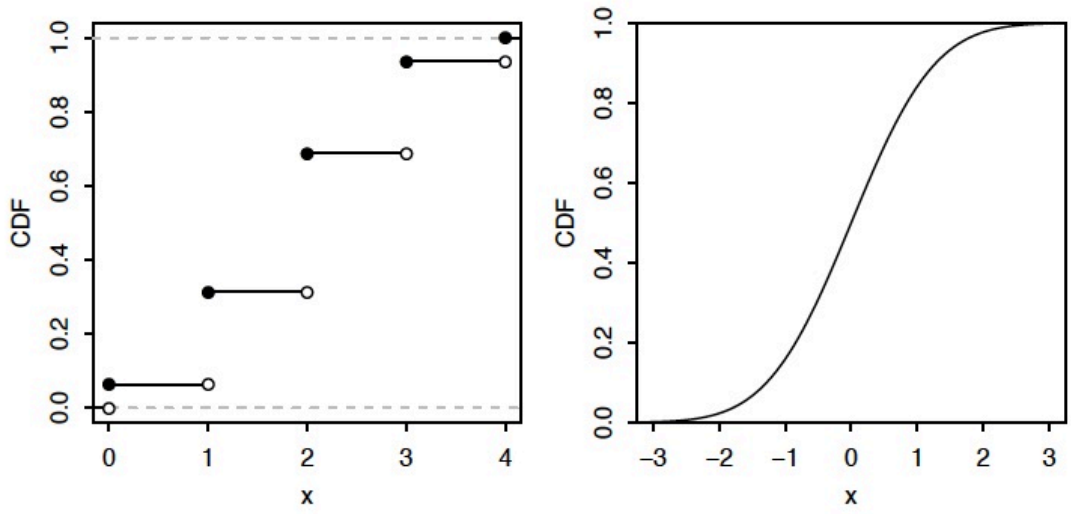

Figure 1: Discrete vs. continuous random variables. Left panel: The CDF of a discrete random variable has jumps at each point in the support. Right panel: The CDF of a continuous random variable increases smoothly. An important way in which continuous random variables differ from discrete random variables is that for a continuous random variable X, $\mathsf{P}(X = x) = 0$ for all x. This is because $\mathsf{P}(X = x) = 0$ is the height of a jump of the CDF at x, but the CDF of X has no jumps.

1 Introduction

We introduce continuous random variables, which can take on any real value in an interval (possibly of infinite length, such as $(0, \infty)$ or the entire real line $\mathbb{R}$).

Definition (Continuous random variables)**.** A random variable has a *continuous distribution* if its CDF is differentiable. We allow there to be finitely many points where the CDF if continuous but not differentiable, as long as the CDF is differentiable everywhere else. A *continuous random variable* is a random variable with a continuous distribution.

Definition (Probability density function)**.** For a continuous random variable X with CDF F, the *probability density function* (PDF) of X is the derivative $f(x) = F'(x) = \frac{dF(x)}{dx}$. The *support* of X, and of its distribution, is the set of all x where $f(x) > 0$.

Proposition (PDF to CDF)**.** *Let X be a continuous random variable with PDF f. Then the CDF of X is*

$$F(x) = \mathsf{P}(X \in (-\infty, x]) = F(x) - F(-\infty) = \int_{-\infty}^{x} f(t)dt.$$

The CDF is the accumulated area under the PDF. To find the probability of X falling into an arbitrary region $A \subseteq \mathbb{R}$, we integrate the PDF of X over A:

$$\mathsf{P}(X \in A) = \int_A f(x)dx.$$

In particular, for an interval $(a, b) \subset \mathbb{R}$, we have

$$\mathsf{P}(a < X < b) = \mathsf{P}(a \leq X < b) = \mathsf{P}(a < X \leq b) = \mathsf{P}(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx.$$

Theorem (Valid PDFs)**.** *The PDF f of a continuous random variable must satisfy the following two criteria:*

- *Nonnegative:* $f(x) \geq 0$.
- *Integrates to* 1*:* $\int_{-\infty}^{\infty} f(x)dx = 1$.

Note that it is possible to have $f(x) > 1$ for some values of x.

Definition (Expectation of a continuous random variable)**.** The *expected value* (also called the *expectation* or *mean*) of a continuous random variable X with PDF f is

$$\mathsf{E}(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

The expected value may or may not exist. The above integral is taken over the entire real line, but if the support of X is not the entire real line we can just integrate over the support.

Theorem (LOTUS, continuous)**.** *If X is a continuous random variable with PDF f and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then*

$$\mathsf{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

2 The Uniform distribution

A Uniform random variable on the interval (a, b) is a completely random number between a and b.

Definition (Uniform distribution)**.** A continuous random variable U is said to have the *Uniform distribution* on the interval (a, b) if its PDF is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

We write $U \sim \mathsf{Unif}(a, b)$. The corresponding CDF is

$$F(x) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{x-a}{b-a} & \text{if } a < x < b, \\ 1 & \text{if } x \geq b. \end{cases}$$

For Uniform distributions, probability is proportional to length.

Proposition. *Let $U \sim \mathsf{Unif}(a, b)$, and let (c, d) be a subinterval of (a, b) of length $l = d - c$. Then the probability of U being in (c, d) is proportional to l.*

Even after conditioning on a Uniform random variable being in a certain subinterval, we still have a Uniform distribution.

Proposition. *Let $U \sim \mathsf{Unif}(a,b)$, and let (c,d) be a subinterval of (a,b). Then the conditional distribution of U given $U \in (c,d)$ is $\mathsf{Unif}(c,d)$.*

Proof. For $u \in (c,d)$, the conditional CDF of u is

$$\mathsf{P}(U \le u \mid U \in (c,d)) = \frac{\mathsf{P}(U \le u, c < U < d)}{\mathsf{P}(U \in (c,d))} = \frac{\mathsf{P}(U \in (c,u])}{\mathsf{P}(U \in (c,d))} = \frac{u-c}{d-c}.$$

□

Next, we derive the mean and the variance of a Uniform random variable using a powerful technique called *location-scale transformation*. This technique works for any family of distributions such that shifting (=a change in location) and scaling (=a change in scale) a random variable whose distribution is in the family produces another random variable whose distribution is also in the same family.

Definition (Location-scale transformation)**.** Let X be a random variable and $Y = \sigma X + \mu$, where σ and μ are constants with $\sigma > 0$. Then we say that Y has been obtained as a *location-scale transformation* of X. Here μ controls how the location is changed, and σ controls how the scale is changed.

Starting with $X \sim \mathsf{Unif}(a,b)$ and transforming to $Y = cX + d$ where c and d are constants with $c > 0$, Uniformity is preserved: $Y \sim \mathsf{Unif}(ca+d, cb+d)$.

To find the expectation and variance of the $\mathsf{Unif}(a,b)$ distribution, we start with $U \sim \mathsf{Unif}(0,1)$. The $\mathsf{Unif}(0,1)$ PDF and CDF are: $f(x) = 1$ and $F(x) = x$ for $0 < x < 1$. Then the mean and the variance of $U \sim \mathsf{Unif}(0,1)$ are:

$$\mathsf{E}(U) = \int_0^1 x dx = \frac{1}{2}, \quad \mathsf{E}(U^2) = \int_0^1 x^2 dx = \frac{1}{3}, \quad \mathsf{Var}(U) = \mathsf{E}(U^2) - (\mathsf{E}U)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

We consider the random variable

$$\tilde{U} = a + (b-a)U \sim \mathsf{Unif}(a,b),$$

that is the location-scale transformation of U with a scaling factor $b-a$ and a location factor a. By the known properties of expectation and variance, we have:

$$\begin{aligned} \mathsf{E}(\tilde{U}) &= a + (b-a)\mathsf{E}(U) = a + \frac{b-a}{2} = \frac{a+b}{2}. \\ \mathsf{Var}(\tilde{U}) &= (b-a)^2 \mathsf{Var}(U) = \frac{(b-a)^2}{12}. \end{aligned}$$

The Uniform distribution has a remarkable property: given a $\mathsf{Unif}(0,1)$ random variable, we can construct another random variable with any continuous distribution we want. We call it the *universality of the Uniform* because it tells us the Uniform is a universal starting point for building random variables with other distributions. This property is known by other names such as: the probability integral transform, inverse transform sampling, the quantile transformation, or the fundamental theorem of simulation.

Theorem (Universality of the Uniform)**.** *Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} (called the quantile function) exists, as a function from $(0,1)$ to $\mathbb{R}$. The following hold:*

1. *Let $U \sim \mathsf{Unif}(0,1)$ and $X = F^{-1}(U)$. Then X is a random variable with CDF F.*
2. *Let X be a random variable with CDF F. Then $F(X) \sim \mathsf{Unif}(0,1)$.*

Proof. 1. Let $U \sim \mathsf{Unif}(0,1)$ and $X = F^{-1}(U)$. For all real x,

$$\mathsf{P}(X \leq x) = \mathsf{P}(F^{-1}(U) \leq x) = \mathsf{P}(U \leq F(x)) = F(x),$$

hence the CDF of X is F, as claimed.

2. Let X have CDF F, and find the CDF of $Y = F(X)$. Since $Y \in (0,1)$, $\mathsf{P}(Y \leq y)$ equals 0 for $y \leq 0$ and equals 1 for $y \geq 1$. For $y \in (0,1)$, we have

$$\mathsf{P}(Y \leq y) = \mathsf{P}(F(X) \leq y) = \mathsf{P}(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y.$$

Thus Y has the $\mathsf{Unif}(0,1)$ CDF. □

Example: Universality with Logistic

The Logistic distribution has CDF

$$F(x) = \frac{e^x}{1+e^x}, \quad x \in \mathbb{R}.$$

The quantile function (the inverse of this CDF) is

$$F^{-1}(u) = \log\left(\frac{u}{1-u}\right).$$

Part 1 of the Universality of the Uniform property says that, if $U \sim \mathsf{Unif}(0,1)$, then

$$F^{-1}(U) = \log\left(\frac{U}{1-U}\right) \sim \text{Logistic}.$$

Conversely, Part 2 of the Universality of the Uniform property says that, if $X \sim$ Logistic, then

$$F(X) = \frac{e^X}{1+e^X} \sim \mathsf{Unif}(0,1).$$

Example: Universality with Rayleigh

The Rayleigh distribution has CDF

$$F(x) = 1 - e^{-x^2/2}, \quad x > 0.$$

The quantile function (the inverse of this CDF) is

$$F^{-1}(u) = \sqrt{-2\log(1-u)}.$$

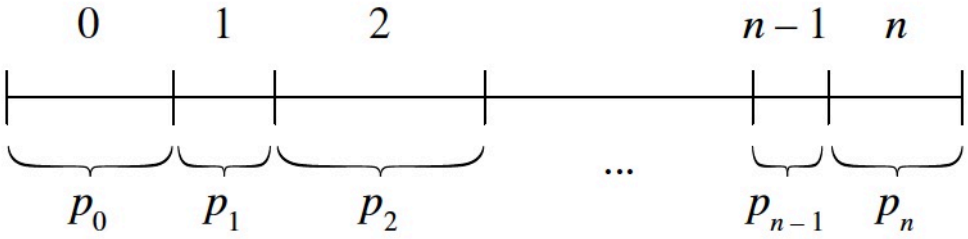

Figure 2: Given a PMF, divide the interval $(0, 1)$ into pieces, with lengths given by the PMF values.

Part 1 of the Universality of the Uniform property says that, if $U \sim \text{Unif}(0, 1)$, then

$$F^{-1}(U) = \sqrt{-2\log(1-U)} \sim \text{Rayleigh}.$$

Conversely, Part 2 of the Universality of the Uniform property says that, if $X \sim$ Rayleigh, then

$$F(X) = 1 - e^{-X^2/2} \sim \text{Unif}(0, 1).$$

In the sequel, we show how to construct (simulate) a random variable with any discrete distribution we want from $U \sim \text{Unif}(0, 1)$. Specifically, we want to construct a discrete random variable X with PMF $p_j = \mathsf{P}(X = j)$ for $j = 0, 1, 2, \ldots, n$. We must have $p_j > 0$ and $\sum_{j=0}^{n} p_j = 1$. Please see Figure 2. To simulate from this discrete distribution, we draw $U \sim \text{Unif}(0, 1)$, then identify the interval of length p_j to which U belongs (such an interval must exist and it is unique). The simulated value is $X = j$. This is correct because, the probability that $X = j$ is the probability that U falls into the interval of length p_j. But for a $\text{Unif}(0, 1)$ random variable, probability is length, so $\mathsf{P}(X = j)$ is equal with p_j.

3 The Normal distribution

The Normal distribution is a famous continuous distribution with a bell-shaped PDF. It is extremely widely used in statistics because of a theorem, the central limit theorem, which says that under very weak assumptions, the sum of a large number of i.i.d. random variables has an approximately Normal distribution, regardless of the distribution of the individual random variables. This means we can start with independent random variables from almost any distribution, discrete or continuous, but once we add up a bunch of them, the distribution of the resulting random variable looks like a Normal distribution.

We begin with the simplest Normal distribution, the *standard Normal*, which is centered at 0 and has variance 1.

Definition (Standard Normal distribution)**.** A continuous random variable Z is said to have the *standard*

Normal distribution if its PDF φ is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

We write this as $Z \sim \mathsf{N}(0,1)$ since Z has mean 0 and variance 1.

The standard Normal CDF is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt.$$

Note that it is mathematically impossible to find a closed-form expression for the antiderivative of ϕ, hence integral in the expression of Φ cannot be eliminated.

Here are several important properties of the standard Normal PDF and CDF.

1) *Symmetry of PDF*: φ satisfies $\varphi(z) = \varphi(-z)$, i.e., φ is an even function.

2) *Symmetry of tail areas*: For all $z \in \mathbb{R}$, we have

$$\Phi(-z) = \int_{-\infty}^{-z} \varphi(t)dt = \int_{z}^{\infty} \varphi(u)du = 1 - \int_{-\infty}^{z} \varphi(u)du = 1 - \Phi(z).$$

3) *Symmetry of Z and $-Z$*: If $Z \sim \mathsf{N}(0,1)$, then $-Z \sim \mathsf{N}(0,1)$ as well. To see this, we write:

$$\mathsf{P}(-Z \le z) = \mathsf{P}(Z \ge -z) = 1 - \Phi(-z) = \Phi(z) = \mathsf{P}(Z \le z).$$

A standard Normal random variable $Z \sim \mathsf{N}(0,1)$ has mean 0 and variance 1:

$$\begin{aligned} \mathsf{E}(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2} dz - \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2} = 0, \\ \mathsf{Var}(Z) &= \mathsf{E}(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z^2 e^{-z^2/2} dz = 1. \end{aligned}$$

The general Normal distribution has two parameters, denoted by μ and σ^2, which correspond to the mean and variance. Starting with a standard Normal random variable $Z \sim \mathsf{N}(0,1)$, we can get a Normal random variable with any mean and variance by a location-scale transformation.

Definition. If $Z \sim \mathsf{N}(0,1)$, then

$$X = \mu + \sigma Z$$

is said to have the *Normal distribution* with mean μ and variance σ^2, for any real μ and σ^2 with $\sigma > 0$. We denote this by $X \sim \mathsf{N}(\mu, \sigma^2)$.

We have:

$$\begin{aligned} \mathsf{E}(X) &= \mathsf{E}(\mu + \sigma Z) = \mathsf{E}(\mu) + \sigma \mathsf{E}(Z) = \mu, \\ \mathsf{Var}(X) &= \mathsf{Var}(\mu + \sigma Z) = \mathsf{Var}(\sigma Z) = \sigma^2 \mathsf{Var}(Z) = \sigma^2. \end{aligned}$$

The process of getting a standard Normal from a non-standard Normal is called *standardization*. For $X \sim \mathsf{N}(\mu, \sigma^2)$, the *standardized version* of X is

$$\frac{X-\mu}{\sigma} \sim \mathsf{N}(0,1).$$

We can use standardization to find the CDF and PDF of $X \sim \mathsf{N}(\mu, \sigma^2)$ in terms of the standard Normal CDF and PDF.

Theorem (Normal CDF and PDF)**.** *Let $X \sim \mathsf{N}(\mu, \sigma^2)$. Then the CDF of X is*

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

and the PDF of X is

$$f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Proof.

$$\begin{aligned} F(x) &= \mathsf{P}(X \le x) = \mathsf{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right), \\ f(x) &= \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}. \end{aligned}$$

□

4 The Exponential distribution

The Exponential distribution is the continuous counterpart to the Geometric distribution. Recall that a Geometric random variable counts the number of failures before the first success in a sequence of Bernoulli trials. The story of the Exponential distribution is analogous, but we are now waiting for a success in continuous time, where successes arrive at a rate of λ successes per unit of time. The average number of successes in a time interval of length t is λt, though the actual number of successes varies randomly. An Exponential random variable represents the waiting time until the first arrival of a success.

Definition (Exponential distribution)**.** A continuous random variable X is said to have the *Exponential distribution* with parameter λ, where $\lambda > 0$, if its PDF is

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

We denote this by $X \sim \mathsf{Expo}(\lambda)$.

We have seen how all Uniform and Normal distributions are related to one other via location-scale transformations, and we might wonder whether the Exponential distribution allows this too. Exponential random variables are defined to have support $(0, \infty)$, and shifting would change the left endpoint (0). But scale transformations work nicely: we can use scaling to get from the simple $\mathsf{Expo}(1)$ to the general $\mathsf{Expo}(\lambda)$.

If $X \sim \mathsf{Expo}(1)$, then

$$Y = \frac{X}{\lambda} \sim \mathsf{Expo}(\lambda),$$

since

$$\mathsf{P}(Y \leq y) = \mathsf{P}\left(\frac{X}{\lambda} \leq y\right) = \mathsf{P}(X \leq \lambda y) = 1 - e^{-\lambda y}, \quad y > 0.$$

Conversely, if $Y \sim \mathsf{Expo}(\lambda)$, then $\lambda Y \sim \mathsf{Expo}(1)$.

The mean and variance of $X \sim \mathsf{Expo}(1)$ are

$$\begin{aligned} \mathsf{E}(X) &= \int_0^\infty x e^{-x} dx = 1, \\ \mathsf{E}(X^2) &= \int_0^\infty x^2 e^{-x} dx = 2, \\ \mathsf{Var}(X) &= \mathsf{E}(X^2) - (\mathsf{E}X)^2 = 1. \end{aligned}$$

For $Y = \frac{X}{\lambda} \sim \mathsf{Expo}(\lambda)$ we then have:

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

The Exponential distribution has a very special property called the *memoryless property*.

Definition (Memoryless property)**.** A distribution is said to have the *memoryless property* if a random variable X with that distribution satisfies

$$\mathsf{P}(X \geq s + t \mid X \geq s) = \mathsf{P}(X \geq t),$$

for all $s, t > 0$.

Here s represents the time you have already spent waiting for an event to happen (e.g., arrival of a bus or arrival of an email). The definition says that after you waited s minutes, the probability that you will have to wait another t minutes is exactly the same as the probability of having to wait t minutes with no previous waiting time. Another way to state the memoryless property is that conditional on $X \geq s$, the additional waiting time $X - s$ is still distributed $\mathsf{Expo}(\lambda)$. This implies

$$\mathsf{E}(X \mid X \geq s) = s + \mathsf{E}(X) = s + \frac{1}{\lambda}.$$

Next we check whether $X \sim \mathsf{Expo}(\lambda)$ has the memoryless property:

$$\mathsf{P}(X \geq s + t \mid X \geq s) = \frac{\mathsf{P}(X \geq s + t)}{\mathsf{P}(X \geq s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathsf{P}(X \geq t).$$

The next result shows that no other continuous distributions on $(0, \infty)$ is memoryless.

Theorem. *If X is a positive continuous random variable with the memoryless property, then X has an Exponential distribution.*

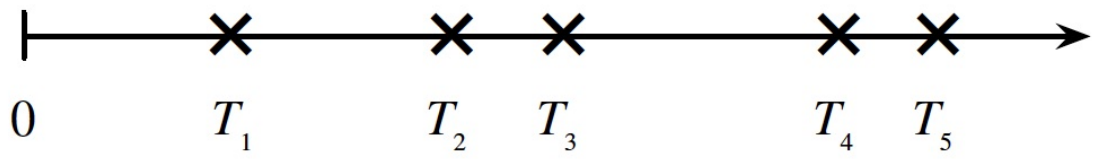

Figure 3: Sketch of a Poisson process. Each × marks the spot of an arrival.

5 Poisson processes

The Exponential distribution is closely connected to the Poisson distribution through *Poisson processes*. A Poisson process is a sequence of arrivals occurring at different time points on a timeline, such that the number of arrivals in a particular time interval of time has a Poisson distribution.

Definition (Poisson process)**.** A process of arrivals in continuous time is called a *Poisson process* with rate λ if the following two conditions hold. (See Figure 3).

1. The number of arrivals that occur in an interval of length t is a $\mathsf{Pois}(\lambda t)$ random variable.
2. The number of arrivals that occur in disjoint time intervals are independent of each other.

Suppose the arrivals are emails landing in an inbox according to a Poisson process. How many emails will arrive in one hour ($t = 1$)? The definition of a Poisson process tells us that the number of emails in an hour follows a $\mathsf{Pois}(\lambda)$ distribution.

But we could also flip the question around and ask: *how long* does it take until the first email arrives, measured relative to some fixed starting point? Let T_1 be the time until the first email arrives. Saying that the waiting time for the first email is greater than t is the same as saying that *no emails* have arrived between 0 and t. Thus, if N_t is the number of emails that arrive at or before time t, then

$$T_1 > t \text{ is the same event as } N_t = 0.$$

This connects the discrete random variable N_t, which counts the number of arrivals with a continuous random variable T_1, which marks the time of the first arrival.

Since $N_t \sim \mathsf{Pois}(\lambda t)$ by the definition of Poisson process, we have

$$\mathsf{P}(T_1 > t) = \mathsf{P}(N_t = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Therefore $\mathsf{P}(T_1 \leq t) = 1 - e^{-\lambda t}$, so $T_1 \sim \mathsf{Expo}(\lambda)$. The time until the first arrival in a Poisson process of rate λ has an Exponential distribution with parameter λ.

Since disjoint intervals in a Poisson process are independent by definition, all the interarrival times $T_j - T_{j-1}$, for $j = 1, 2, \ldots$ are i.i.d. $\mathsf{Expo}(\lambda)$ random variables ($T_0 = 0$). However, we note that the total time

until the second arrival, $T_2 = T_1 + (T_2 - T_1)$ does not follow an Exponential distribution even if it is the sum of two independent $\mathsf{Expo}(\lambda)$ random variables. Instead, T_2 follows a Gamma distribution, to be introduced later.

Example: Minimum of independent Exponentials

Let $X_1, X_2, \ldots, X_n$ be independent, with $X_j \sim \mathsf{Expo}(\lambda_j)$. Let $L = \min(X_1, X_2, \ldots, X_n)$. Show that $L \sim \mathsf{Expo}(\lambda_1 + \lambda_2 + \ldots + \lambda_n)$.

Proof.

$$\mathsf{P}(L > t) = \mathsf{P}(X_1 > t, \ldots, X_n > t) = \mathsf{P}(X_1 > t) \cdot \ldots \cdot \mathsf{P}(X_n > t) = e^{-\lambda_1 t} \cdot \ldots \cdot e^{-\lambda_n t} = e^{-(\lambda_1 + \ldots + \lambda_n)t}.$$

□