Stat 134 Fall 2011: The expectation of the maximum of exponentials

Michael Lugo

October 28, 2011

Let $X_1, \ldots, X_n$ be independent random variables, each exponential with rate 1. We found in class that the CDF of their maximum, $X = max(X_1, \ldots, X_n)$, was

$$\begin{aligned} F_X(x) &= P(max(X_1, \ldots, X_n) < x) \\ &= P(X_1 < x, X_2 < x, \ldots, X_n < x) \\ &= P(X_1 < x)P(X_2 < x) \cdots P(X_n < x) \\ &= (1 - e^{-x})^n \end{aligned}$$

We also derived the formula $E(X) = \int_0^\infty 1 - F_X(x)\,dx$. We can apply this here to get

$$E(X) = \int_0^\infty 1 - \left(1 - e^{-x}\right)^n \,dx$$

and we can use this formula directly with small values of n. For example we have if $n = 2$

$$E(X) = \int_0^\infty 1 - (1 - e^{-x})^2 \,dx = \int_0^\infty 2e^{-x} - e^{-2x} \,dx = 2 - \frac{1}{2} = \frac{3}{2}$$

and if $n = 3$ we have

$$E(X) = \int_0^\infty 1 - (1 - e^{-x})^3 \,dx = \int_0^\infty 3e^{-x} - 3e^{-2x} + e^{-3x} \,dx = 3 - \frac{3}{2} + \frac{1}{3} = \frac{11}{6}.$$

If we try this for various small values of n, we get values which agree with the formula $E(X) = 1 + 1/2 + \cdots + 1/n$. For example $1 + 1/2 = 3/2$ and $1 + 1/2 + 1/3 = 11/6$. Furthermore, we can argue that this formula holds as follows. Consider n Poisson point processes of rate 1, each on the space $[0, \infty)$, where the points in each process are of a different color. Then X_i is the position of the leftmost point of color i, and X is the position of the rightmost of these minima. But we can also decompose X as a sum $X = W_1 + W_2 + \cdots + W_n$. Let $T_k = W_1 + \cdots + W_k$.

Here W_1 is the distance from 0 to the first point of any color – call this color c_1. Then W_2 is the distance from T_1 to the first point of a color other than c_1; call this color c_2, and note that this point is at $T_1 + W_2 = T_2$. Next W_3 is the distance from T_2 to the first point of a color other than c_1 or c_2; call this color c_3, and note that this point is at $T_2 + W_3 = T_3$. In general W_k is the distance from T_{k-1} to the first point of a color other than $c_1, \ldots, c_{k-1}$, which we call c_k.

Now imagine finding W_k. This is the distance from T_{k-1} to the first point beyond it which is of a color other than $c_1, \ldots, c_{k-1}$. There are $n - (k-1)$ such colors, so those points form a Poisson process of rate $n - (k-1)$. The location of the first point, W_k, is therefore exponential with rate $n - (k-1)$, and so $E(W_k) = 1/(n - (k-1))$. Then finally

$$\begin{aligned} E(X) &= E(W_1) + E(W_2) + \cdots + E(W_n) \\ &= \frac{1}{n-(1-1)} + \frac{1}{n-(2-1)} + \cdots + \frac{1}{n-(n-1)} \\ &= \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1}. \end{aligned}$$

This is the formula we wanted; furthermore since the W_k are independent we actually know the distribution of X.

But perhaps we want to prove that $E(X) = 1 + 1/2 + \cdots + 1/n$ directly from the PDF, without the detour via Poisson processes. (In fact this is a reasonable goal, as making the arguments about Poisson processes given here rigorous takes some work.) We can use the binomial theorem to get

$$E(X) = \int_0^\infty 1 - (1 - e^{-x})^n \, dx = \int_0^\infty \sum_{k=1}^{n} (-1)^{k+!} e^{kx} \, dx$$

and pull the integral inside the sum to get

$$E(X) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \int_0^\infty e^{-kx} \, dx.$$

Finally, doing the integral gives

$$E(X) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{1}{k}.$$

This is a useful formula if we want to compute $E(X)$ for any particular n, but it's not at all obvious that $E(X) = 1 + 1/2 + \cdots + 1/n$. To prove this we need to use some generating-function trickery.

First, we construct a generating function in two variables:

$$F(z, u) = \sum_{n=0}^{\infty} \left(\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} u^k \right) z^n.$$

The inner sum can be rewritten to give

$$F(z,u) = \sum_{n=0}^{\infty} \left(1 - \sum_{k=0}^{n} (-1)^k \binom{n}{k} u^k \right) z^n$$

and the inner sum can now be simplified using the binomial theorem to get

$$F(z,u) = \sum_{n=0}^{\infty} (1 - (1-u)^n) z^n.$$

Now we can write this as a difference of geometric series,

$$F(z,u) = \sum_{n=0}^{\infty} z^n - \sum_{n=0}^{\infty} (z(1-u))^n$$

and summing the series gives

$$F(z,u) = \frac{1}{1-z} - \frac{1}{1-z(1-u)} = \frac{zu}{(1-z)(1-z+zu)}.$$

But of course $E(X)$ has $1/k$ in it, which will cause some trouble. We saw in class that differentiating generating functions brings down powers of k; for example if some random variable X has the generating function $f(z) = \sum_{k=0}^{\infty} P(X=k) z^k$, then we have $f'(z) = \sum_{k=0}^{\infty} P(X=k) k z^{k-1}$ and so $E(X) = f'(1)$. Similarly, to deal with factors of $1/k$ in a summand we need to integrate.

In particular, note that $\int_0^u t^k/t \, dt = u^k/k$. We can use this to get

$$\begin{aligned}
\int_0^u \frac{F(z,t)}{t} \, dt &= \int_0^u \frac{\sum_{n=0}^{\infty} \left(\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} t^k \right) z^n}{t} \, dt \\
&= \int_0^u \sum_{n=0}^{\infty} \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} t^{k-1} z^n \, dt \\
&= \sum_{n=0}^{\infty} \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \left(\int_0^u t^{k-1} \, dt \right) z^n \\
&= \sum_{n=0}^{\infty} \left(\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{u^k}{k} \right) z^n
\end{aligned}$$

If we let $u = 1$, then, we get

$$\int_0^1 \frac{F(z,t)}{t} \, dt = \sum_{n=0}^{\infty} \left(\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{1}{k} \right) z^n$$

and so the coefficient of z^n on the right-hand side is exactly the sum we want.

Now we actually do the integral:

$$
\begin{aligned}
\int_0^1 \frac{F(z,t)}{t}\,dt &= \int_0^1 \frac{z}{(1-z)(1-z+zt)}\,dt \\
&= \frac{z}{1-z}\int_0^1 \frac{1}{1-z+zt}\,dt \\
&= \frac{z}{1-z}\,\frac{\log(1-z+zt)}{z}\bigg|_{t=0}^{1} \\
&= \frac{1}{1-z}\log(1-z+zt)_{t=0}^{1} \\
&= \frac{1}{1-z}(\log(1-z+z)-\log(1-z)) \\
&= \frac{-\log(1-z)}{1-z}.
\end{aligned}
$$

So the sum we are looking for is just the coefficient of z^n in the Taylor series of $-\log(1-z)/(1-z)$. We have

$$\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \cdots$$

and if we replace z with $-z$ and add a minus sign we get

$$-\log(1-z) = z + \frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \cdots .$$

Furthermore if $A(z) = a_0 + a_1 z + a_2 z^2 + \cdots$, then we have

$$
\begin{aligned}
\frac{A(z)}{1-z} &= A(z)(1 + z + z^2 + z^3 + \cdots) \\
&= (a_0 + a_1 z + a_2 z + a_3 z^3 + \cdots) + (a_0 z + a_1 z^2 + a_2 z^3 + \cdots) \\
&\quad + (a_0 z^2 + a_1 z^3 + \cdots) + (a_0 z^3 + \cdots) + \cdots
\end{aligned}
$$

If we combine like terms then we get

$$\frac{A(z)}{1-z} = a_0 + (a_0 + a_1)z + (a_0 + a_1 + a_2)z^2 + (a_0 + a_1 + a_2 + a_3)z^3 + \cdots$$

and so the coefficients of $A(z)/(1-z)$ are just the partial sums of the coefficients of $A(z)$. In particular, the coefficients of $-\log(1-z)/(1-z)$ are the partial sums of the coefficients of $-\log(1-z)$. That is, the z^n coefficient of $-\log(1-z)/(1-z)$ is exactly $1+1/2+1/3+\cdots+1/n$, which is what we wanted to show.