Chapter 1

Some reminders in probability

Several important probability distributions arise naturally from the Poisson process: the Poisson distribution, the exponential distribution, and the gamma distribution. This chapter aims to recall basic definitions and properties in probability theory necessary to define and study Poisson processes.

I Poisson random variable

The Poisson distribution is a discrete probability distribution commonly used to model the number of (rare) events occurring in a fixed interval of time or space. For instance, an insurer might want to model the number of claims during a month, or in quality control, one might want to model the number of failures of a manufacturing machine during a week.

Definition 1.1.

Let N be a real random variable. Then N is said to have a Poisson distribution with parameter $\lambda > 0$, denoted $N \sim \mathcal{P}(\lambda)$, if N is discrete with values in $\mathbb{N}$ such that, for $k \in \mathbb{N}$,

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

The parameter λ represents the (constant) mean rate of occurence. The Poisson distribution satisfies the following properties.

Proposition 1.2.

Let $N \sim \mathcal{P}(\lambda)$ with $\lambda > 0$.

1. *The Poisson distribution is characterized by its moment-generating function (m.g.f.), or Laplace transform:*
$$\forall\, t \leq 0, \quad \mathbb{E}\left[e^{tN}\right] = \exp\left(\lambda\left(e^t - 1\right)\right).$$
2. *The Poisson distribution is also characterized by its characteristic function:*
$$\forall\, t \in \mathbb{R}, \quad \mathbb{E}\left[e^{itN}\right] = \exp\left(\lambda\left(e^{it} - 1\right)\right).$$

Proof. In exercise. Let $N \sim \mathcal{P}(\lambda)$, with $\lambda > 0$. Let $t \leq 0$.

1. Compute $\mathbb{E}\left[e^{tN}\right]$.

Solution.

Le m.g.f. of N equals

$$\mathbb{E}\left[e^{tN}\right] = \sum_{k\in\mathbb{N}} \left[e^{tk}\left(\frac{\lambda^k}{k!}e^{-\lambda}\right)\right] = \sum_{k\in\mathbb{N}} \left[\frac{(\lambda e^t)^k}{k!}\right] e^{-\lambda} = \exp\left(\lambda e^t - \lambda\right),$$

which ends the proof.

2. Same computation as in 1.

□

Note that both the moment-generating function and the characteristic function characterize the distribution of a random variable. Hence, if two random variables have the same moment-generating function, or the same characteristic function, then they have the same distribution.

Property 1.3.

Let $N \sim \mathcal{P}(\lambda)$ with $\lambda > 0$. Then $\mathbb{E}[N] = \mathrm{Var}(N) = \lambda$.

Proof. In exercise. Recall that if $L : t \mapsto \mathbb{E}\left[^{tN}\right]$ denotes the moment-generating function, then, for all k in $\mathbb{N}^*$, the kth-order moment of N is linked to the kth-order derivative of L as follows:

$$\mathbb{E}\left[N^k\right] = L^{(k)}(0).$$

Deduce $\mathbb{E}[N]$ and $\mathrm{Var}(N)$.

Solution.

- Expectation: Since for all $t \leq 0$,

$$L'(t) = \lambda e^t \exp\left(\lambda\left(e^t - 1\right)\right) = \lambda \exp\left(\lambda e^t + t - \lambda\right),$$

we directly obtain

$$\mathbb{E}[N] = L'(0) = \lambda.$$

- Variance: Since for all $t \leq 0$,

$$L''(t) = \lambda(\lambda e^t + 1)\exp\left(\lambda e^t + t - \lambda\right),$$

we directly obtain

$$\mathbb{E}\left[N^2\right] = L''(0) = \lambda(\lambda + 1),$$

and thus

$$\mathrm{Var}(N) = \mathbb{E}\left[N^2\right] - \mathbb{E}[N]^2 = \lambda.$$

□

II Exponential random variable

The exponential distribution is a continuous probability distribution commonly used to model lifetimes, or more generally time elapsed between events.

Definition 1.4.
Let X be a real random variable. Then X is said to have an exponential distribution with parameter $\lambda > 0$, denoted $X \sim \mathcal{E}(\lambda)$, if X has a density w.r.t. the Lebesgue measure defined on $\mathbb{R}$ by

$$x \mapsto \lambda e^{-\lambda x} \mathbb{1}_{\{x>0\}}.$$

The exponential distribution can be characterized by the following functions.

Proposition 1.5.
Let $X \sim \mathcal{E}(\lambda)$ with $\lambda > 0$.

1. *Its moment-generating function (m.g.f.), or Laplace transform, equals for all $t < \lambda$,*

$$\mathbb{E}\left[e^{tX}\right] = \frac{\lambda}{\lambda - t}.$$

2. *Its distribution functions equals:* $F : t \in \mathbb{R} \mapsto \left(1 - e^{-\lambda t}\right) \mathbb{1}_{\{t \geq 0\}}$.
3. *Its survival function satisfies:* $\forall t \geq 0, \quad \mathbb{P}(X > t) = e^{-\lambda t}$.

Proof. In exercise.

1. a) Let $t < \lambda$. Prove that $\mathbb{E}\left[e^{tX}\right] = \lambda \int_0^{+\infty} e^{-(\lambda - t)x} dx$.
   b) Deduce that $\mathbb{E}\left[e^{tX}\right]$ is well-defined only if $\lambda - t > 0$ and compute the Laplace transform.

   Solution.

$$\mathbb{E}\left[e^{tX}\right] = \int_0^{+\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-(\lambda - t)x} dx \overset{(*)}{=} \lambda \left[\frac{-1}{\lambda - t} e^{-(\lambda - t)x}\right]_0^{+\infty} = \frac{\lambda}{\lambda - t}.$$

   Note that the integral $^{(*)}$ is well-defined only if $\lambda - t > 0$.

   Note that in this case, the Laplace transform is positive as expected.

2. Compute $\mathbb{P}(X \leq t)$ by distinguishing the cases $t < 0$ and $t \geq 0$.

   Solution.

   - Since X takes positive values, if $t < 0$, then $\mathbb{P}(X \leq t) = 0$.
   - Let $t \geq 0$. Then

$$\mathbb{P}(X \leq t) = \int_0^t \lambda e^{-\lambda s} ds = \left[-e^{-\lambda s}\right]_0^t = 1 - e^{-\lambda t}.$$

3. Immediate from 2 since $\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t)$.

□

Property 1.6.
Let $X \sim \mathcal{E}(\lambda)$ with $\lambda > 0$. Then $\mathbb{E}[X] = 1/\lambda$ and $\mathrm{Var}(X) = 1/\lambda^2$.

Proof. In exercice, using the moment-generating function (m.g.f.) as in the proof of Property 1.3.

Solution.

Let $X \sim \mathcal{E}(\lambda)$ and denote $L : t \mapsto \mathbb{E}[tX] = \lambda/(\lambda - t)$ its m.g.f.
Then L is differentiable on $]-\infty; \lambda[$, with

$$L'(t) = \frac{\lambda}{(\lambda - t)^2} \qquad \text{and} \qquad L''(t) = \frac{2\lambda}{(\lambda - t)^3}.$$

Hence

$$\mathbb{E}[X] = L'(0) = \frac{1}{\lambda}, \quad \mathbb{E}\left[X^2\right] = L''(0) = \frac{2}{\lambda^2}, \quad \text{and} \quad \mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}.$$

□

Property 1.7.

Let $X_1 \sim \mathcal{E}(\lambda_1)$ and $X_2 \sim \mathcal{E}(\lambda_2)$ be two independent exponential random variables. Then

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

Proof. In exercise. *Hint:* $\mathbb{P}(X_1 < X_2) = \mathbb{E}\left[\mathbb{1}_{\{X_1 < X_2\}}\right]$.

Solution.

$$\begin{aligned}
\mathbb{P}(X_1 < X_2) &= \mathbb{E}\left[\mathbb{1}_{\{X_1 < X_2\}}\right] \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x_1 < x_2\}} \left(\lambda_1 e^{-\lambda_1 x_1} \mathbb{1}_{\{x_1 > 0\}}\right) \left(\lambda_2 e^{-\lambda_2 x_2} \mathbb{1}_{\{x_2 > 0\}}\right) dx_1 dx_2 \\
&= \int_0^{+\infty} \lambda_1 e^{-\lambda_1 x_1} \left(\int_{x_1}^{+\infty} \lambda_2 e^{-\lambda_2 x_2} dx_2\right) dx_1 \\
&= \int_0^{+\infty} \lambda_1 e^{-\lambda_1 x_1} \left[-e^{-\lambda_2 x_2}\right]_{x_1}^{+\infty} dx_1 \\
&= \int_0^{+\infty} \lambda_1 e^{-(\lambda_1 + \lambda_2) x_1} dx_1 \\
&= \left[\frac{-\lambda_1}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2) x_1}\right]_0^{+\infty} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

□

Example: A system consists of two independent components in series (the system fails if at least one of the components fails). Denote X_i the time (in years) to failure of component i and assume that $X_1 \sim \mathcal{E}(2)$ and $X_2 \sim \mathcal{E}(6)$. In average, component 1 runs 6 months before failure, whereas component 2 runs only 2 months before failure. So in average, component 2 fails before component 1. In fact, according to Property 1.7, the probability that the first component fails before the second one equals 1/4 (it is coherent).

Proposition 1.8.

Let $(X_i)_{1\leq i\leq n}$ be independent exponential r.v. such that for all $1 \leq i \leq n$, $X_i \sim \mathcal{E}(\lambda_i)$. Then

$$\min_{1\leq i\leq n} X_i \sim \mathcal{E}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Proof. In exercise. *Hint:* Compute the survival function of $\min_{1\leq i\leq n} X_i$.

Solution.

Let $t > 0$. The survival function equals

$$\begin{aligned}\mathbb{P}\left(\min_{1\leq i\leq n} X_i > t\right) &= \mathbb{P}\left(\bigcap_{1\leq i\leq n} \{X_i > t\}\right) \overset{(*)}{=} \prod_{1\leq i\leq n} \mathbb{P}(X_i > t) \\ &= \prod_{1\leq i\leq n} e^{-\lambda_i t} = \exp\left(-\left[\sum_{i=1}^{n} \lambda_i\right] t\right),\end{aligned}$$

where $^{(*)}$ comes from the independence between the X_i's. We recognize the survival function of an exponential r.v. with parameter $(\sum_{i=1}^{n} \lambda_i)$.

□

Example: A system consists of 12 independent components in series with exponential time to failure (in years) $\mathcal{E}(2)$. Individually, the average time to failure of each component is 6 months. However, the time to failure of the system has an exponential distribution $\mathcal{E}(24)$. In particular, its average time to failure is about 15 days.

III Memoryless distribution

Definition 1.9.

A random variable X is said to have a *memoryless distribution* if for all $t, s > 0$,

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)\,. \tag{1.1}$$

Note that this definition is well defined if for all $t > 0$, $\mathbb{P}(X > t) > 0$.

Example: In reliability theory, if we think of X being the time to failure of a component, then (1.1) states that the probability that the component still runs after $s + t$ hours knowing that it is still working after t hours is the same as the initial probability that it runs for at least s hours. In other words, if the component is still working at time t, then the distribution of the remaining amount of time that it works is the same as the original operating distribution; that is, the instrument does not "remember" that is has already been used for a time t.

Proposition 1.10.

The exponential distribution is the only continuous memoryless distribution, i.e.,

1. *If $X \sim \mathcal{E}(\lambda)$, then X has a memoryless distribution.*
2. *If X has a continuous and memoryless distribution, then there exists $\lambda > 0$ such that $X \sim \mathcal{E}(\lambda)$.*

The proof of Proposition 1.10 relies on the following Lemma.

Lemma 1.11.

The functional equation

$$f(x+y) = f(x)f(y) \quad \forall x, y \in \mathbb{R}$$

has a unique continuous solution that is $f : x \in \mathbb{R} \mapsto [f(1)]^x$.

Proof of Proposition 1.10. In exercise.

1. Let $X \sim \mathcal{E}(\lambda)$. Let $t, s > 0$ and compute $\mathbb{P}(X > t+s | X > t)$.

2. Let X be a continuous r.v. with a memoryless distribution. Denote $G : t \mapsto \mathbb{P}(X > t)$ its survival function.

   a) Justify that G is a continuous function.

   Solution.

   Since X is a continuous r.v., then its cumulative distribution function, and thus its survival function, are continuous functions.

   b) Let $s, t > 0$. Prove that $G(t+s) = G(s)G(t)$.

   Solution.

   By definition of the memoryless property,

$$\begin{aligned} G(t+s) = \mathbb{P}(X > t+s) &= \mathbb{P}(X > t+s | X > t)\, \mathbb{P}(X > t) \\ &= \mathbb{P}(X > s)\, \mathbb{P}(X > t) = G(s)G(t). \end{aligned}$$

   c) Deduce from Lemma 1.11, that for all $t > 0$, $G(t) = e^{-\lambda t}$ where λ is a well-chosen constant (to determine).

   Solution.

   Since G is a continuous function satisfying the functional equality of Lemma 1.11, we can deduce that for all t,

$$G(t) = [G(1)]^t = \exp\left(\ln[G(1)]t\right) = \exp\left(-\lambda t\right),$$

   where $\lambda = -\ln[G(1)] > 0$ since $G(1) \in]0, 1[$.

   d) Conclude.

   Solution.

   We recognize the survival function of an exponential distribution $\mathcal{E}(\lambda)$.

□

Comment 1.12. Discrete memoryless distributions are geometric distributions.

IV Hazard rate

In this section, we only consider non-negative continuous real-valued random variables.

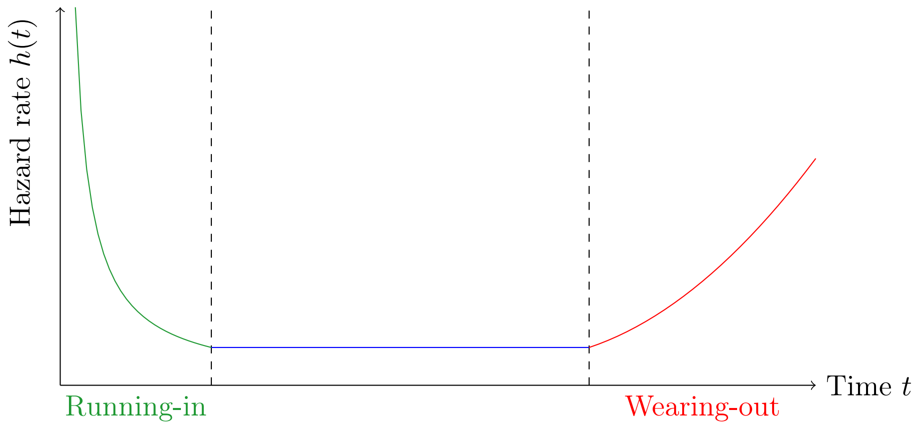
Definition 1.13.
Let X be a non-negative real random variable with density function f w.r.t. the Lebesgue measure. The hazard rate function h is defined by

$$h(t) = \begin{cases} \dfrac{f(t)}{\mathbb{P}(X > t)} & \text{if } \mathbb{P}(X > t) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Example: Notice that, heuristically,

$$h(t)dt = \frac{f(t)dt}{\mathbb{P}(X > t)} \approx \frac{\mathbb{P}(X \in]t, t + dt])}{\mathbb{P}(X > t)} = \mathbb{P}(X \in]t + dt] \, | X > t) \, .$$

For instance, in reliability theory, when X denotes the time to failure of a component, the hazard rate corresponds to the probability of a failure occurring juste after time t given that the device is still working at time t. The survival function $R : t \mapsto \mathbb{P}(X > t)$ is called *reliability function*, and the hazard rate is called *failure rate* and is usually denoted λ. In particular, if the failure rate h is decreasing, it means that the probability of failure decreases with the age. On the contrary, if the failure rate h is increasing, it means that the probability of failure increases with the age. In the reliability theory, the failure rate is shaped as below.



Proposition 1.14.
Let X is a continuous r.v. Then the hazard rate of X is h, if and only if, for all $t > 0$,

$$\mathbb{P}(X > t) = \exp\left(-\int_0^t h(x)dx\right) .$$

In particular, if X and Y are two continuous r.v with the same hazard rate, then they have the same distribution.

Proof. In exercise. Denote $G : t \mapsto \mathbb{P}(X > t)$ the survival function of X.

$\boxed{\Rightarrow}$ Denote h the hazard rate of X.

a) Prove that for all $t > 0$, $\mathbb{P}(X \le t) = \int_0^t h(s)\mathbb{P}(X > s)\,ds$, and deduce that

$$G(t) = 1 - \int_0^t h(s)G(s)ds. \tag{1.2}$$

Solution.

Since the hazard rate of X is h, then its density equals for all $s > 0$, $f(s) = h(s)\mathbb{P}(X > s)$. In particular, its c.d.f equals

$$\mathbb{P}(X \le t) = \int_0^t f(s)ds = \int_0^t h(s)\mathbb{P}(X > s)\,ds,$$

that is $1 - G(t) = \int_0^t h(s)G(s)ds$.

b) Deduce that G is differentiable and satisfies

$$\begin{cases} G(0) = g_0 \\ G'(t) = -h(t)G(t), \end{cases}$$

where g_0 is a constant to determine.

Solution.

From (1.2), we directly deduce that G is differentiable and satisfies $-G'(t) = h(t)G(t)$. Moreover, $G(0) = \mathbb{P}(X > 0) = 1$.

c) Deduce the expression of $G(t)$.

Solution.

G is the unique solution of the homogeneous linear differential equation with initial condition $G(0) = 1$. Hence,

$$G(t) = \underbrace{G(0)}_{1} \exp\left(-\int_0^t h(s)ds\right).$$

$\Leftarrow$ Assume that there exists a function h with survival function $G(t) = \mathbb{P}(X > t) = \exp\left(-\int_0^t h(x)dx\right)$.

i) Compute the cumulative distribution function of X and prove that its density function is $f : t \mapsto f(t) = h(t)G(t)$.

Solution.

The c.d.f. of X equals for all $t > 0$

$$F(t) = 1 - G(t) = 1 - \exp\left(-\int_0^t h(x)dx\right).$$

Hence, the density of X equals for all $t > 0$

$$f(t) = F'(t) = -\left[-h(t)\ \exp\left(-\int_0^t h(x)dx\right)\right] = h(t)G(t).$$

ii) Deduce the hazard rate of X.

Solution.

Hence, by definition, the hazard rate of X equals

$$\frac{f(t)}{G(t)} = h(t),$$

which ends the proof.

□

Property 1.15.

Let $\lambda > 0$. The random variable $X \sim \mathcal{E}(\lambda)$ if and only if its hazard rate function is constant: $h(t) = \lambda$ for all $t > 0$.

Proof. In exercise.

$\boxed{\Rightarrow}$ Assume $X \sim \mathcal{E}(\lambda)$ and compute the hazard rate function of X.

Solution.

Recall the density function f of X equals for all $t > 0$, $f(t) = \lambda e^{-\lambda t}$, and the survival function is $\mathbb{P}(X > t) = e^{-\lambda t}$. Hence the hazard rate function of X equals

$$h(t) = \frac{f(t)}{\mathbb{P}(X > t)} = \lambda.$$

$\boxed{\Leftarrow}$ Assume the hazard rate function of X is constant equal to λ. Prove that $X \sim \mathcal{E}(\lambda)$ using Proposition 1.14 without any computation.

Solution.

By Proposition 1.14, the hazard rate characterizes the distribution. Hence a r.v. with constant hazard rate function equal to λ has the same hazard rate function, and thus the same distribution than an exponential r.v. (by $\boxed{\Rightarrow}$), that is $\mathcal{E}(\lambda)$.

□

V On the gamma and Erlang's distributions

The Erlang distribution generalizes the exponential distribution. It is commonly used to model the waiting time between more than one events. It was developed by A. K. Erlang to examine the number of telephone calls which might be made at the same time to the operators of the switching stations. More precisely, the Erlang distribution is the distribution of the sum of n i.i.d. r.v. with exponential distribution. The gamma distribution then generalizes the Erlang distribution by allowing n to be any positive real number.

Definition 1.16.

- **Gamma distribution.** Let $\lambda > 0$ and $\alpha > 0$. A non-negative real-valued random variable

X with density (w.r.t. the Lebesgue measure)

$$x \mapsto \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}\lambda e^{-\lambda x}\mathbb{1}_{\{x>0\}},$$

where $\Gamma(\alpha) = \int_0^{+\infty} e^{-t}t^{\alpha-1}dt$ is the gamma function, is said to have a gamma distribution with parameter (α, λ). We denote $X \sim \Gamma(\alpha, \lambda)$.

- **Erlang distribution.** If $\alpha = n \in \mathbb{N}^*$, then the gamma distribution $\Gamma(n, \lambda)$ is called Erlang's distribution. In particular, since $\Gamma(n) = (n-1)!$, the density of a r.v. $X \sim \Gamma(n, \lambda)$ equals

$$x \mapsto \frac{\lambda^n x^{n-1}}{(n-1)!}e^{-\lambda x}\mathbb{1}_{\{x>0\}}.$$

The gamma distribution can be characterized by its moment-generating function.

Proposition 1.17.

Let $\alpha, \lambda > 0$ and consider $X \sim \Gamma(\alpha, \lambda)$.

1. *The moment-generating function, or Laplace transform of X equals for all $t < \lambda$,*

$$\mathbb{E}\left[e^{tX}\right] = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}.$$

2. *We deduce that $\mathbb{E}[X] = \alpha/\lambda$ and $\mathrm{Var}(X) = \alpha/\lambda^2$.*

Proof. In exercise (Worksheet 1, Exercise 2). □

Proposition 1.18.

Links with the exponential and the Chi-square distributions.

i) If $X \sim \Gamma(\alpha, \lambda)$, then, for all $\mu > 0$, $\mu X \sim \Gamma(\alpha, \lambda/\mu)$.

ii) $\Gamma(1, \lambda) \sim \mathcal{E}(\lambda)$.

iii) If $X_1, \ldots, X_n$ are i.i.d. with distribution $\mathcal{E}(\lambda)$, then $\sum_{i=1}^{n} X_i \sim \Gamma(n, \lambda)$.

iv) $\Gamma(n, 1/2) \overset{(d)}{=} \chi^2(2n)$.

Proof. In exercise.

i) Use the m.g.f. to prove this point.

Solution.

For all $t > 0$, the m.g.f. of μX equals

$$\mathbb{E}\left[e^{t(\mu X)}\right] = \mathbb{E}\left[e^{(\mu t)X}\right] = \left(\frac{\lambda}{\lambda - \mu t}\right)^{\alpha} = \left(\frac{\lambda/\mu}{\lambda\mu - t}\right)^{\alpha}.$$

We recognize the m.g.f. of a $\Gamma(\alpha, \lambda/\mu)$ distribution.

ii) By taking $\alpha = 1$ in the definition, since $\Gamma(1) = 0! = 1$, we recognize the density of an exponential r.v. with parameter λ.

iii) Worksheet 1, Exercise 2.

iv) Immediate since the density of a $\chi^2(d)$ is

$$x \mapsto \frac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1} e^{-x/2} \mathbb{1}_{\{x>0\}}.$$

□