Expected Value Road Trip

As a result of a questionable decision to combine a cross-country move with a four-week stay at a summer math program, I found myself driving solo from San Francisco to Salt Lake City in a single day during the summer of 2007. To combat the lethargic mid-afternoon stretch I decided to mentally reconstruct the solution to a classic problem which dates back at least to Laplace [N]; namely, to determine how many numbers, chosen uniformly at random from the interval $[0,1]$, one must select before their sum exceeds 1. The delightful answer, as we all know, is e.

This question has been the subject of renewed interest lately. In [N], Ćurgus and Jewett examine the function $\alpha(t)$, defined as the expected number of random selections from $[0,1]$ that must be made before the cumulative total exceeds t. They go on to find an explicit formula for $\alpha(t)$ using the theory of delay functions. It stands to reason that $\alpha(t) \approx 2t$, since each selection is equal to $\frac{1}{2}$, on average. Their surprising observation is that

$$\lim_{t\to\infty}(\alpha(t) - 2t) = \frac{2}{3}.$$

One can independently show [N] that the value of $\alpha(t)$ may be expressed in terms of powers of e for integral values of t. We are thus led to statements such as

$$e^6 - 5e^5 + 8e^4 - \tfrac{9}{2}e^3 + \tfrac{2}{3}e^2 - \tfrac{1}{120}e \approx 12\tfrac{2}{3},$$

by considering $\alpha(6)$, for example. The two quantities agree to six digits past the decimal point.

Our purpose here is to pursue a thought which struck as yet another low mountain range passed by to the south somewhere in Nevada. We wish to know how many numbers, chosen uniformly at random from the interval $[1,e]$, one must select before their product exceeds e. To cut the suspense, the answer is the somewhat unlikely expression

$$e^{\frac{1}{e-1}} + \frac{e-1}{e}.$$

Readers may test their intuition by predicting whether this quantity is greater than or less than e.

The assault proceeds in predictable fashion: for $n \geq 1$ let q_n be the probability that a product of n numbers chosen from $[1,e]$ is *not* greater than e. (It will be convenient to define $q_0 = 1$ as well.) Then the probability that the product exceeds e for the first time at the nth selection is

$$(1-q_n) - (1-q_{n-1}) = q_{n-1} - q_n,$$

valid for $n \geq 1$. The expected value we seek is

$$\begin{aligned} &\sum_{n=1}^{\infty} n(q_{n-1} - q_n) \\ = \; &\sum_{n=1}^{\infty} q_{n-1} + \sum_{n=1}^{\infty}(n-1)q_{n-1} - \sum_{n=1}^{\infty} nq_n \\ = \; &\sum_{n=0}^{\infty} q_n. \end{aligned}$$

The fact that all sums converge absolutely will be clear once we obtain a closed form expression for q_n. The ticklish part is finding the formula.

The region R_n within the n-cube $[1,e]^n$ consisting of points $(x_1, \ldots, x_n)$, the product of whose coordinates is at most e, is described by

$$\begin{aligned} 1 \leq \; &x_1 \leq e \\ 1 \leq \; &x_2 \leq \frac{e}{x_1} \\ &\vdots \\ 1 \leq \; &x_n \leq \frac{e}{x_1 \cdots x_{n-1}}. \end{aligned}$$

We may then compute q_n via

$$q_n = \frac{1}{(e-1)^n}\int_{R_n} dx_n \cdots dx_1.$$

We focus our attention on the integral by defining $\theta_n = \int_{R_n} dx_n \cdots dx_1$. The first six values are $e-1$, 1, $\frac{1}{2}e - 1$, $-\frac{1}{3}e + 1$, $\frac{3}{8}e - 1$, and $-\frac{11}{30}e + 1$, found by electronic means. Evaluating these integrals directly becomes an increasingly arduous affair as n grows, so it comes as a pleasant surprise to discover that their value may be ascertained with hardly any effort. Indeed, we find that

$$\theta_{n+1} = \int_{R_n}\left(\frac{e}{x_1 \cdots x_n} - 1\right) dx_n \cdots dx_1,$$

by performing the easy integration with respect to x_{n+1}. It follows that

$$\theta_{n+1} + \theta_n = \int_{R_n} \frac{e}{x_1 \cdots x_n}\, dx_n \cdots dx_1.$$

Making the change of variables $y_k = \ln x_k$ so that $dy_k = \frac{1}{x_k}dx_k$ we find that

$$\theta_{n+1} + \theta_n = \int_{R_n'} e\, dy_1 \cdots dy_n.$$

It is clear that the region R'_n consists of those points $(y_1, \ldots, y_n)$ satisfying $y_k \geq 0$ and $y_1 + \cdots + y_n \leq 1$, i.e. R'_n is a right unit n-simplex. It is well-known that the volume of this region is $\frac{1}{n!}$, so we have shown that

$$\theta_{n+1} + \theta_n = \frac{e}{n!}.$$

It is conceivable that one might stumble upon this approach while navigating northern Nevada by car. In fact, I initially headed down another road by asking how many numbers from the interval $[1, 2]$ would be required to obtain a product that exceeded 2. Several mental integrations later it became clear that I had taken a wrong turn when the accumulation of $\ln 2$'s became overwhelming. By the time the question had been properly formulated the Great Salt Flats beckoned, and the interesting task of answering it was postponed until a later time.

Resuming the argument, we now find an expression for θ_n. For $n \geq 1$ let b_n be the nth partial sum of the usual series for e^{-1}, so that

$$b_n = 1 - \frac{1}{1} + \frac{1}{2} - \cdots + (-1)^{n-1}\frac{1}{(n-1)!}.$$

We claim that $\theta_n = (-1)^n(1 - b_n e)$. The quantities agree for $n = 1$, and for $n \geq 2$ we compute

$$\begin{aligned}
\theta_n + \theta_{n+1} &= (-1)^n(1 - b_n e) + (-1)^{n+1}(1 - b_{n+1} e) \\
&= e(-1)^n(b_{n+1} - b_n) \\
&= e(-1)^n\left(\frac{(-1)^n}{n!}\right) \\
&= \frac{e}{n!},
\end{aligned}$$

as desired. This provides the sought after expression for q_n, namely

$$q_n = \frac{(-1)^n(1 - b_n e)}{(e-1)^n}$$

for $n \geq 1$. Clearly $|q_n| \leq (e-1)^{-n}$, which justifies the manipulations of the infinite series above.

The final step of summing the q_n presents a very satisfying exercise involving geometric and exponential series. To begin,

$$\begin{aligned}
\sum_{n=0}^{\infty} q_n &= 1 + \sum_{n=1}^{\infty}\frac{(-1)^n}{(e-1)^n} - \sum_{n=1}^{\infty}\frac{(-1)^n(b_n e)}{(e-1)^n} \\
&= \frac{e-1}{e} - \sum_{n=1}^{\infty}\frac{(-1)^n(b_n e)}{(e-1)^n}.
\end{aligned}$$

We evaluate the remaining term by writing b_n as a sum and interchanging the order of summation, obtaining

$$\begin{aligned}
-\sum_{n=1}^{\infty}\frac{(-1)^n(b_n e)}{(e-1)^n} &= -e\sum_{n=1}^{\infty}\frac{(-1)^n}{(e-1)^n}\sum_{k=0}^{n-1}\frac{(-1)^k}{k!} \\
&= -e\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\sum_{n=k+1}^{\infty}\frac{(-1)^n}{(e-1)^n} \\
&= -e\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\cdot\frac{(-1)^{k+1}}{e(e-1)^k} \\
&= \sum_{k=0}^{\infty}\frac{1}{k!}\cdot\frac{1}{(e-1)^k} \\
&= e^{\frac{1}{e-1}}.
\end{aligned}$$

This completes the computation.

It is natural to revisit the phenomenon described by Ćurgus and Jewett in this multiplicative context. So let $\beta(t)$ be the expected number of random selections that must be made from the interval $[1, e]$ before the product exceeds e^t. Since the average value of $\ln x$ on the interval $[1, e]$ is $\frac{1}{e-1}$, we anticipate that

$$\beta(t) \approx e^{\frac{t}{e-1}}.$$

But can a more precise statement be made regarding the relationship between these two quantities, similar to the observation that $(\alpha(t) - 2t) \to \frac{2}{3}$? So far we have established that

$$\beta(1) - e^{\frac{1}{e-1}} = \frac{e-1}{e}.$$

Incidentally, since $\ln x$ is concave down, a number x chosen uniformly from $[1, e]$ will yield a value for $\ln x$ that is closer to 1, on average, than we would have obtained by simply having chosen a number uniformly from $[0, 1]$. In other words, a random selection from $[1, e]$ contributes "more" to a product than a selection from $[0, 1]$ will contribute to a sum. So we expect that fewer selections are required for our product to exceed a given value; i.e. we predict that $\beta(t) < \alpha(t)$ for all positive t. As anticipated,

$$\beta(1) \approx 2.42169 < e,$$

settling the matter raised earlier. And this remark will serve nicely as our Salt Lake City, concluding the journey, or at least this leg of the trip.