Paul Nahin's Problem

This is a probability problem that I found in Paul Nahin's book "Dueling Idiots." Suppose that we have a sequence of identical, independently drawn random variables $X_0, X_1, X_2, \ldots$. How many steps can we expect to wait in order to get a value of X that beats the initial "record" set by X_0? That is, if we define the stopping time N as the smallest index for which $X_N > X_0$, what is the expectation value of N?

Better yet: what is the probability distribution of N? It is easy to get the cumulative distribution function:

$$P\left[N > n\right] = P\left[X_0 \text{ is the largest of } \{X_0, X_1, X_2, \ldots X_n\}\right] = \frac{1}{n+1}. \tag{1}$$

and therefore

$$P\left[N = n\right] = P\left[N > n-1\right] - P\left[N > n\right] = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)} \tag{2}$$

You will notice that we did not need to know the distribution of the X_j at all! The expectation value of N follows easily.

$$\langle N \rangle = \sum_{n=1}^{\infty} n \cdot P\left[N = n\right] = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty \tag{3}$$

The mean waiting time diverges! I find this truly shocking. Although the record set by the first value X_0 will always be broken eventually, the mean time until the record is broken is not well defined. Suppose we simulate this process one hundred thousand times and calculate the sample mean of N. Figure 1 shows the results of doing that whole experiment twenty times. Random variables such as N, which have no well-defined mean value, will be subject to large fluctuations of the kind shown in the figure.

The effects of blurring (ties)

The divergence of the expectation value of N stems from the non-negligible probability of waiting very long before the initial record is broken. Perhaps this event will turn out less likely if we make our model slightly more "realistic" in some way. For instance, maybe we should say that two values are practically indistinguishable if they are closer together than some cutoff ϵ that determines the resolution with which we can measure X_i. An easier way to implement the same idea is to make the X_i a discrete random variable. For now, let's suppose that X_i takes the values $\{1, 2, \ldots M\}$ with a uniform distribution. It is intuitively clear that the expected record-breaking time N should be finite. In the particular case $M = 2$ we would just be flipping coins and waiting for the first heads, say.

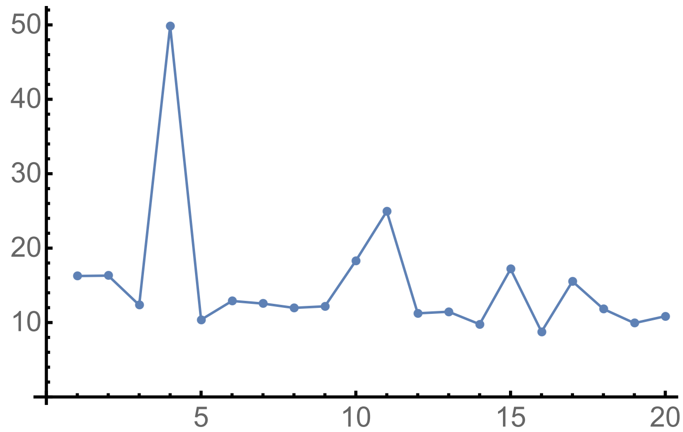

Figure 1: Results from simulating the record-breaking time N one hundred thousand times, and calculating the sample mean. Twenty independent sample means are shown.

Our previous calculation can be repeated with discrete random variables:

$$\begin{aligned} P[N > n] &= \sum_{i=1}^{M} P[N > n \,|\, X_0 = i] \cdot P[X_0 = i] \\ &= \sum_{i=1}^{M} P[X_k \leq i, k = 1, 2, \ldots n] \cdot P[X_0 = i] \\ &= \frac{1}{M} \cdot \sum_{i=1}^{M} \left(\frac{i}{M}\right)^n \end{aligned}$$

Once again we have

$$\begin{aligned} P[N = n] &= P[N > n-1] - P[N > n] \\ &= \frac{1}{M} \cdot \sum_{i=1}^{M} \left(\frac{i}{M}\right)^{n-1} - \frac{1}{M} \cdot \sum_{i=1}^{M} \left(\frac{i}{M}\right)^{n} \\ &= \frac{1}{M} \cdot \sum_{i=1}^{M} \left(\frac{i}{M}\right)^{n-1} \cdot \left(1 - \frac{i}{M}\right) \\ &= \frac{1}{M} \cdot \sum_{i=1}^{M-1} \left(\frac{i}{M}\right)^{n-1} \cdot \left(1 - \frac{i}{M}\right) \qquad (4) \end{aligned}$$

Notice the last line, where we leave out the term for $i = M$ because it vanishes. This will allow us to avoid worrying about dividing by zero in a later step. The expectation value of N follows after switching the order of the sums.

$$\begin{aligned}\langle N \rangle &= \sum_{n=1}^{\infty} n \cdot P\left[N=n\right] \\ &= \frac{1}{M} \cdot \sum_{i=1}^{M-1} \left(1 - \frac{i}{M}\right) \sum_{n=1}^{\infty} n \cdot \left(\frac{i}{M}\right)^{n-1} \end{aligned} \tag{5}$$

Now we use the series

$$\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2} \tag{6}$$

to get the final result:

$$\langle N \rangle = \frac{1}{M} \cdot \sum_{i=1}^{M-1} \left(1 - \frac{i}{M}\right)^{-1} = \sum_{i=1}^{M-1} \frac{1}{M-i} = \sum_{i=1}^{M-1} \frac{1}{i} \tag{7}$$

There is a nice similarity with the (diverging) expression for $\langle N \rangle$ in the case of continuous random variables. Here we see that the number of distinguishable states, M, acts as a cutoff on the harmonic sum. This result also indicates that $\langle N \rangle$ grows logarithmically with M.

Looking back, I think I could have done this more simply using the conditional expectation values.