Musings on the Exponential Distribution

2024-04-01

This article presents a rigorous mathematical proof demonstrating why exponential distributions naturally emerge in memoryless physical processes. We show that the exponential distribution is uniquely characterized by its memoryless property through two distinct approaches. The first proof utilizes the survival function and its homomorphism properties, extending from natural numbers to rational and real domains. The second, more physics-oriented derivation employs differential equations to arrive at the same conclusion. We conclude by connecting these results to the hazard function formalism, providing an intuitive framework for understanding why exponential distributions are fundamental to describing microscopic systems such as radioactive decay.

blog: https://tetraquark.vercel.app/posts/expdist/

email: quarktetra@gmail.com

Many processes in physics are defined by exponential distribution. Have you ever wondered why that is the case? That is because it is the only one available for microscopic systems. A microscopic system, such as a radioactive atom has no internal memory. And the only continuous distribution with no memory is the exponential distribution. Let' prove this. Consider a random variable X defined by its probability function $F_X(x) = P(X \leq x)$, which is read as probability of X being less than a given number x. Typically X is associated with some kind of event or failure, and that is why a counter part of F_X is defined as the survival function S_X Define the survival function $S(t)$:

$$S_X(x) = 1 - F_X(x) = 1 - P(X \leq x) = P(X > x). \tag{1}$$

What is the conditional probability of $X > x + s$ given that it survived until x?

$$P(X > x + s | X > x) = \frac{P(X > x + s)}{P(X > x)} = \frac{S_X(x + s)}{S_X(x)}. \tag{2}$$

The critical observation is this: if the process has no memory, the conditional probability can only depend on the difference in observation points. Think about it this way: the probability of an atom to decay before $t = 1$ year given that it is intact at $t = 0$ is the same as the

probability of it decaying within the $t = 11^{th}$ year given that it is intact at $t = 10$ year. Atoms don't remember how old they are. Therefore we require that the right hand side of Eq. 2 has no x dependence, that is:

$$\frac{S_X(x+s)}{S_X(x)} = S_X(s) \implies S_X(x+s) = S_X(x)S_X(s), \tag{3}$$

which is begging for the exponential function due to its homomorphism property mapping multiplication to addition. To show this explicitly, consider the repeated application of $S(x)$ p times

$$[S_X(x)]^p = \underbrace{S_X(x)S_X(x) \cdots S_X(x)}_{p \text{ times}} = S_X(px), \tag{4}$$

where p is a natural number. This certainly holds for natural numbers, but we can do better. Now consider another counting number q and apply $S(x/q)$ q times

$$\left[S_X\left(\frac{x}{q}\right)\right]^q = \underbrace{S_X\left(\frac{x}{q}\right)S_X\left(\frac{x}{q}\right)\cdots S_X\left(\frac{x}{q}\right)}_{q \text{ times}} = S_X\left(q\frac{x}{q}\right) = S_X(x) \implies S_X\left(\frac{x}{q}\right) = [S_X(x)]^{\frac{1}{q}}. \tag{5}$$

Since we can make this work for p and $1/q$, it also works for p/q:

$$S_X\left(\frac{p}{q}x\right) = S_X\left(p\frac{x}{q}\right) = \left[S_X\left(\frac{x}{q}\right)\right]^p = \left[(S_X(x))^{\frac{1}{q}}\right]^p = [S_X(x)]^{\frac{p}{q}}. \tag{6}$$

That is

$$S_X(ax) = [S_X(x)]^a, \tag{7}$$

where $a = \frac{p}{q}$ is a rational number. What about irrational numbers? Since the rationals are a dense subset of the real numbers, for every real number we can find rational numbers arbitrarily close to it[1]. The continuity of S ensure that we can upgrade a from a rational number to a real number. One last thing to do is to show that we are getting our old friend e out of this. Setting $x = 1$ in Eq. 7 gives

$$S_X(a) = [S_X(1)]^a = (\exp\{-\ln[S_X(1)]\})^a \equiv e^{-\lambda a}, \tag{8}$$

where $\lambda = -\ln[S_X(1)] > 0$.

This derivation is very satisfying, particularly for a non-mathematician like me. But it is a bit too rigorous. We can do a physicist version of this derivation. We start from Eq. 3, take the log and the derivative of both sides with respect to s:

$$\frac{d}{ds}\log\left(S_X(x+s)\right) = \frac{S_X'(x+s)}{S_X(x+s)}$$

$$\frac{d}{ds}\log\left(S_X(x)S_X(s)\right) = \frac{d}{ds}\left[\log\left(S_X(x)\right)+\log\left(S_X(s)\right)\right] = \frac{S_X'(s)}{S_X(s)}. \tag{9}$$

They are equal to each other:

$$\frac{S_X'(x+s)}{S_X(x+s)} = \frac{S_X'(s)}{S_X(s)} \equiv -\lambda, \tag{10}$$

where we realized that the terms are functions of s or $x+s$, and they cannot possibly be equal to each other unless they are equal to a constant. Integrating, we get:

$$\frac{S_X'(s)}{S_X(s)} = \frac{d}{ds}\left[\log\left(S_X(s)\right)\right] = -\lambda \implies S_X(s) = e^{-\lambda s}, \tag{11}$$

which completes the proof.

One of the difficulties with such proofs is that, they are done backwards. In fact, life would have been much easier if we started from the so called hazard function. I talked about this in one of my earlier posts: Musings on Weibull distribution. If we start from a hazard function (ratio of failures to the survivors) h and define everything else on top, this is how things go:

$$h(t) \equiv \frac{f(t)}{1-F(t)} = \frac{\frac{d}{dt}F(t)}{1-F(t)} = -\frac{d}{dt}\left[ln\left(1-F(t)\right)\right] \implies F(t) = 1-e^{-\int_0^t d\tau h(\tau)}. \tag{12}$$

If you go with the simplest assumption (memory-less, constant) $h = \lambda$, you get the exponential distribution.

[1]   N. L. Biggs, *Discrete mathematics.* OUP Oxford, 2002 [Online]. Available: https://books.google.com/books?id=vrSdQgAACAAJ