1 Review of the exponential distribution

The exponential probability distribution has many nice/special properties; we review them here. They are fundamental in the use of the famous Poisson point process, and more generally in continuous-time Markov chains; both topics that we will study soon.

1.1 Basic definition

A r.v. X has an exponential distribution at rate $\lambda > 0$, denoted by $X \sim exp(\lambda)$, if X is non-negative with c.d.f. $F(x) = P(X \leq x),\ x \geq 0$, tail $\overline{F}(x) = P(X > x) = 1 - F(x)$ and density $f(x) = F'(x)$ given by

$$\begin{aligned} F(x) &= 1 - e^{-\lambda x},\ x \geq 0, \\ \overline{F}(x) &= e^{-\lambda x},\ x \geq 0, \\ f(x) &= \lambda e^{-\lambda x},\ x \geq 0. \end{aligned}$$

It is easily seen that

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ E(X^2) &= \frac{2}{\lambda^2} \\ Var(X) &= \frac{1}{\lambda^2}. \end{aligned}$$

For example,

$$\begin{aligned} E(X) &= \int_0^\infty x f(x) dx \\ &= \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}. \text{ Alternatively:} \\ E(X) &= \int_0^\infty \overline{F}(x) dx \ \text{ (integrating the tail method)} \\ &= \int_0^\infty e^{-\lambda x} dx \\ &= \frac{1}{\lambda}. \end{aligned}$$

If the rate is λ, then the mean is $1/\lambda$.

Similarly,

$$\begin{aligned}
E(X^2) &= \int_0^\infty x^2 f(x)dx \\
&= \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}. \quad \text{Alternatively:} \\
E(X^2) &= \int_0^\infty 2x\overline{F}(x)dx \quad \text{(integrating the tail method)} \\
&= \int_0^\infty e^{-\lambda x} dx \\
&= \frac{2}{\lambda^2}.
\end{aligned}$$

$$\text{Hence } Var(X) = E(X^2) - E^2(X) = 1/\lambda^2.$$

The most important property of the exponential distribution is the *memoryless property*,

$$P(X - y > x | X > y) = P(X > x), \text{ for all } x \geq 0 \text{ and } y \geq 0,$$

which can also be written as

$$P(X > x + y) = P(X > x)P(X > y), \text{ for all } x \geq 0 \text{ and } y \geq 0.$$

The memoryless property asserts that the residual (remaining) lifetime of X given that its age is at least y, namely $X - y$, has the same distribution as X originally did, and is independent of its age: X forgets its age or past and starts all over again. If X denotes the lifetime of a light bulb, then this property implies that if you find this bulb burning sometime in the future, then its remaining lifetime is the same as a new bulb and is independent of its age. So you could take the bulb and sell it as if it were brand new. Even if you knew, for example, that the bulb had already burned for 3 years, this would be so. We say that X (or its distribution) is memoryless.

The fact that if $X \sim exp(\lambda)$, then it is memoryless is immediate from the basic definition of conditional probability, $P(A \mid B) = P(AB)/P(B)$, where below, $A = \{X > x + y\}$ and $B = \{X > y\}$, where we also observe that $P(X - y > x, X > y) = P(X > x + y, X > x) = P(X > x + y)$:

$$\begin{aligned}
P(X - y > x | X > y) &= \frac{P(X > x + y)}{P(X > y)} \\
&= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} \\
&= e^{-\lambda x} \\
&= P(X > x).
\end{aligned}$$

Note that equivalently we can write that as

$$P(X > x + y) = P(X > x)P(X > y), \ x \geq 0, y \geq 0.$$

The converse is also true:

Proposition 1.1 *A non-negative r.v. X (which is not identically 0) has the memoryless property if and only if it has an exponential distribution.*

Proof : One direction was proved above already, so we need only prove the other. Letting $g(x) = P(X > x)$, we have $g(x+y) = g(x)g(y),\ x \geq 0,\ y \geq 0$, and we proceed to show that such a function must be of the form $g(x) = e^{-\lambda x}$ for some λ. To this end, observe that by using $x = y = 1$ it follows that $g(2) = g(1)g(1)$ and more generally $g(n) = g(1)^n,\ n \geq 1$. Noting that

$$1 = \frac{1}{m} + \frac{1}{m} + \cdots + \frac{1}{m} \quad (m \text{ summands}),$$

we see that $g(1) = g(1/m)^m$ yielding $g(1/m) = g(1)^{1/m}$.

Thus for any rational number $r = n/m$,

$$r = \frac{n}{m} = \frac{1}{m} + \frac{1}{m} + \cdots + \frac{1}{m} \quad (n \text{ summands}),$$

yielding $g(r) = g(1/m)^n = g(1)^{n/m} = g(1)^r$. Finally, we can, for any irrational $x > 0$, choose a decreasing sequence of rational numbers, $r_1 > r_2 > \cdots$, such that $r_n \to x$, as $n \to \infty$. Since g is the tail of a c.d.f., it is right-continuous in x and hence

$$\begin{aligned} g(x) &= \lim_{n\to\infty} g(r_n) \\ &= \lim_{n\to\infty} g(1)^{r_n} \\ &= g(1)^x. \end{aligned}$$

We conclude that $g(x) = g(1)^x,\ x \geq 0$, and since $g(x) = P(X > x) \to 0$ as $x \to \infty$, we conclude that $0 \leq g(1) < 1$. But $g(1) > 0$, for otherwise $g(x) = P(X > x) = 0,\ x \geq 0$ implies that $P(X = 0) = 1$, a contradiction to the assumption that X not be identically 0. Thus $0 < g(1) < 1$. Since $g(1)^x = e^{x\ln(g(1))}$ we finally obtain $g(x) = e^{-\lambda x}$, where $\lambda \stackrel{\text{def}}{=} -\ln(g(1)) > 0$.
■

1.2 Relation to the geometric distribution

In a discrete r.v. setting, the memoryless property is given by

$$P(X - k > n | X > k) = P(X > n),$$

for non-negative integers k, n. The only discrete distribution with this property is the geometric distribution; $P(X = n) = (1-p)^{n-1}p,\ n \geq 1$ (success probability p). Thus the exponential distribution can be viewed as the continuous analog of the geometric distribution. To make this rigorous: Fix n large, and perform, using (tiny) success probability $p_n = \lambda/n$, an independent Bernoulli trial at each time point $i/n,\ i \geq 1$. Thus we are performing such a Bernoulli trial every (tiny) $1/n$ units of time. Let Y_n denote the *time* at which the first success occurred. Then $Y_n = K_n/n$ where K_n denotes the number of trials until the first success, and has the geometric distribution with success probability p_n; $P(K_n = k) = (1-p_n)^{k-1}p_n,\ k \geq 1$. As $n \to \infty$, Y_n converges in distribution to a r.v. Y having the exponential distribution with rate λ (we use the tail probabilities):

$$\begin{aligned} P(Y_n > x) &= P(K_n > nx) \\ &= (1-p_n)^{nx} \\ &= (1-(\lambda/n))^{nx} \\ &\to e^{-\lambda x},\ n \to \infty \\ &= P(Y > x). \end{aligned}$$

The limit is computed from elementary calculus by taking natural logarithms and using L'Hô pital's rule: $(1-(\lambda/n))^{nx} \to e^{-\lambda x}$ if and only if $nx \ln(1-(\lambda/n)) \to -\lambda x$, as $n \to \infty$. We re-write $nx \ln(1-(\lambda/n))$ as $f(n)/g(n)$ where $f(n) = \ln(1-(\lambda/n))$ and $g(n) = 1/(nx)$, then obtain the limit via $\lim_{n\to\infty} f'(n)/g'(n) = -\lambda x$.

In essence, intuitively, we can construct an exponential rv Y at rate λ, by performing iid Bernoulli trials every (infinitesimal) dx units of time with (infinitesimally small) success probability λdx, and define Y as the time until the first success.

1.3 Useful properties of the exponential distribution

Other useful properties of the exponential distribution are given by

Proposition 1.2 *If X_1 has an exponential distribution with rate λ_1, and X_2 has an exponential distribution with rate λ_2 and the two r.v.s. are independent, then*

1. *The minimum of X_1 and X_2, $Z = \min\{X_1, X_2\}$, has an exponential distribution with rate $\lambda = \lambda_1 + \lambda_2$;*
$$P(Z > x) = e^{-(\lambda_1+\lambda_2)x},\ x \geq 0.$$

2. $$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

3. *The r.v. $Z = \min\{X_1, X_2\}$ is independent of the two events $\{Z = X_1\} = \{X_1 < X_2\}$ and $\{Z = X_2\} = \{X_2 < X_1\}$: Z is independent of which one of the two r.v.s. is in fact the minimum. This means that*
$$\begin{aligned} P(Z > x|X_1 < X_2) &= e^{-(\lambda_1+\lambda_2)x},\ x \geq 0, \\ P(Z > x|X_2 < X_1) &= e^{-(\lambda_1+\lambda_2)x},\ x \geq 0. \end{aligned}$$
This implies that $E(Z \mid X_1 < X_2) = E(Z \mid X_2 < X_1) = E(Z) = \frac{1}{\lambda_1+\lambda_2}$.

Proof : 1. Observing that $Z > x$ if and only if both $X_1 > x$ and $X_2 > x$, we conclude that $P(Z > x) = P(X_1 > x, X_2 > x) = P(X_1 > x)P(X_2 > x)$ (from independence) $= e^{-\lambda_1 x}e^{-\lambda_2 x} = e^{-(\lambda_1+\lambda_2)x}$.

2. Let $f_1(x) = \lambda_1 e^{-\lambda_1 x}$ denote the density function for X_1.

$P(X_1 < X_2|X_1 = x) = P(X_2 > x|X_1 = x) = P(X_2 > x)$ (from independence) $= e^{-\lambda_2 x}$, and thus

$$\begin{aligned} P(X_1 < X_2) &= \int_0^\infty P(X_1 < X_2|X_1 = x) f_1(x) dx \\ &= \int_0^\infty e^{-\lambda_2 x}\lambda_1 e^{-\lambda_1 x} dx \\ &= \lambda_1 \int_0^\infty e^{-(\lambda_1+\lambda_2)x} dx \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

3. $P(Z > x|Z = X_1) = P(X_1 > x|X_1 < X_2)$ so we must prove that $P(X_1 > x|X_1 < X_2) = e^{-(\lambda_1+\lambda_2)x}$, $x \geq 0$. To this end we condition on $X_2 = y$ for all values $y > x$, noting that $P(x < X_1 < X_2|X_2 = y) = P(x < X_1 < y)$:

$$\begin{aligned}
P(Z > x|Z = X_1) &= P(X_1 > x|X_1 < X_2) \\
&= \frac{P(X_1 > x, X_1 < X_2)}{P(X_1 < X_2)} \\
&= \frac{P(x < X_1 < X_2)}{P(X_1 < X_2)} \\
&= \frac{\int_x^\infty \{P(x < X_1 < y)\lambda_2 e^{-\lambda_2 y}\}dy}{P(X_1 < X_2)} \\
&= \frac{\int_x^\infty \{(e^{-\lambda_1 x} - e^{-\lambda_2 y})\lambda_2 e^{-\lambda_2 y}\}dy}{P(X_1 < X_2)} \\
&= \frac{e^{-(\lambda_1+\lambda_2)x}\frac{\lambda_1}{\lambda_1+\lambda_2}}{P(X_1 < X_2)} \\
&= \frac{e^{-(\lambda_1+\lambda_2)x}\frac{\lambda_1}{\lambda_1+\lambda_2}}{\frac{\lambda_1}{\lambda_1+\lambda_2}} \quad \text{from } \textit{2.} \text{ above} \\
&= e^{-(\lambda_1+\lambda_2)x}.
\end{aligned}$$

Hence, given $Z = X_1, Z$ is (still) exponential with rate $\lambda_1 + \lambda_2$. Similarly if $Z = X_2$. The point here is that the minimum is exponential at rate $\lambda_1 + \lambda_2$ regardless of knowing which of the two is the minimum.

■

Examples

Here we illustrate, one at a time, each of *1,2,3* of Proposition 1.2.

Suppose you have two computer monitors (independently) one in your office having lifetime X_1 exponential with rate $\lambda_1 = 0.25$ (hence mean = 4 years), and the other at home having lifetime X_2 exponential with $\lambda_2 = 0.5$ (hence mean = 2 years). As soon as one of them breaks, you must order a new monitor.

1. What is the expected amount of time until you need to order a new monitor?

   The amount of time is given by $Z = \min\{X_1, X_2\}$ and has an exponential distribution at rate $\lambda_1 + \lambda_2$; $E(Z) = 1/(\lambda_1 + \lambda_2) = 1/(0.75) = 4/3$ years.

2. What is the probability that the office monitor is the first to break?

   $P(X_1 < X_2) = \lambda_1/(\lambda_1 + \lambda_2)$
   $= 0.25/(0.25 + 0.50) = 1/3$.

3. Given that the office monitor broke first, what was its expected lifetime?

   The lifetime is given by $Z = \min\{X_1, X_2\}$ and has an exponential distribution at rate $\lambda_1 + \lambda_2$ regardless of knowing that $X_1 < X_2$; thus the answer remains $E(Z \mid X_1 < X_2) = E(Z) = 4/3$, as in Question 1.

Remark 1 *1.* above generalizes to any finite number of independent r.v.s.: $\min\{X_1, X_2, \ldots, X_n\} \sim exp(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ if $X_i \sim exp(\lambda_i),\ 1 \le i \le n$.

$P(\min\{X_1, X_2, \ldots, X_n\} > x) = P(X_1 > x, X_2 > x, \ldots, X_n > x)$
$= P(X_1 > x)P(X_2 > x) \cdots P(X_n > x)$
$= e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)x}$.

1.4 Simulating samples X with the exponential distribution

Recalling the *inverse transform method*, the inverse function, $F^{-1}(y),\ y \in (0,1)$, of $F(x) = 1 - e^{-\lambda x},\ x \ge 0$ is given by solving

$$y = 1 - e^{-\lambda x},$$

for x in terms of y yielding

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y),$$

where $\ln(x)$ denotes the natural logarithm of $x > 0$.

Thus for U uniform over $(0,1)$, we can set $X = -\frac{1}{\lambda}\ln(1-U)$ to get an $X \sim exp(\lambda)$. But since $1 - U$ is itself uniform over $(0,1)$, we can simplify the algorithm to

Algorithm for simulating/generating $X \sim exp(\lambda)$:

1. Generate U (uniform over $(0,1)$).
2. Set $X = -\frac{1}{\lambda}\ln(U)$.

This immediately leads to a simulation algorithm for a so-called *Poisson process*. If a random sequence of times $0 < t_1 < t_2 < \cdots$, are defined by $t_n = X_1 + \cdots + X_n$, where the X_i are iid with an exponential distribution at rate λ, the *point process*, $\{t_n : n \ge 1\}$ is called a Poisson point process, or just a Poisson process for short.

If you imagine t_n as the time that the n^{th} call to your mobile phone comes in, then a Poisson process can approximate $\{t_n : n \ge 1\}$ under suitable conditions. We will study such processes in detail later but let us observe how we can easily simulate $\{t_n : n \ge 1\}$ out to any desired n or out to any desired time T:

Algorithm for simulating a Poisson process at rate λ up to the n^{th} point, $n \ge 1$:

1. Generate $U_1, \ldots, U_n$ (iid uniforms over $(0,1)$).
2. Set $t_0 = 0$ and then recursively set $t_{i+1} = t_i - \frac{1}{\lambda}\ln(U_{i+1}),\ 0 \le i \le n-1$.

Algorithm for simulating a Poisson process at rate λ up to time $T > 0$:

1. Set $t_0 = 0 = i = N$
2. Generate U. Set $i = i + 1$. set $t_i = t_{i-1} - \frac{1}{\lambda}\ln(U)$.
3. If $t_i \le T$, then set $N = N + 1$ and go back to (2); otherwise if $t_i > T$, then stop. If $N = 0$ output "No points by time T"; otherwise output N and $(t_1, \ldots, t_N)$.