The choice problem and the M(j) process - a different view

February 6, 2021

1 Definitions

$$\{a_1, \ldots, a_n\} \sim U\left[0,1\right] \tag{1}$$

$$M\left(j\right) = \max\left(\{a_1, \ldots, a_j\}\right), 1 \leq j \leq n \tag{2}$$

$$\{k_1, \ldots, k_n\} \in [0,1] \tag{3}$$

2 Stopping probabilities

We stop at round R_j if $a_j > k_j$ and $a_i < k_i \forall i < j$. We then determine if we win or lose; if we did not stop the maximum will only be revealed at round R_n.

Round	Stop	Continue
1	$1-k_1$	k_1
2	$k_1 \cdot (1-k_2)$	$k_1 \cdot k_2$
3	$k_1 \cdot k_2$	0

Round	Stop	Continue
1	$1-k_1$	k_1
2	$k_1 \cdot (1-k_2)$	$k_1 \cdot k_2$
3	$k_1 \cdot k_2 \cdot (1-k_3)$	$k_1 \cdot k_2 \cdot k_3$
4	$k_1 \cdot k_2 \cdot k_3 \cdot (1-k_4)$	$k_1 \cdot k_2 \cdot k_3 \cdot k_4$
5	$k_1 \cdot k_2 \cdot k_3 \cdot k_4$	0

Table 1: Stopping probabilities for n=3 and n=5

Following the integration procedures from the original paper, we find that the loss probability on round R_n is somewhere between $\left(\frac{n-1}{n}\right) \cdot k_{n-1}^n$ and $\left(\frac{n-1}{n}\right) \cdot k_1^n$ as expected. The best way to understand the formula is to decompose:

$$\left(\frac{n-1}{n}\right) = \sum_{j=1}^{n-1} \frac{1}{j \cdot (j+1)} \tag{4}$$

And then:

$$\left(\frac{n-1}{n}\right) \cdot k_1^n = \sum_{j=1}^{n-1} \frac{k_1^n}{j \cdot (j+1)} \tag{5}$$

As k_j decreases, at every round we carve a bit of the previous unrecognized max CDF when we recognize a loss for $k_j < a_j < M(n)$; this ends up being equivalent to rewrite the formula above as:

$$P_n^L = \sum_{j=1}^{n-1} \frac{k_j^j \cdot \prod_{i=j}^{n-1} k_i}{j \cdot (j+1)} \tag{6}$$

Where on each term of the sum we have a product of ks with the same sum of exponents but with each product decreasing in value. And because:

$$P_n^W + P_n^L = \prod_{i=j}^{n-1} k_i \tag{7}$$

We have the win probability for the last round:

$$P_n^W = \prod_{i=j}^{n-1} k_i - \sum_{j=1}^{n-1} \frac{k_j^j \cdot \prod_{i=j}^{n-1} k_i}{j \cdot (j+1)} \tag{8}$$

3 n=3

From simulation of 10 million draws we can see the distribution of $M(j)$. let's start with just the running maximum, no stopping:

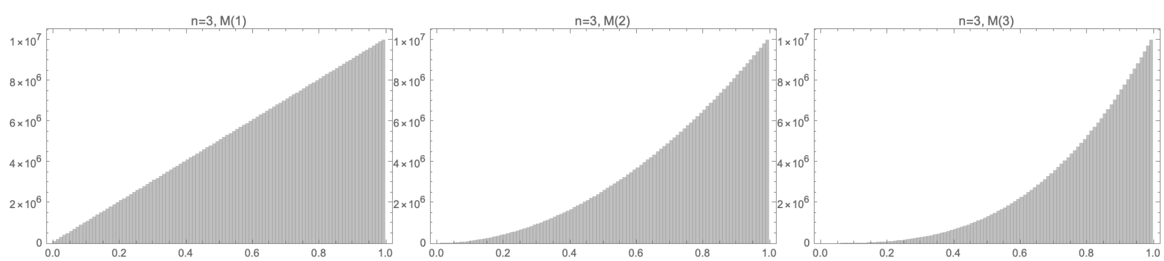

Figure 1: Unconditional CDF for M(j) - running maximum, n=3

Defining the stopping regions for rounds 1, 2 and 3 as:

$$R_1 : a_1 > k_1 = 0.689897948556636 \tag{9}$$

$$R_2 : a_2 > k_2 = 0.5 \tag{10}$$

$$R_3 : a_3 > k_3 = 0 \tag{11}$$

And assuming that if we do not stop at rounds $\{1, \dots, n-1\}$ we will know the maximum only after round n (delay knowledge of losses), $M(j)$ will only be stopped with wins and immediate losses on regions R_j.

The formulas for the CDFs and Ps below were calculated by Mathematica using the Probability command (code on GitHub). For R_1:

$$CDF[M(1)]_{WIN} = \begin{cases} \frac{\left(z^3 - k_1^3\right)}{3} & z > k_1 \\ 0 & otherwise \end{cases} \tag{12}$$

$$CDF[M(1)]_{LOSS} = \begin{cases} \frac{\left(3 \cdot z - z^3 - 3 \cdot k_1 + k_1^3\right)}{3} & z > k_1 \\ 0 & otherwise \end{cases} \tag{13}$$

$$P_1^W = \frac{\left(1 - k_1^3\right)}{3} \tag{14}$$

$$P_1^L = \frac{\left(2 - 3 \cdot k_1 + k_1^3\right)}{3} \tag{15}$$

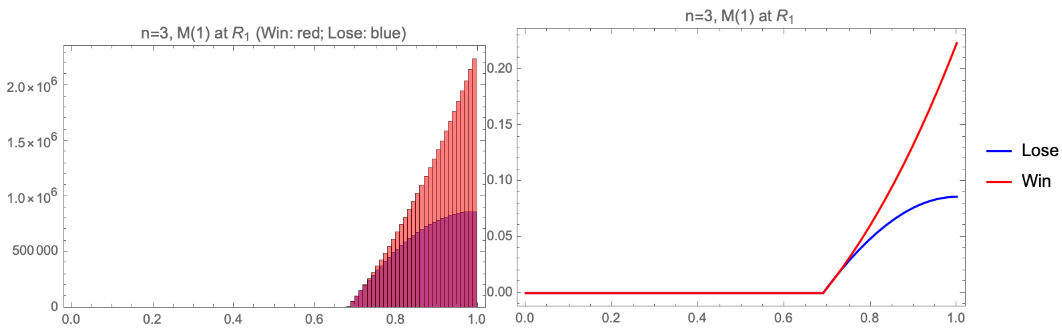

Figure 2: Wins and (immediate) losses at R_1

Which will truncate further realizations of $M(j)$ - we cut the stopped realizations from the further rounds.

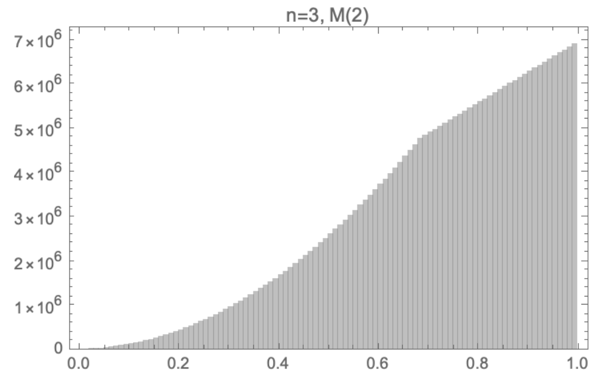

Figure 3: New M(2)

For R_2:

$$CDF\left[M\left(2\right)\right]_{WIN}=\begin{cases}\frac{\left(3\cdot z^2\cdot k_1-2\cdot k_2^3-k_1^3\right)}{6} & z>k_1\\ \frac{\left(z^3-k_1^3\right)}{3} & k_1>z>k_2\\ 0 & otherwise\end{cases} \tag{16}$$

$$CDF\left[M\left(2\right)\right]_{LOSS}=\begin{cases}\frac{\left(6\cdot z\cdot k_1-3\cdot z^2\cdot k_1+2\cdot k_2^3+k_1^3-6\cdot k_1\cdot k_2\right)}{6} & z>k_1\\ \frac{\left(3\cdot z^2-z^3-3\cdot z^2\cdot k_2+k_2^3\right)}{3} & k_1>z>k_2\\ 0 & otherwise\end{cases} \tag{17}$$

$$P_2^W=\frac{\left(3\cdot k_1-k_1^3-2\cdot k_2^3\right)}{6} \tag{18}$$

$$P_2^L=\frac{\left(3\cdot k_1+k_1^3+2\cdot k_2^3-6\cdot k_1\cdot k_2\right)}{6} \tag{19}$$

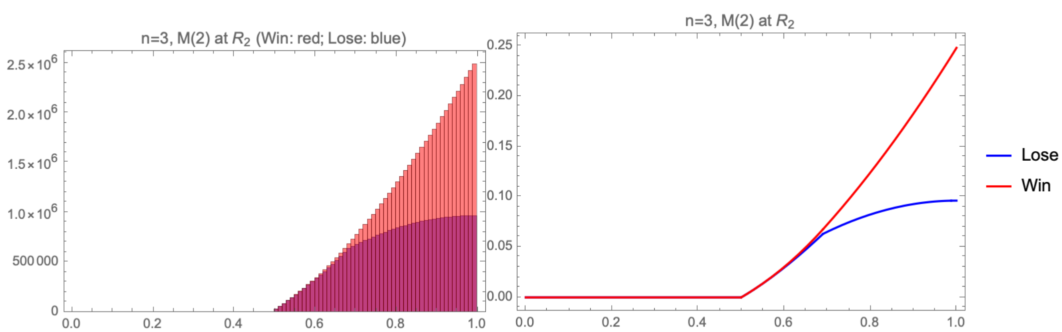

Figure 4: Wins and (immediate) losses at R_2

On to round 3 and the reckoning with the hidden losses:

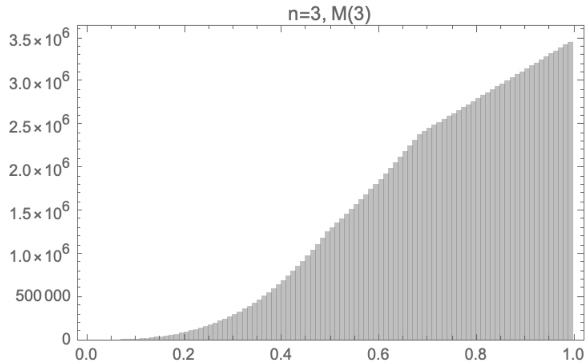

Figure 5: New M(3)

For R_3:

$$CDF\left[M\left(3\right)\right]_{WIN} = \begin{cases} \frac{\left(6\cdot z\cdot k_1\cdot k_2-3\cdot k_1^2\cdot k_2-k_2^3\right)}{6} & z > k_1 \\ \frac{\left(3\cdot z^2\cdot k_2-k_2^3\right)}{6} & k_1 > z > k_2 \\ \frac{z^3}{3} & otherwise \end{cases} \tag{20}$$

$$CDF\left[M\left(3\right)\right]_{LOSS} = \begin{cases} \frac{\left(3\cdot k_1^2\cdot k_2+k_2^3\right)}{6} & z > k_1 \\ \frac{\left(3\cdot z^2\cdot k_2+k_2^3\right)}{6} & k_1 > z > k_2 \\ \frac{2\cdot z^3}{3} & otherwise \end{cases} \tag{21}$$

$$P_3^W = \frac{\left(6\cdot k_1\cdot k_2 - 3\cdot k_1^2\cdot k_2 - k_2^3\right)}{6} \tag{22}$$

$$P_3^L = \frac{\left(3\cdot k_1^2\cdot k_2 + k_2^3\right)}{6} \tag{23}$$

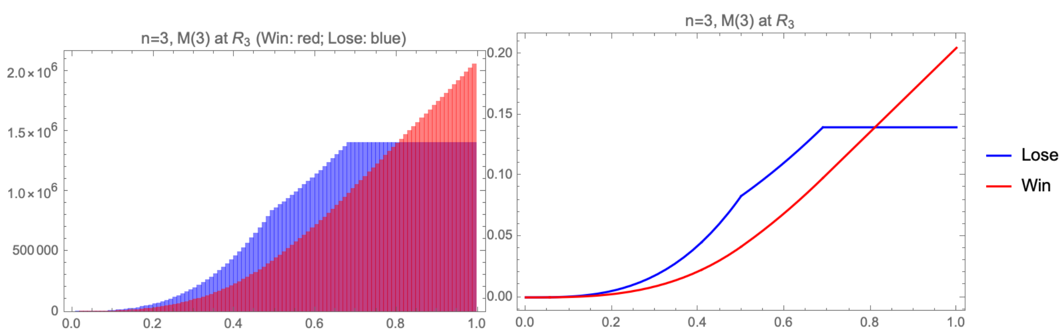

Figure 6: Wins and (previously unrevealed) losses at R_3

So although the attribution of losses and continuations per round is different, the win probabilities are the same as in the original paper. Cross terms will appear on losses starting at round 2 and on wins starting at round 3 for all values of $n \geq 3$.