For the inaugural post of this blog, I'd like to talk about the exponential distribution, i.e., a continuous random variable X with probability density function (henceforth, pdf),

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0 \tag{1}$$

I use e here, but will use exp if the formulae are too cluttered. Note that some authors write $x > 0$, but since X is a continuous random variable, this distinction is innocuous. The cumulative distribution function (henceforth, cdf) or simply, distribution function is,

$$F(y) = P(X \leq y) = \int_0^y \mathrm{d}s\, \lambda e^{-\lambda s} = 1 - e^{-\lambda y} \tag{2}$$

Note that $F = 0$ for $y \leq 0$ and $F(\infty) = 1$ as it should be. Also, notation, such as, $X \sim F$ in literature should be read as, X is distributed as F (i.e., X has distribution function F).

As is obvious, this is mathematically a fairly simple distribution; however, a very useful one to know. For example, it is used to model the time of first occurrence of a certain event, say, the failure of a device or an electronic component. In this sense, it is a continuous analog of the (discrete) geometric distribution. The exponential distribution also shows up in discussions of hazard rate. Therefore, λ is often referred to as the rate parameter.

The exponential distribution has the following property, known as memoryless, which is formally written as follows

$$P(X > y + z \mid X > y) = P(X > z) \tag{3}$$

where the probability on the LHS (left hand side) denotes a conditional probability, i.e., probability of A given B, $P(A \mid B)$. In words, if a component has not failed for y units of time, the probability of not failing for an additional z units of time does not carry the "memory" of not having failed for y. It is easy to check this for an exponential distribution.

First, note that

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{4}$$

The event $X > y + z \subset$ of the event $X > y$ and so using the above, equation (3) reduces to

$$\frac{P(X > y + z)}{P(X > y)} = P(X > z) \tag{5}$$

Recall that $P(X > y) = 1 - F(y)$. So we have

$$\frac{P(X > y + z)}{P(X > y)} = \frac{e^{-\lambda(y+z)}}{e^{-\lambda y}} = e^{-\lambda z} = P(X > z) \tag{6}$$

Trying the same thing with a Uniform distribution, for e.g., would not work. Here $F(x) = x$ assuming $X \in (0, 1)$. We have,

$$\frac{P(X > y + z)}{P(X > y)} = \frac{1 - y - z}{1 - y} \neq P(X > Z) = 1 - z \tag{7}$$

assuming all expressions involving y and z are in $(0, 1)$. For a continuous random variable, it can be shown that the exponential distribution is the only distribution with this memoryless property. Partly because of this, the exponential distribution also shows up in the theory of continuous time Markov chains (CTMCs), modelling the transition time between discrete states of the chain.

Calculation of the **mean**, **variance**, can be done as follows. The mean (or expectation, E) is,

$$\mathrm{E}(X) = \int_0^\infty \mathrm{d}s\, s\, \lambda e^{-\lambda s} = \frac{1}{\lambda} \tag{8}$$

Writing $\beta = 1/\lambda$, one also finds the pdf of the exponential distribution written as,

$$f(x) = \frac{1}{\beta} \exp(-x/\beta) \tag{9}$$

Written this way, one can easily recognize this distribution as an $\alpha = 1$ special case of a gamma distribution parameterized as $\mathrm{Gamma}(\alpha, \beta)$. The definition of a variance is

$$\mathrm{var}(X) = \mathrm{E}[(X - \beta)^2] \tag{10}$$

With some elementary algebra, this reduces to

$$\mathrm{var}(X) = \mathrm{E}(X^2) - \beta^2 \tag{11}$$

Texts on probability and statistics usually present the above formula as,

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \tag{12}$$

which reduces to the form in equation (11), as $\mathrm{E}(X) = \beta$.

One can of course calculate the second order moment $\mathrm{E}(X^2)$ using

$$\int_0^\infty \mathrm{d}s\, s^2\, \lambda e^{-\lambda s} \tag{13}$$

Such integrals can be easily evaluated using integration by parts or by looking into tables of integration. If neither of those methods sound appealing, one can also resort to a software that will do the calculation. However, here I want to talk about the moment generating function (henceforth mgf),

$$M(t) = \mathrm{E}(\exp(tX)) = \int_0^\infty \mathrm{d}s e^{ts} \lambda e^{-\lambda s} = \int_0^\infty \mathrm{d}s\, \lambda e^{(t-\lambda)s} = \frac{1}{\lambda - t} \tag{14}$$

It is important to note that this easy to do integral only exists for $t < \lambda$. To see why, note that the integral has an upper limit of infinity and will only converge if the integrand is a decaying exponential, which happens if $t - \lambda < 0$.

From the theory of mgf's, we know that n^{th} order moments

$$\mathrm{E}(X^n) \tag{15}$$

can be obtained by derivatives of $M(t)$. So we have $M(0) = 1$ as it should be (since the pdf is normalized); $M'(0) = 1/\lambda$, giving the mean; and,

$$\mathrm{E}(X^2) = M^{''}(0) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\lambda}{\lambda - t}\right)|_{t=0} \tag{16}$$