \sum_{p=1}^{n-1}\frac{1}{n^p}\binom{n-1}{p} = \sum_{p=0}^{n-1}\binom{n-1}{p}\frac{1}{n^p} = \left(1+\frac{1}{n}\right)^{n-1}.$$

2. Bonus exercise

2.1. Theoretical solution

With the previous notation, n is equal to 6 and the first player wins if and only if the random variable N takes an odd value. Therefore

$$\begin{aligned}
\mathbb{P}(\text{First player wins}) &= \sum_{p=0}^{2}\mathbb{P}(N = 2p+1)\\
&= \sum_{p=0}^{2}\mathbb{P}(N > 2p) - \mathbb{P}(N > 2p+1)\\
&= \sum_{p=0}^{2}\mathbb{P}(S_{2p} < 6) - \mathbb{P}(S_{2p+1} < 6)\\
&= \sum_{p=0}^{2}\frac{1}{6^{2p}}\binom{5}{2p} - \frac{1}{6^{2p+1}}\binom{5}{2p+1} \qquad \text{by (1)}\\
&= \sum_{p=0}^{2}\frac{1}{6^{2p}}\left(\binom{5}{2p} - \frac{1}{6}\binom{5}{2p+1}\right)\\
&= \frac{3125}{7776}\\
&\approx 0.402.
\end{aligned}$$

Consequently $\mathbb{P}(\text{Second player wins}) = \frac{4651}{7776} \approx 0.598$ and you should play second.

2.2. Numerical simulation

The following piece of code repeats the experiment 20 000 times. The output is 0.401 which confirms that our computations are correct.

```
import numpy as np

n = 6
n_reps = 20000
np.random.seed(2024)

X = np.random.randint(1, n+1, (n_reps, n))
X = np.cumsum(X, axis=1) >= 6
X = np.argmax(X, axis=1) - 1 #subtract 1 to convert zero-based into one-based indexing
print(np.mean(X % 2)) #proportion of wins by the first player
```