# Project suggestion
# Expectations of Probability

John McCuan

September 27, 2023

In my discussion of Problems 1-3 of Assignment 2 there was something I found interesting, but I didn't mention explicitly. In Problem 2 part (b) there is a question of the form "How many trials would you expect to execute in order to see some particular outcome at least once?" I suggested furthermore that one framework in which to try to answer such a question is using a probability measure space $\mathbb{N}$ modeling the number of trials.

From this point, my son suggested an answer to such a question might be "the smallest number of trials so that the cumulative probability of that particular outcome is at least 1/2." Another answer might be "the expectation for the number of trials corresponding to seeing that particular outcome." I found it striking that these two values were different. You may recall that the first came out to be 9, and the second gave an answer of about 13.

The question of what is going on here can be made precise: Let $\alpha : \wp(S) \to [0, 1]$ be an adolescent probability measure on a set $S$ where $S = \mathbb{N}$ or $S = \mathbb{N}_0$. As phrased above we might restrict attention to $S = \mathbb{N}$, but I'll explain below why one might wish to consider a more general situation. In this framework, the two statistical values

$$E_f = \min\left\{k \in \mathbb{N} : F(k) \geq \frac{1}{2}\right\}$$

where $F : \mathbb{R} \to [0, 1]$ is the CMF associated with $\alpha$ and

$$E = \int_S \mathrm{id}_{\mathbb{R}}$$

where the integral is with respect to $\alpha$ are well-defined. The second is called the **expectation**, and gives the average value of numbers $\xi \in S$ with respect to the measure $\alpha$.

The basic project I'm suggesting is to try to understand the relation between these two statistical values. Under what conditions are they the same? Under what conditions can you say one is smaller than the other?

Gaining a comprehensive understanding of the answers to these questions may be a project that is a bit too ambitious for most of you. But at the very least, any one of you should be able to see what happens in some special cases in order to get some idea of what is going on. In particular, it should not be difficult to find situations in which $E_f = E$. Also, example situations in which $E_f \neq E$ are not to difficult to find and explain.

It would be nice to discuss the meaning(s) of $E_f$ and $E$ and explain why they might be expected to be the same and/or why they should be expected to be different.

I think one should consider the geometric distribution for different values of $p$. Under what conditions can the difference between $E_f$ and $E$ be larger than 1? If I've computed correctly, the difference can be as large as $13 - 9 = 4$. Is $E$ always more "conservative," i.e., larger, than $E_f$?

Thinking about this a little bit, it occurred to me that this is also a question that makes sense for the binomial induced measure. In that case, one naturally starts with $k = 0$, and this is why I suggested including $S = \mathbb{N}_0$. Of course, one could also translate the binomial distribution to the right by one and just use $\{1, 2, \dots, n+1\} \subset \mathbb{N}$, but I think it is better to see the result using the natural domain. In any case, what happens in this situation is quite interesting.

For the binomial and geometric distributions, I was able to animate an image showing the graph of the CMF (at integer values) with a horizontal line at height $p = 1/2$ superimposed to see the value of $E_f$ easily and a vertical line at $\xi = E$. The animation was executed on the base probability value.

Finally, the discussion might be extended to various "continuous" distributions, that is to say integral measures where $E_f$ is defined by

$$\int_{-\infty}^{E_f} \delta(\omega)\, d\omega = \frac{1}{2}$$

and

$$E = \mu = \int_{-\infty}^{\infty} \omega\delta(\omega)\, d\omega.$$

In a certain sense, more or less the same questions can be asked in a different way: What can you say about

$$\alpha(\{\xi \in \mathbb{R} : \xi \leq E\})?$$