# Chapter 6

# Relationships between Random Variables

When describing and reasoning about random phenomena, we often need to exploit the relationship between two or more random variables that are defined in the same probability model. In this chapter, we discuss a few common scenarios in which different random variables can be related to one another in a deterministic fashion (e.g., one can be described as a function of others). Non-deterministic relationships between random variables (namely, statistical dependence and independence) will be discussed in a later chapter.

## 6.1 Functions of random variables

Suppose we have a random variable $Y$ that can be represented as a function of a simpler random variable $X$. More specifically, let $Y = h(X)$ where $h$ is a function mapping real numbers to real numbers. How is the distribution of $Y$ related to the distribution of $X$? Can we derive the distribution of $Y$ if we know the distribution of $X$? Can we compute expected value, variance, and other characteristics of $Y$?

**Example 6.1.1** (Points in a game)**.** In a game of chance, the number of points you gain is determined by rolling a die. If you get a 1, you lose 1 point; if you get a 5 or a 6, you gain 1 point; otherwise you neither gain nor lose any points.

Let $D$ denote the number shown on the die, and $G$ your gain. According to the description of the game,

$$G = \begin{cases} -1 & \text{if } D = 1, \\ +1 & \text{if } D \in \{5, 6\}, \\ 0 & \text{if } D \in \{2, 3, 4\}. \end{cases}$$

In other words, $G = h(D)$ where $h$ is the function given by[1]

$$h(x) = \begin{cases} -1 & \text{if } x = 1, \\ +1 & \text{if } x \in \{5, 6\}, \\ 0 & \text{if } x \in \{2, 3, 4\}. \end{cases}$$

Q What is the distribution of $G$?

A The possible values of $G$ are $-1$, $0$ and $1$. We have

$$\begin{aligned} \mathbb{P}(G = -1) &= \mathbb{P}(D = 1) = 1/6 , \\ \mathbb{P}(G = 0) &= \mathbb{P}(D \in \{2, 3, 4\}) = \mathbb{P}(D = 2) + \mathbb{P}(D = 3) + \mathbb{P}(D = 4) = 1/6 + 1/6 + 1/6 = 1/2 , \\ \mathbb{P}(G = 1) &= \mathbb{P}(D \in \{5, 6\}) = \mathbb{P}(D = 5) + \mathbb{P}(D = 6) = 1/6 + 1/6 = 1/3 . \end{aligned}$$

Hence, the pmf of $G$ is

$$p_G(y) = \begin{cases} 1/6 & \text{if } y = -1, \\ 1/2 & \text{if } y = 0, \\ 1/3 & \text{if } y = 1, \\ 0 & \text{otherwise.} \end{cases}$$

As a sanity check, let us note that $p_G(-1) + p_G(0) + p_G(1) = 1/6 + 1/2 + 1/3 = 1$, as it should be.

[1] Let us spell out the model more precisely: The set of possible outcomes is $\Omega := \{1, 2, 3, 4, 5, 6\}$ and the measure of probabilities is given by $\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = 1/6$. The two random variables $D$ and $G$ are functions on $\Omega$ defined by $D(a) := a$ and $G(a) := h(D(a))$ for each $a \in \Omega$.

Ⓠ What is the expected value of $G$?

A1 (using the distribution of $G$)
Since we already know the pmf of $G$, we can write

$$
\begin{aligned}
\mathbb{E}[G] &= \sum_g \mathbb{P}(G = g)g \\
&= \mathbb{P}(G = -1) \cdot (-1) + \mathbb{P}(G = 0) \cdot 0 + \mathbb{P}(G = 1) \cdot 1 \\
&= (1/6) \cdot (-1) + (1/2) \cdot 0 + (1/3) \cdot 1 \\
&= 1/6 \,.
\end{aligned}
$$

A2 (using the distribution of $D$)
As we saw earlier, in order to calculate the expected value of $G = h(D)$, we can directly use the pmf of $D$. Namely,

$$
\begin{aligned}
\mathbb{E}[G] &= \sum_d \mathbb{P}(D = d)h(d) \\
&= \mathbb{P}(D = 1) \cdot (-1) + \mathbb{P}(D = 2) \cdot 0 + \mathbb{P}(D = 3) \cdot 0 + \mathbb{P}(D = 4) \cdot 0 + \mathbb{P}(D = 5) \cdot 1 + \mathbb{P}(D = 6) \cdot 1 \\
&= (1/6) \cdot (-1) + (1/6) \cdot 0 + (1/6) \cdot 0 + (1/6) \cdot 0 + (1/6) \cdot 1 + (1/6) \cdot 1 \\
&= 1/6 \,.
\end{aligned}
$$

The advantage of this approach is that it does not require deriving the pmf of $G$. ○

**Example 6.1.2** (Square root of an exponential RV). Let $T$ be an exponential random variable with rate $\lambda$, and suppose $S := \sqrt{T}$.

Ⓠ What is the distribution of $S$?

A Let us start by finding the cdf of $S$. The possible values of $S$ are all the non-negative numbers, hence $F_S(x) = 0$ if $x < 0$. For $x \geq 0$, we have

$$
F_S(x) = \mathbb{P}(S \leq x) = \mathbb{P}(\sqrt{T} \leq x) = \mathbb{P}(T \leq x^2) = 1 - \mathrm{e}^{-\lambda x^2} \,.
$$

(*Warning*. Note that the latter computation is not valid if $x < 0$, because in that case, $\sqrt{T} \leq x$ is impossible while $T \leq x^2$ is not.) Put together, we get

$$
F_S(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x^2} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}
$$

To find the pdf of $S$, we differentiate $F_S(x)$:

$$
f_S(x) = \begin{cases} 2\lambda x \mathrm{e}^{-\lambda x^2} & \text{if } x > 0, \\ 0 & \text{if } x < 0. \end{cases}
$$

Let us note that $F_S(x)$ is not differentiable at $x = 0$, hence $f_S(x)$ is not defined at $x = 0$.

Ⓠ What is expected value of $S$?

A1 (using the pdf of $S$)
Since we have already derived the pdf of $S$, we can write

$$
\begin{aligned}
\mathbb{E}[S] &= \int_{-\infty}^{\infty} x f_S(x)\,\mathrm{d}x \\
&= \int_0^{\infty} 2\lambda x^2 \mathrm{e}^{-\lambda x^2} \\
&= \overset{0}{\cancel{-x\mathrm{e}^{-\lambda x^2}\Big|_0^{\infty}}} + \int_0^{\infty} \mathrm{e}^{-\lambda x^2}\,\mathrm{d}x \,. && \left(\text{integration by parts } \begin{cases} u(x) := x \\ v(x) := -\mathrm{e}^{-\lambda x^2} \end{cases}\right)
\end{aligned}
$$

To compute the remaining integral, we can proceed as in the exercise after Example 5.2.3. Alternatively, we can observe that $\mathrm{e}^{-\lambda x^2}$ resembles the pdf of the normal distribution. More specifically,

$$
\begin{aligned}
\mathbb{E}[S] &= \int_0^{\infty} \mathrm{e}^{-\lambda x^2}\,\mathrm{d}x \\
&= \sqrt{\frac{\pi}{\lambda}} \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}\,\mathrm{d}z && \left(\text{change of variable } x = \tfrac{1}{\sqrt{2\lambda}}z\right) \\
&= \tfrac{1}{2}\sqrt{\tfrac{\pi}{\lambda}} \,,
\end{aligned}
$$

where the latter follows once we recall that $\varphi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ (the pdf of the standard normal distribution) is symmetric with $\int_{-\infty}^{\infty} \varphi(z)\, dz = 1$.

A2 (using the pdf of $T$)
Since $S$ is a function of $T$, we can directly calculate the expected value of $S$ using the pdf of $T$.

$$\begin{aligned}\mathbb{E}[S] &= \int_{-\infty}^{\infty} \sqrt{t} f_T(t)\, dt \\ &= \int_0^{\infty} \sqrt{t} \lambda e^{-\lambda t}\, dt \\ &= \int_0^{\infty} 2\lambda x^2 e^{-\lambda x^2}\, dx && \text{(change of variable } t = x^2,\ x > 0)\end{aligned}$$

which is the same integral as in the previous answer. ○

**Example 6.1.3** (Normal distribution)**.** In this example, we derive the pdf of the general normal distribution based on the pdf of the standard normal distribution introduced in the previous chapter.

Recall from Section 5.3 that a normal random variable with mean $\mu$ and standard deviation $\sigma > 0$ is a random variable of the form $X = \sigma Z + \mu$ where $Z$ is a standard normal random variable.

Q What is the pdf of $X$?

A Let us start by deriving the cdf of $X$ in terms of the cdf of the standard normal distribution $\Phi$. By definition, the cdf of $X$ at a given value $x \in \mathbb{R}$ is

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(\sigma Z + \mu \le x) = \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Differentiating with respect to $x$, we obtain

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Figure 6.1 depicts the graph of the pdf of $X$ in the case $\mu = 3$ and $\sigma = 2$. As expected, the pdf of $X$ has the same shape as the pdf of $Z$, except it has a different location and a different scale.

Q What is the maximum value of $f_X(x)$?

A1 In the derived expression above, the maximum is clearly achieved when $x = \mu$. We have $f_X(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$.

A2 Compared to the pdf of $Z$, the pdf of $X$ is horizontally scaled up by a factor $\sigma$. In order for the total area under the pdf to remain 1, the pdf of $X$ must be vertically scaled down by a factor $\sigma$. Hence, the maximum value of $f_X(x)$ is $1/\sigma$ the maximum value of $\varphi(z)$, that is, $\frac{1}{\sigma\sqrt{2\pi}}$.

○

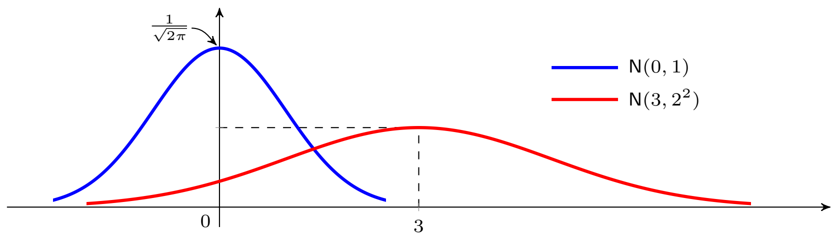



Figure 6.1: The pdfs of two normal distributions.

**Example 6.1.4** (Chi-squared distribution with 1 degree of freedom)**.** Consider a random variable of the form $X := Z^2$, where $Z \sim \mathsf{N}(0, 1)$ is a standard normal random variable. Random variables of this type often show up in statistical contexts, and hence deserve a name. We call $X$ a *chi-squared random variable* with 1 *degree of freedom*. A more general concept of a chi-squared distribution will be discussed in the following section.

Q What is the pdf of $X$?

A Let us start with deriving the cdf of $X$.

Observe that $X$ takes only non-negative values, hence $F_X(x) = 0$ for $x < 0$. For $x \geq 0$, we have

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(Z^2 \leq x) = \mathbb{P}\big( -\sqrt{x} \leq Z \leq \sqrt{x}\big)$$
$$= \mathbb{P}\left(Z \leq \sqrt{x}\right) - \mathbb{P}\left(Z < \sqrt{x}\right) = \Phi\big(\sqrt{x}\big) - \Phi\big( -\sqrt{x}\big) \ .$$

where $\Phi$ denotes the cdf of the standard normal distribution. Let us note that for the last equality, we have used the fact that $\Phi$ is continuous.

Differentiating with respect to $x$, we obtain

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \dfrac{1}{2\sqrt{x}}\varphi(\sqrt{x}) + \dfrac{1}{2\sqrt{x}}\varphi(-\sqrt{x}) & \text{if } x > 0. \end{cases}$$
$$= \begin{cases} 0 & \text{if } x < 0, \\ \dfrac{1}{\sqrt{2\pi x}} \mathrm{e}^{-x/2} & \text{if } x > 0, \end{cases}$$

where, as usual, $\varphi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}$ denotes the pdf of the standard normal distribution.

Q What is the expected value of $X$?

A Since $X = Z^2$, we have

$$\mathbb{E}[X] = \mathbb{E}[Z^2] = \mathbb{V}\mathrm{ar}[Z] = 1 \ .$$

○

*Exercise.* Find the variance of a chi-squared random variable with 1 degree of freedom. (*Hint:* The answer is 2.)

While a function of a discrete random variable is necessarily a discrete random variable, a function of continuous random variable may or may not be continuous.

**Example 6.1.5** (Bernoulli as a function of uniform)**.** Let $U$ be a continuous random variable uniformly distributed over the interval $[0, 1]$. In other words, suppose that $U$ has pdf

$$f_U(x) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $p \in [0, 1]$ be an arbitrary number. Define a new random variable $X$, where $X = 0$ if $U < 1 - p$ and $X = 1$ if $U \geq 1 - p$.

Q What is the distribution of $X$?

A Bernoulli with $\mathbb{P}(X = 0) = 1 - p$ and $\mathbb{P}(X = 1) = 1 - (1 - p) = p$. ○

It turns out that *every* random variable can be expressed as a function of a uniformly distributed continuous random variable. In computer simulations, this can be used to generate (pseudo-)random numbers with any prescribed distribution.

*Exercise.* Let $U$ be a random variable that is uniformly distributed over $[0, 1]$. Define a new random variable $X$ as a function of $U$ such that $\mathbb{P}(X = -1) = 1/6$, $\mathbb{P}(X = 0) = 1/2$ and $\mathbb{P}(X = 1) = 1/3$.

## 6.2 Sums of independent random variables

In this section, we consider the scenario in which a random variable $Y$ can be represented as a sum of two or more simpler random variables, that is, $Y = X_1 + X_2 + \cdots + X_n$. In Chapter 4, we saw how such a representation can help us compute the expected value or the variance of $Y$ (Examples 4.4.1, 4.4.2 and 4.4.4). Here, we focus on the case in which $X_1, X_2, \ldots, X_n$ are independent, and discuss how the distribution of $Y$ is related to the distribution of $X_1, X_2, \ldots, X_n$. This will also be an excuse to introduce some important families of distributions.

**Example 6.2.1** (Sum of i.i.d. geometric RVs)**.** Let $N_1$ and $N_2$ be two independent geometric random variables with parameter $p$, and let $M := N_1 + N_2$.

Ⓠ What are the possible values of $M$?

A $2, 3, 4, \ldots$.

Ⓠ What is the distribution of $M$?

A1 (conceptual approach)
The random variable $M$ resembles (but is not exactly the same as) a negative binomial random variable with parameters 2 and $p$ (see Example 4.4.4).

Indeed, we can simulate $N_1$, $N_2$ and $M$ using a coin with bias parameter $p$. Suppose we repeat flipping the coin until two tails come up. Let $N_1'$ denote the number of flips until the first head, and $N_2'$ denote the number of flips after the first head until the second head, and define $M' := N_1' + N_2'$.

$$\underbrace{\overbrace{\texttt{T T T T T H}}^{N_1'}\overbrace{\texttt{T T T H}}^{N_2'}}_{M'}$$

Clearly,

- $N_1'$ has the same distribution as $N_1$,
- $N_2'$ has the same distribution as $N_2$,
- $N_1'$ and $N_2'$ are independent, just like $N_1$ and $N_2$.

Therefore, $M'$ has the same distribution as $M$.

Note that $M'$ is the number of flips until the second head comes up. The number of *tails* until the second head (i.e., $M' - 2$) is, by definition, a negative binomial random variable with parameters 2 and $p$. Thus, we find the pmf of $M$ to be

$$\begin{aligned} p_M(k) &:= \mathbb{P}(M = k) = \mathbb{P}(M' = k) \\ &= \mathbb{P}(M' - 2 = k - 2) = \begin{cases} \binom{k-1}{1} p^2 (1-p)^{k-2} & \text{if } k = 2, 3, \ldots, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

A2 (computational approach)
To compute $\mathbb{P}(M = k)$, we can use the principle of total probability by breaking the possibilities based on the value of $N_1$ (or $N_2$). Namely, for every $k$, we can write

$$\begin{aligned} \mathbb{P}(M = k) &= \sum_{\ell} \mathbb{P}(N_1 = \ell \text{ and } M = k) \\ &= \sum_{\ell} \mathbb{P}(N_1 = \ell \text{ and } N_2 = k - \ell) && \text{(since } M = N_1 + N_2\text{)} \\ &= \sum_{\ell} \mathbb{P}(N_1 = \ell)\, \mathbb{P}(N_2 = k - \ell)\,. && \text{(since } N_1 \text{ and } N_2 \text{ are independent)} \end{aligned}$$

Since the possible values of $N_1$ and $N_2$ are $1, 2, \ldots$, we have

- $\mathbb{P}(N_1 = \ell) = 0$ unless $\ell$ is an integer with $\ell \geq 1$,
- $\mathbb{P}(N_2 = k - \ell) = 0$ unless $k - \ell$ is an integer with $k - \ell \geq 1$, that is, $\ell \leq k - 1$.

Thus, for $k = 2, 3, \ldots$, we have

$$\begin{aligned} \mathbb{P}(M = k) &= \sum_{\ell=1}^{k-1} \mathbb{P}(N_1 = \ell)\, \mathbb{P}(N_2 = k - \ell) \\ &= \sum_{\ell=1}^{k-1} \underbrace{(1-p)^{\ell-1} p \times (1-p)^{k-\ell-1} p}_{(1-p)^{k-2} p^2} \\ &= (k-1)(1-p)^{k-2} p^2\,. \end{aligned}$$

In summary, the pmf of $M$ is given by

$$p_M(k) := \mathbb{P}(M = k) = \begin{cases} (k-1)(1-p)^{k-2} p^2 & \text{if } k = 2, 3, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

○

The idea of using the principle of total probability to write the pmf of $M = N_1 + N_2$ in terms of the pmfs of $N_1$ and $N_2$ is not specific to the above example. Let us spell it out in general terms.

**Distribution of sum of independent RVs (discrete case).** Let $X$ and $Y$ be two independent discrete random variables with pmfs $p_X$ and $p_Y$. Then, using the principle of total probability as in the above example, the pmf of $Z := X + Y$ can be written as

$$p_Z(k) = \mathbb{P}(Z = k) = \sum_i \mathbb{P}(X = i)\,\mathbb{P}(Y = k - i) = \sum_i p_X(i) p_Y(k - i)$$

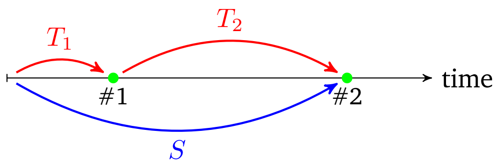
The above sum is often called the *discrete convolution* of the two integer functions $p_X$ and $p_Y$. The summation runs over the values for which $p_X(i)$ and $p_Y(k - i)$ are non-zero.

**Example 6.2.2** (Sum of i.i.d. exponential RVs)**.** Let $T_1$ and $T_2$ be two independent exponential random variables with rate $\lambda$, and consider $S := T_1 + T_2$.

Recall that a exponential random variable can be a reasonable model for the waiting time till the arrival of the next spam email in your mailbox.

Q In the latter context, what could be the interpretation of the sum of two independent exponential random variables with rate $\lambda$?

A The waiting time till the arrival of two spam emails in your mailbox.




Namely, think of $T_1$ as the waiting time from now till the arrival of the first spam email, and $T_2$ as the waiting time between the first two spam emails. Then, $S = T_1 + T_2$ is simply the waiting time till the arrival of the second spam email.

Q What are the possible values of $S$?

A All non-negative real numbers.

Q What is the distribution of $S$?

A In analogy with the convolution formula for the pmf of the sum of two independent discrete random variables, we may guess that the pdf of $S$ satisfies

$$f_S(x) = \int_t f_{T_1}(t) f_{T_2}(x - t)\,\mathrm{d}t\,.$$

This is indeed true, although it takes more to justify it than a simply application of the law of total probability (see below).

Since $T_1$ and $T_2$ can only take non-negative values, we have

- $f_{T_1}(t) = 0$ unless $t \geq 0$,
- $f_{T_2}(x - t) = 0$ unless $x - t \geq 0$, that is, $t \leq x$.

Therefore, for $x \geq 0$, we have

$$\begin{aligned} f_S(x) &= \int_{t=0}^{x} f_{T_1}(t) f_{T_2}(x - t)\,\mathrm{d}t \\ &= \int_{t=0}^{x} \underbrace{\lambda \mathrm{e}^{-\lambda t} \times \lambda \mathrm{e}^{-\lambda(x-t)}}_{\lambda^2 \mathrm{e}^{-\lambda x}}\,\mathrm{d}t \\ &= \int_{t=0}^{x} \lambda^2 \mathrm{e}^{-\lambda x}\,\mathrm{d}t \\ &= \lambda^2 x \mathrm{e}^{-\lambda x}\,. \end{aligned}$$

Altogether, we obtain

$$f_S(x) = \boxed{\begin{cases} \lambda^2 x \mathrm{e}^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}}$$

Due to its natural interpretation, the distribution of $S$ deserves a name: it is called the *gamma distribution* with parameters 2 and $\lambda$. More generally, the distribution of the sum of $r$ independent exponential random variables with rate $\lambda$ is referred to as the *gamma distribution* with parameters $r$ (the *shape* parameters) and $\lambda$ (the *rate*). As an example, if the average rate of arrivals of spam emails in your mailbox is $\lambda$, then the arrival time of the $r$'th spam email in your mailbox is a gamma random variable with shape $r$ and rate $\lambda$.

Repeating the above computation recursively, one can find the pdf of the gamma distribution with shape $r$ (a positive integer) and rate $\lambda > 0$ to be

$$f_S(x) = \begin{cases} \frac{\lambda^r}{(r-1)!} x^{r-1} \mathrm{e}^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

○

*Exercise.* Derive the above formula for the pdf of the gamma distribution.

**Distribution of sum of independent RVs (continuous case).** Let $X$ and $Y$ be two <u>independent</u> <u>continuous</u> random variables with pdfs $f_X$ and $f_Y$. Then, the pdf of $Z := X + Y$ satisfies

$$f_Z(z) = \int_x f_X(x) f_Y(z - x)\,\mathrm{d}x \;.$$

The integral is called the *convolution* of $f_X$ and $f_Y$.

Skip on first read!

Ⓠ What is the justification for the above identity?

A1 (intuitive but non-rigorous)

We would like to use the principle of total probability as in the discrete case. However, there are two obstacles to this: (1) $f_Z(z)$, $f_X(x)$ and $f_Y(z-x)$ indicate probability densities rather than probabilities, and (2) the possible values of $X$ are uncountable. To circumvent these, we use approximations that can be made arbitrarily sharp.

First, let $\Delta z > 0$ be small. Then

$$f_Z(z)\Delta z \approx \mathbb{P}(z \leq Z \leq z + \Delta z) \tag{⌘}$$

(see Figure 6.2a). Next, let $\Delta x > 0$ be small, and divide the possible values of $X$ into intervals $[x_i, x_{i+1})$ of length $\Delta x$ (see Figure 6.2b). We now partition the outcome space based on which tiny interval $X$ belongs to, and apply the principle of total probability to write

$$\mathbb{P}(z \leq Z \leq z + \Delta z) = \sum_{i=-\infty}^{\infty} \mathbb{P}(x_i \leq X < x_i + \Delta x \text{ and } z \leq Z \leq z + \Delta z) \;. \tag{⌘}$$

If we choose $\Delta x$ to be much smaller than $\Delta z$, then the condition

$$\big\langle x_i \leq X < x_i + \Delta x \text{ and } z \leq Z \leq z + \Delta z \big\rangle$$

becomes almost equivalent to

$$\big\langle x_i \leq X < x_i + \Delta x \text{ and } z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i \big\rangle \;.$$

Indeed, if $x_i \leq X < x_i + \Delta x$ and $z \leq Z \leq z + \Delta z$, then $z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i$, and conversely, if $x_i \leq X < x_i + \Delta x$ and $z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i$, then $z - \Delta x < Z < z + \Delta z + \Delta x$, which is not much different from $z \leq Z \leq z + \Delta z$. Hence,

$$\begin{aligned} &\mathbb{P}(x_i \leq X < x_i + \Delta x \text{ and } z \leq Z \leq z + \Delta z) \\ &\qquad \approx \mathbb{P}(x_i \leq X < x_i + \Delta x \text{ and } z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i) \;. \end{aligned} \tag{⌘}$$

The smaller $\Delta x$, the better this approximation. Since $X$ and $Y$ are independent,

$$\begin{aligned} &\mathbb{P}(x_i \leq X < x_i + \Delta x \text{ and } z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i) \\ &\qquad = \mathbb{P}(x_i \leq X < x_i + \Delta x)\,\mathbb{P}(z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i) \end{aligned} \tag{⌘}$$

Note that, since $\Delta x$ and $\Delta z$ are small,

$$\mathbb{P}(x_i \leq X < x_i + \Delta x) \approx f_X(x_i)\Delta x \;, \tag{⌘}$$

$$\mathbb{P}(z - x_i - \Delta x \leq Y \leq z + \Delta z - x_i) \approx f_Y(z - x_i)(\Delta z + \Delta x) \;. \tag{⌘}$$

Combining the equations tagged with (⌘), we find the approximation

$$
\begin{aligned}
f_Z(z)\Delta z &\approx \sum_{i=-\infty}^{\infty} f_X(x_i)\Delta x f_Y(z-x_i)(\Delta z+\Delta x) \\
&= (\Delta z+\Delta x)\underbrace{\sum_{i=-\infty}^{\infty} f_X(x_i) f_Y(z-x_i)\Delta x}_{◈} \,.
\end{aligned}
$$

Note that ◈ resembles a Riemann sum for the integral $\int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x$.[2] Hence, it is plausible (and indeed true) that

$$
f_Z(z)\Delta z \approx (\Delta z+\Delta x)\int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x \,.
$$

Sending first $\Delta x \to 0$ and then $\Delta z \to 0$, all the approximations sharpen, and in the limit, we obtain

$$
f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x \,,
$$

as claimed.

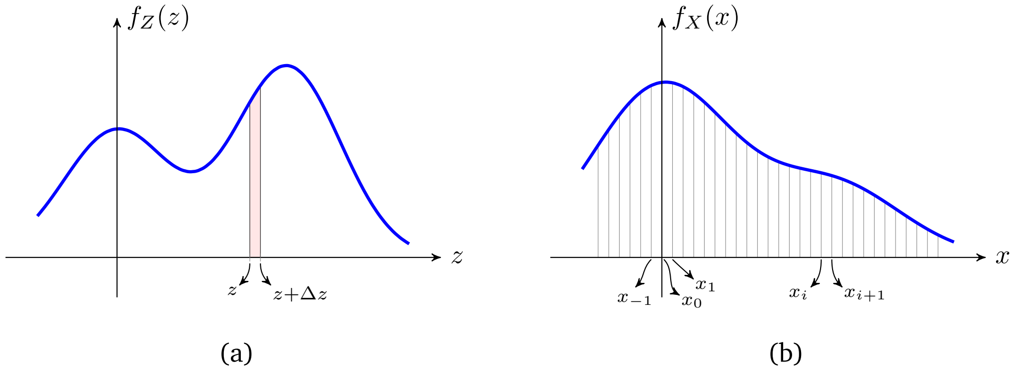



Figure 6.2: Illustration of the intuitive argument for the convolution formula. (a) The pdf of $Z$ at a point $z$ is approximately $\mathbb{P}(z \le Z \le z+\Delta z)/\Delta z$. (b) The possible values of $X$ are partitioned into small intervals of length $\Delta x$. Here, $x_{i+1}-x_i = \Delta x$ for every $i$.

$\boxed{\text{A2}}$ (rigorous but less intuitive)
We postpone this till Chapter 8, where we see a calculus-based argument for a more general version of the claimed identity.

[To be moved to Chapter 8:] Let us start with the cdf of $Z$. For every $z$, we have

$$
\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \le z) \\
&= \mathbb{P}(X+Y \le z) \\
&= \iint_{x+y\le z} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{x=-\infty}^{\infty}\left(\int_{y=-\infty}^{z-x} f_{X,Y}(x,y)\,\mathrm{d}y\right)\mathrm{d}x && \text{(Fubini's theorem)} \\
&= \int_{x=-\infty}^{\infty}\left(\int_{w=-\infty}^{z} f_{X,Y}(x,w-x)\,\mathrm{d}w\right)\mathrm{d}x && \text{(change of variable } w=y+x\text{)} \\
&= \int_{w=-\infty}^{z}\left(\int_{x=-\infty}^{\infty} f_{X,Y}(x,w-x)\,\mathrm{d}x\right)\mathrm{d}w && \text{(Fubini's theorem)} \qquad (☆)
\end{aligned}
$$

[2] Recall that Riemann sums are finite sums and are only meant to make sense of integrals on bounded closed intervals. Here, we have an infinite sum and an improper integral.

Now, using the fundamental theorem of calculus, we conclude that

$$f_Z(z) = F_Z'(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\,\mathrm{d}x$$

at every point $z$ at which $z \mapsto \displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,\mathrm{d}x$ is continuous.[3]

**Example 6.2.3** (Sum of independent normal RVs)**.** Let $Y = X_1 + X_2$ where $X_1$ and $X_2$ are two independent random variables with $X_1 \sim \mathsf{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$.

Q What are the expected value and the variance of $Y$?

A We have $\mathbb{E}[Y] = \mu_1 + \mu_2$ (by the linearity of expectation) and $\mathbb{V}\mathrm{ar}[Y] = \sigma_1^2 + \sigma_2^2$ (since $X_1$ and $X_2$ are independent).

Q What is the pdf of $Y$?

A1 (tedious integration)
As discussed above, the pdf of $Y$ is the convolution of the pdfs of $X_1$ and $X_2$. From Example 6.1.3, we know that

$$f_{X_1}(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \qquad f_{X_2}(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

Computing the convolution of these two functions, one can find that

$$f_Y(y) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \mathrm{e}^{-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}. \qquad (📖)$$

A2 (geometric approach)
We postpone this till Chapter 8.

Q From its pdf, do you recognize the distribution of $Y$?

A The function in (📖) is the pdf of the normal distribution with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. ○

*Exercise.* In the above example, derive the pdf of $Y$ by computing the convolution integral.

The above example highlights an important property of the normal distribution.

**Stability of the normal distribution.** Every non-degenerate linear combination of independent normal random variables is again a normal random variable. Specifically,

- If $X$ is a normal random variable and $a, b \in \mathbb{R}$ with $a \neq 0$, then $aX + b$ is again a normal random variable.
- If $X$ and $Y$ are independent normal random variables, then $X + Y$ is again a normal random variable.

**Example 6.2.4** (Chi-squared distribution)**.** Let $r$ be a positive integer, and consider the random variable $Y = Z_1^2 + Z_2^2 + \cdots + Z_r^2$, where $Z_1, Z_2, \ldots, Z_r$ are independent standard normal random variables. Such a random variable $Y$ is called a *chi-squared random variable* with $r$ *degrees of freedom*. The chi-squared distribution with $r$ degrees of freedom is often denoted by $\chi^2(r)$.[4]

We discussed the case $r = 1$ in Example 6.1.4. Let us now consider the case $r = 2$.

Q What is the pdf of $Y$ when $r = 2$?

A We can write $Y = X_1 + X_2$ where $X_1 := Z_1^2$ and $X_2 := Z_2^2$ are independent chi-squared random variables with $1$ degree of freedom. From Example 6.1.4, we know that $X_1$ and $X_2$ have the following pdf:

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \dfrac{1}{\sqrt{2\pi x}} \mathrm{e}^{-x/2} & \text{if } x > 0. \end{cases}$$

[3] To be precise, the fundamental theorem of calculus concerns proper integrals, whereas the outer integral on the right-hand side of (☆) is improper. However, the extension of the fundamental theorem of calculus to this type of improper integrals is straightforward. Namely, fix $z_0$ and write $\int_{v=-\infty}^{z} = \int_{v=-\infty}^{z_0} + \int_{v=z_0}^{z}$. The complete argument is left to the mathematically conscientious readers.

[4] Note: $\chi$ is the Greek letter "chi", not to be confused with the Latin letter $X$.

The pdf of $Y$ is thus the convolution of the latter function with itself.

Since $Y$ is always non-negative, we have $f_Y(y) = 0$ for every $y < 0$. For $y \geq 0$, we have

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_X(x) f_X(y-x)\,\mathrm{d}x \\
&= \int_0^y f_X(x) f_X(y-z)\,\mathrm{d}x && \text{(why?)} \\
&= \frac{1}{2\pi} \int_0^y \frac{1}{\sqrt{x(y-x)}} \mathrm{e}^{-x/2} \mathrm{e}^{-(y-x)/2}\,\mathrm{d}x \\
&= \frac{1}{2\pi} \mathrm{e}^{-y/2} \underbrace{\int_0^y \frac{1}{\sqrt{x(y-x)}}\,\mathrm{d}z}_{\pi} && \text{(standard integration)} \\
&= \frac{1}{2} \mathrm{e}^{-y/2}\,.
\end{aligned}$$

In summary,

$$f_Y(y) = \begin{cases} \frac{1}{2}\mathrm{e}^{-y/2} & \text{if } y \geq 0, \\ 0 & \text{if } y < 0. \end{cases}$$

Q Do you recognize the latter pdf?

A It is the pdf of the exponential distribution with rate $1/2$.

You may find it curious that $\chi^2(r=2)$ is nothing but the exponential distribution with rate $1/2$. Iterating the convolution recursively, one can find the pdf of the chi-squared distribution for $r > 2$. The general formula for the pdf of $\chi^2(r)$ turns out to be

$$f_Y(y) = \begin{cases} \langle\text{some constant}\rangle\, y^{r/2-1}\mathrm{e}^{-y/2} & \text{if } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The unspecified constant is such that the integral of $f_Y$ is $1$.

Let us next discuss a geometric interpretation of the chi-squared distribution. For simplicity, let us focus on the case $r = 2$. As before, let $Z_1$ and $Z_2$ be two independent standard normal random variables. Consider a point $Q$ on the plane with coordinates $Z_1$ and $Z_2$. Let $D$ denote the (random) distance between $Q$ and the origin.

Q What is the distribution of $D$?

A By the Pythagorean theorem, $D^2 = Z_1^2 + Z_2^2$. Thus, $D^2$ has the chi-squared distribution with $2$ degrees of freedom. As we saw above, the chi-squared distribution with $2$ degrees of freedom happens to be the same as the exponential distribution with rate $1/2$. Recall Example 6.1.2 in which we derived the pdf of the square root of an exponential random variable with rate $\lambda$. Setting $\lambda = 1/2$, we obtain

$$f_D(x) = \begin{cases} x\mathrm{e}^{-x^2/2} & \text{if } x > 0, \\ 0 & \text{if } x < 0. \end{cases}$$

For general $r$, the chi-squared distribution can be interpreted similarly. If $Q$ is a point in the $r$ dimensional Euclidean space whose coordinates are independent standard normal random variables, then the square of the distance between $Q$ and the origin has the $\chi^2(r)$ distribution. ○

**Gamma vs. chi-squared distribution.** You may have noticed a similarity between the pdf of a gamma distribution and the pdf of a chi-squared distribution:

$$\mathsf{Gamma}(s,\lambda): \qquad f(x) = \begin{cases} \langle\text{some constant}\rangle\, x^{s-1}\mathrm{e}^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$\chi^2(r): \qquad f(x) = \begin{cases} \langle\text{some constant}\rangle\, x^{r/2-1}\mathrm{e}^{-x/2} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

| Bernoulli process | $\longleftrightarrow$ | Poisson process |
|---|---|---|
| geometric distribution | $\longleftrightarrow$ | exponential distribution |
| negative binomial distribution | $\longleftrightarrow$ | gamma distribution |
| binomial distribution | $\longleftrightarrow$ | ? |

Table 6.1: The analogy between Bernoulli and Poisson processes.

From the pdfs, one can observe that, when $r$ is even, $\chi^2(r) = \mathsf{Gamma}(r/2, 1/2)$, that is, the chi-squared distribution with $r$ degrees of freedom is the same as the gamma distribution with shape $r/2$ and rate $1/2$. This is consistent with the earlier observation that the chi-squared distribution with 2 degrees of freedom is the same as the exponential distribution with rate $1/2$ (Example 6.2.4).

Does the identity $\chi^2(r) = \mathsf{Gamma}(r/2, 1/2)$ still hold when $r$ is odd? Our earlier definition of $\mathsf{Gamma}(s, \lambda)$ as the distribution of the sum of $s$ independent $\mathsf{Exp}(\lambda)$ random variables does not give a meaning to gamma distributions with non-integer shapes. Nevertheless, given arbitrary real numbers $\alpha > 0$ and $\lambda > 0$, we can consider a pdf of the form

$$f(x) = \begin{cases} \langle\text{some constant}\rangle\, x^{\alpha-1}\mathrm{e}^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where the constant is chosen such that the integral of $f$ is 1. By extension, the distribution defined by the latter pdf is called the *gamma distribution* with *shape* $\alpha$ and *rate* $r$, and is denoted by $\mathsf{Gamma}(\alpha, \lambda)$. With this extended notion of gamma distribution, the identity $\chi^2(r) = \mathsf{Gamma}(r/2, 1/2)$ holds for every positive integer $r$.

*Exercise.* Recall that, for a positive integer $r$, the constant in front of the pdf of $\mathsf{Gamma}(r, \lambda)$ is $\frac{\lambda^r}{(r-1)!}$. The constant in front of the pdf of $\chi^2(r)$ can be shown to be $\frac{1}{2^{r/2}\Gamma(r/2)}$, where $\Gamma(\cdot)$ is the gamma function. Find an expression for the constant in front of the pdf of $\mathsf{Gamma}(\alpha, \lambda)$ for arbitrary $\alpha > 0$ in terms of the gamma function. Verify that your expression is consistent with the two special cases.

## 6.3 Mixtures of random variables

`[To be written!]`

*Exercise.* Let $X$ be a continuous RV with cdf $F$ and $B$ be a Bernoulli RV with parameter $p$, and suppose that $X$ and $B$ are independent. Derive a formula for the distribution of $Y = B + X$ in terms of $F$ and $p$.

## 6.4 Bernoulli and Poisson processes

You may have noticed that there is an analogy between some discrete random variables and some continuous random variables (see Table 6.1). On the one side, we have random variables that can be defined in terms of a coin-flipping experiment: geometric, negative binomial and binomial random variables. On the other side, we have exponential, gamma, and perhaps other types of random variables. Exponential random variables resemble geometric random variables: both can be interpreted as the time until some event happens, and both are "memoryless". Likewise, gamma random variables are analogous to negative binomial random variables: a gamma random variable can be defined as a sum of independent exponential random variables the same way a negative binomial random variable can (almost) be thought of as a sum of independent geometric random variables.

Q In this analogy, which distribution corresponds to the binomial distribution?

What ties all these distributions together and clarifies the analogy is the concepts of Bernoulli and Poisson processes. A Bernoulli process is the mathematical model of a coin-flipping experiment. A Poisson process is a "continuous-time" analog of a Bernoulli process as we shall clarify shortly.

**Bernoulli process.** A *Bernoulli trial* refers to an elementary random experiment with two possible outcomes: "success" and "failure." The paradigmatic example of a Bernoulli trial is the experiment of flipping a coin, in which we can declare *heads* as "success" and *tails* as "failure", or vice versa. An infinite sequence of independent Bernoulli trials all with the same probability of success is called a *Bernoulli process* (Figure 6.3).

A variety of different experiments and phenomena can be modeled as or in terms of Bernoulli processes. Here are a few examples:

- We repeatedly flip a coin.

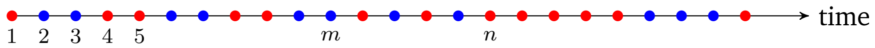



Figure 6.3: A realization of a Bernoulli process with success probability $1/2$. Blue indicates "success" and red indicates "failure."

- We send a sequence of information bits through a noisy communication channel. The fate of each bit of information can be thought of as a Bernoulli trial: either the bit is communicated correctly ("success") or is corrupted by noise ("failure").
- You and your friend repeatedly play a game of chance. You winning a game could count as "success" and you losing a game as "failure."

Mathematically, a Bernoulli process can be described by a sequence of independent Bernoulli random variables

$$X_1, X_2, X_3, \ldots$$

all with the same parameter $p$. In this formulation, $X_i = 1$ indicates that the $i$-th trial has been successful and $X_i = 0$ indicates that the $i$-th trial has been unsuccessful.

Observe that the geometric, negative binomial, and binomial distributions appear naturally in the context of a Bernoulli process. Let $p$ denote the probability of success in each trial.

Q What is the distribution of the number of trials until the first success?

A Geometric with parameter $p$.

More generally, starting from any trial, the number of trials until the next success is geometric.

Q Let $r \geq 1$ be an integer. What is the distribution of the number of failures until there are $r$ successful trials?

A Negative binomial with parameters $r$ and $p$.

Q Let $n \geq m \geq 1$ be integers. What is the distribution of the number of successful trials between trial number $m$ and trial number $n$?

A Binomial with parameters $n - m + 1$ and $p$.

**Poisson process.** A *Poisson process* is a model of "events" occurring over time independently of one another at a certain average rate (Figure 6.4). For instance, the arrivals of spam emails in your mailbox can be modeled as a Poisson process: spam emails are virtually independent of one another, yet arrive at more or less constant rate. Here are a few examples of other processes that can be modeled as a Poisson process:

- Arrival of unique visitors at a popular website. An "event" means the arrival of a visitor.
- Arrival of data packets at a server of a computer network. An "event" means the arrival of a data packet.
- The clicks of a Geiger counter.[5] An "event" means a clicking sound on the device, which in turn corresponds a burst of radiation (alpha particles, beta particles, or gamma rays) hitting the detector of the Geiger counter.

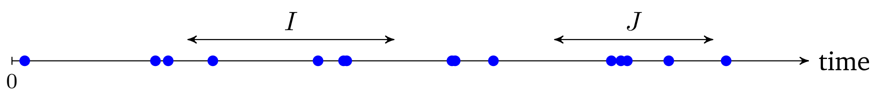



Figure 6.4: A realization of a Poisson process. Blue dots indicate arrival times of the "events."

A Poisson process can be thought of as a "continuous-time" analog of a Bernoulli process, where "events" correspond to "successful" trials. However, the mathematical formulation of a Poisson process is more cumbersome. Indeed, one has to resort to an indirect formulation by imitating the key aspects of a Bernoulli process.

A Poisson process with rate $\lambda > 0$ can indirectly be formulated as follows (see Figure 6.4).

(P1) Each time interval $I$ contains a random number of *events*, which we denote by $N(I)$.

[5] A Geiger counter is an instrument used for detecting ionized radiations. The portable Geiger counters often make a clicking sound upon detection of each ionization event.

(P2) The expected number of events occurring in a time interval $I$ is $\mathbb{E}[N(I)] = \lambda \times \text{length}(I)$. This is a different way to say that the average number of events per unit time is $\lambda$.

(P3) The number of events occurring in disjoint time intervals are independent. Namely, if $I$ and $J$ are two disjoint time intervals, then the two random variables $N(I)$ and $N(J)$ are independent. More generally, if $I_1, I_2, \ldots, I_k$ are disjoint time intervals, then the random variables $N(I_1), N(I_2), \ldots, N(I_k)$ are independent.

Observe that a Bernoulli process has properties analogous to the above three properties.

Consider a Poisson process with rate $\lambda$.

(Q) What is the distribution of the arrival time of the first event?

A Exponential with rate $\lambda$. (See below for a justification.)

More generally, starting from any point in time, the arrival time of the next event is exponential. In particular, the *inter-arrival* times between consecutive events are independent and exponentially distributed with the same rate $\lambda$ (see Figure 6.5).

(Q) Let $r \geq 1$ be an integer. What is the distribution of the arrival time of the $r$-th event?

A Gamma with parameters $r$ (the shape) and $\lambda$ (the rate).

Indeed, the arrival time of a the $r$-th event is a sum of $r$ independent exponential random variables, each with rate $\lambda$. This is what we earlier called a gamma random variable (see Example 6.2.2 and the paragraph after that).

(Q) Let $I$ be a time interval. What is the distribution of $N(I)$ (i.e., the number of events occurring in $I$)?

A Poisson with parameter $\lambda \times \text{length}(I)$. (See below for a justification.)

This completes the analogy in Table 6.1. The missing distribution is the Poisson distribution!

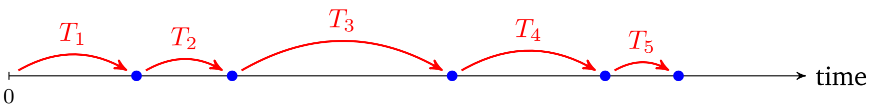



Figure 6.5: The inter-arrival times in a Poisson process.

A justification of the above claims can be given through approximation of a Poisson process with a Bernoulli process.

**Approximation of a Poisson process with a Bernoulli process.** Suppose we divide the time into tiny intervals of length $\Delta t$ (see Figure 6.6). Let us call the endpoints of these intervals $t_0, t_1, t_2, \ldots$, so that $t_0 = 0$, $t_1 = \Delta t$, $t_2 = 2\Delta t$, and so on. We can associate a Bernoulli trial to each of these tiny intervals $[t_{k-1}, t_k)$. The trial associated to $[t_{k-1}, t_k)$ will be considered "successful" if an event has occurred during that interval.

(Q) What is the probability of "success" for the trial at interval $[t_{k-1}, t_k)$?

A Approximately $\lambda \Delta t$.

Indeed, since $\Delta t$ is very small, it is very unlikely for the interval $[t_{k-1}, t_k)$ to contain more than one event. Thus, $N\big([t_{k-1}, t_k)\big)$, the number of events in this interval, is in effect a Bernoulli random variable, indicating whether the trial has been "successful" or not. The parameter of this Bernoulli random variable (the probability of "success") coincides with its expected value, which is $\lambda \Delta t$.

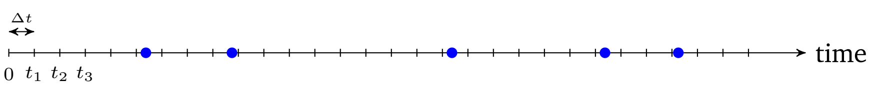
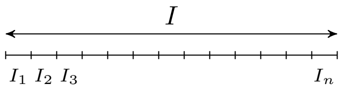



Figure 6.6: An approximation of a Poisson process with a Bernoulli process. There is a trial corresponding to each tiny interval, which is considered "successful" if an event has occurred during that interval.

This approximation can be used to derive the distribution of the arrival time $T_1$ of the first event and the distribution of the number $N(I)$ of events occurring within a time interval $I$. Let us start with the latter.

Ⓠ Why does $N(I)$ have a Poisson distribution?

A Let $\ell$ denote the length of $I$. The interval $I$ is partitioned into roughly $n \approx \ell/\Delta t$ tiny intervals $I_1, I_2, \ldots, I_n$, each with length $\Delta t$.




Clearly, $N(I) = N(I_1) + N(I_2) + \cdots + N(I_n)$. Since the tiny intervals $I_1, I_2, \ldots, I_n$ are disjoint, the variables $N(I_1), N(I_2), \ldots, N(I_n)$ are independent. Furthermore, as argued above, each $N(I_i)$ is approximately a Bernoulli random variable with parameter $\lambda\Delta t$. Thus, the distribution of $N(I)$ is approximately binomial with parameters $n$ and $\lambda\Delta t$.

Now, let $k$ be a non-negative integer. Assuming $\Delta t$ is small enough so that $n \approx \ell/\Delta t \geq k$, we have

$$\begin{aligned}
\mathbb{P}\left(N(I) = k\right) &\approx \binom{n}{k}(\lambda\Delta t)^k(1-\lambda\Delta t)^{n-k} \\
&\approx \frac{n!}{k!(n-k)!}\left(\lambda\frac{\ell}{n}\right)^k\left(1-\lambda\frac{\ell}{n}\right)^{n-k} && \text{(since } \ell \approx n\Delta t\text{)} \\
&= \frac{(\lambda\ell)^k}{k!}\underbrace{\left(1-\frac{\lambda\ell}{n}\right)^n}_{\approx \mathrm{e}^{-\lambda\ell}}\underbrace{\frac{n!}{n^k(n-k)!}}_{\approx 1}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\approx 1} && \text{(rearrangement)} \\
&\approx \frac{(\lambda\ell)^k}{k!}\mathrm{e}^{-\lambda\ell}\,, && (🦈)
\end{aligned}$$

where the latter approximations follow from the facts that:

- $\left(1-\frac{\lambda\ell}{n}\right)^n \to \mathrm{e}^{-\lambda\ell}$ as $n \to \infty$,[6]
- $\frac{n!}{n^k(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \to 1$ as $n \to \infty$,
- $\left(1-\frac{\lambda}{n}\right)^{-k} \to 1$ as $n \to \infty$.

Note that the expression (🦈) is simply the probability mass function of the Poisson distribution with parameter $\lambda\ell$.

Ⓠ Why does $T_1$ (the arrival time of the first event) have an exponential distribution?

A1 Let us derive the cumulative distribution function $F(t)$ of $T_1$. Clearly, $T_1$ cannot be negative, hence $F(t) = 0$ for $t < 0$.

Let $t \geq 0$. Let $K_1$ denote the number of trials until the first "success" so that the first event arrives during the $K_1$-th tiny interval. Clearly, $K_1$ is a geometric random variable with parameter $\lambda\Delta t$. Observe that $T_1 \approx K_1\Delta t$. Thus, defining $m := t/\Delta t$, we have

$$\begin{aligned}
F(t) &= \mathbb{P}(T_1 \leq t) \\
&\approx \mathbb{P}(K_1\Delta t \leq t) \\
&= \mathbb{P}(K_1 \leq t/\Delta t) \\
&\approx 1 - (1-\lambda\Delta t)^{t/\Delta t} && \text{(since } K_1 \sim \mathsf{Geom}(\lambda\Delta t)\text{)} \\
&\approx 1 - \left(1 - \frac{\lambda t}{m}\right)^m && \text{(using } t = m\Delta t\text{)} \\
&\approx 1 - \mathrm{e}^{-\lambda t}\,, && (☎)
\end{aligned}$$

where we have again used the fact that

- $\left(1-\frac{\lambda t}{m}\right)^m \to \mathrm{e}^{-\lambda t}$ as $m \to \infty$.[6]

Note that the expression (☎) is simply the cumulative distribution function of the exponential distribution with rate $\lambda$.

[6] The proof of the identity $\lim_{n\to\infty}(1 - x/n)^n = \mathrm{e}^{-x}$ can be found in most calculus textbooks.

A2 For $t \geq 0$, observe that the event $\{T_1 > t\}$ (the first event arrives after time $t$) is the same as the event $\{N([0,t]) = 0\}$ (there are no events during the interval $[0,t]$). As we saw earlier, $N([0,t])$ is a Poisson random variable with parameter $\lambda t$. Thus,

$$\mathbb{P}(T_1 > t) = \mathbb{P}\left(N([0,t] = 0\right) = \mathrm{e}^{-\lambda t} .$$

Thus, the cumulative distribution function of $T_1$ is

$$F(t) = \mathbb{P}(T \leq t) = 1 - \mathbb{P}(T > t) = \begin{cases} 1 - \mathrm{e}^{-\lambda t} & \text{if } t \geq 0, \\ 0 & \text{if } t < 0. \end{cases}$$

Hence, $T_1$ is exponential with rate $\lambda$.

## 6.5 Poisson processes: An example

**Example 6.5.1** (Customers). Starting from its opening hours, customers arrive at a shop at random times with an average rate of 2 customers per hour. It is reasonable to model the arrival of customers with a Poisson process with rate $\lambda = 2\,\mathsf{hour}^{-1}$.[7]

Q What is the probability that the first customer arrives within the first 15 minutes?

A Let $T_1$ denote the arrival time of the first customer (see Figure 6.5). We know that $T_1$ has an exponential distribution with rate $\lambda = 2\,\mathsf{hour}^{-1}$. Denoting the cdf of $T_1$ by $F_{T_1}$, we can write

$$\mathbb{P}(T_1 \leq 15\,\mathsf{min}) = \mathbb{P}(T_1 \leq 0.25\,\mathsf{hour}) = F_{T_1}(0.25\,\mathsf{hour}) = 1 - \mathrm{e}^{-2\times 0.25} = 1 - \mathrm{e}^{-0.5} \approx 0.3935 .$$

Q What is the probability that the second customer does not arrive within the first 2 hours?

A1 Let $T_2$ denote the inter-arrival time between the first and the second customers (see Figure 6.5), and $S_2 \coloneqq T_1 + T_2$. We are looking for the probability that $S_2 > 2$.

We know that $S_2$ has a Gamma distribution with shape $r = 2$ and rate $\lambda = 2\,\mathsf{hour}^{-1}$. Denoting the cdf of $S_2$ by $F_{S_2}$, we have $\mathbb{P}(S_2 > 2) = 1 - F_{S_2}(2)$. Using a computer software such as R, we can find $F_{S_2}(2) \approx 0.90842$, thus $\mathbb{P}(S_2 > 2) \approx 1 - 0.90842 = 0.09158$.

Alternatively, if we let $f_{S_2}$ denote the pdf of $S_2$ given in ($\blacklozenge$), we can write

$$\begin{aligned}
\mathbb{P}(S_2 > 2) &= \int_2^\infty f_{S_2}(x)\,\mathrm{d}x \\
&= \int_2^\infty 4x\mathrm{e}^{-2x}\,\mathrm{d}x \\
&= -2x\mathrm{e}^{-2x}\Big|_{x=2}^{\infty} + \int_2^\infty 2\mathrm{e}^{-2x}\,\mathrm{d}x && \text{(integration by parts)} \\
&= 5\mathrm{e}^{-4} \\
&\approx 0.09158 .
\end{aligned}$$

A2 The event that the second customer does not arrive within the first two hours can equivalently be described as $N([0,2]) \leq 1$, where $N([0,2])$ denotes the number of customers arriving within the first two hours.

We know that $N([0,2])$ is a Poisson random variable with parameter

$$\mu = \lambda \times \operatorname{length}([0,2]) = 2\,\mathsf{hour}^{-1} \times 2\,\mathsf{hour} = 4 .$$

Therefore,

$$\begin{aligned}
\mathbb{P}(N([0,2]) \leq 1) &= \mathbb{P}(N([0,2]) = 0) + \mathbb{P}(N([0,2]) = 1) \\
&= \mathrm{e}^{-\mu} + \mathrm{e}^{-\mu}\frac{\mu}{1!} \\
&= 5\mathrm{e}^{-4} \\
&\approx 0.09158 .
\end{aligned}$$

[7] Why is a Poisson process a good model in this scenario? Does the process of the arrival of customers satisfy the defining conditions of a Poisson process with rate $\lambda$?

Ⓠ What is the probability that, among the first $5$ customers, no two arrive within $15$ minutes of each other?

A For $k = 2, 3, 4, \ldots$, let $T_k$ denote the inter-arrival duration between the $(k-1)$st and the $k$th customers (see Figure 6.5). We are looking for the probability that $(T_2 > 15\,\mathsf{min})$ and $(T_3 > 15\,\mathsf{min})$ and $(T_4 > 15\,\mathsf{min})$ and $(T_5 > 15\,\mathsf{min})$.

We know that $T_2, T_3, T_4, T_5$ are independent, each having an exponential distribution with rate $\lambda = 2\,\mathsf{hour}^{-1}$. Therefore,

$$\begin{aligned}
&\mathbb{P}\big((T_2 > 15\,\mathsf{min}) \text{ and } (T_3 > 15\,\mathsf{min}) \text{ and } (T_4 > 15\,\mathsf{min}) \text{ and } (T_5 > 15\,\mathsf{min})\big) \\
&\quad = \mathbb{P}(T_2 > 15\,\mathsf{min})\,\mathbb{P}(T_3 > 15\,\mathsf{min})\,\mathbb{P}(T_4 > 15\,\mathsf{min})\,\mathbb{P}(T_5 > 15\,\mathsf{min}) \\
&\quad = \left(\mathrm{e}^{-2\times 0.25}\right)^4 \\
&\quad = \mathrm{e}^{-2} \\
&\quad \approx 0.1353\,.
\end{aligned}$$

○

## 6.6 More on Poisson processes

[To be written!]