13. Continuous random variables

Imagine a stick of unit length that is oriented horizontally from left to right. Consider choosing a point x uniformly at random between 0 and 1 and breaking the stick x units from the left end. In this example, the sample space can be regarded as $\Omega = [0, 1)$. For finitely many subintervals (non-overlapping) $J_1, \ldots, J_n$, with $E := \bigcup_{i=1}^n J_i$, we define

$$P[E] := P\left[\bigcup_{i=1}^n J_i\right] = \sum_{i=1}^n \text{length}(J_i).$$

The intuition here is that there are too many possible values of X, so each individual value has an excessively small (in fact, zero) probability of occurring. Think back to our interpretation of the event space $(\Omega, \mathcal{E})$ at the beginning of the semester. The elements $E \in \mathcal{E}$ are the events that we *can obtain information about*. In the above example, even if we could generate a uniform number x to infinite precision, you could never know exactly what the number is; therefore, even an observation such as $x = 0.1542847$ only tells you that x is in the interval $[0.15428465, 0.15428475)$.

By this definition, we have $P[[0, 1)] = P[\Omega] = 1 - 0 = 1$, $P[[0, 1/2)] = 1/2$ and $P[[x, x]] = P[\{x\}] = x - x = 0$ for every $x \in [0, 1)$. This example emphasizes the importance of *countable* additivity of P. In particular, we can write $\Omega = \bigcup_{\omega \in [0,1)} \{\omega\}$, which is an uncountable union, and

$$1 = P[\Omega] = P\left[\bigcup_{\omega \in \Omega} \{\omega\}\right] \neq \sum_{\omega \in \Omega} P[\{\omega\}] = 0.$$

Definition 13.1 (Continuous random variables)**.** *Let $(\Omega, \mathcal{E}, P)$ be a probability model. A function $X : \Omega \to \mathbb{R}$ has a* continuous distribution *is there exists a function $f : \mathbb{R} \to [0, \infty)$ such that*

$$P\{X \in B\} = \int_B f(x)dx$$

for all events $B \in \mathcal{E}$. Specifically, f satisfies

$$P\{X \in \mathbb{R}\} = P\{-\infty < X < \infty\} = \int_{-\infty}^{\infty} f(x)dx = 1.$$

We call f the density *of X.*

If f is continuous at x, then

$$\lim_{\Delta \downarrow 0} \frac{P\{X \in [x, x + \Delta]\}}{\text{length}[x, x + \Delta]} = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_x^{x+\Delta} f(y)dy = f(x).$$

Therefore, we interpret $f(x)dx$ as the probability that X is in an *infinitesimally small* window $[x, x + dx]$ of x.

The cumulative distribution function $F_X : \mathbb{R} \to [0, 1]$ of X is defined by

$$F_X(x) := P\{X \leq x\} = P\{X \in (-\infty, x]\} = \int_{-\infty}^{x} f(y)dy$$

and, for $x < z$, $P\{x < X \leq z\} = F_X(z) - F_X(x)$.

Conversely, if F is continuously differentiable, or everywhere continuous and differentiable at all but finitely many points, then

$$F(z) - F(x) = \int_x^z F'(y)dy,$$

so X has density $f = F'$ by the fundamental theorem of calculus.

13.1. **Uniform random variables.** Define

$$f(x) := \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then f is the density of a *Uniform*$(0,1)$ *random variable*. If $U \sim \text{Unif}(0,1)$, then

$$\begin{aligned} F_U(u) := P\{U \le u\} &= \int_0^u dx = u, \quad 0 \le u \le 1, \\ \mathbb{E}U &= \int_0^1 u du = 1/2, \\ \mathbb{E}U^2 &= \int_0^1 u^2 du = 1/3, \quad \text{and} \\ \text{Var}(U) &= \mathbb{E}U^2 - [\mathbb{E}U]^2 = 1/3 - (1/2)^2 = 1/12. \end{aligned}$$

For $a < b$, $W = (b-a)U + a$ is uniformly distributed on (a,b). The density of W is

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

Also,

$$\begin{aligned} \mathbb{E}W &= (b-a)\mathbb{E}U + a = \frac{1}{2}(a+b) \\ \text{Var}(W) &= \frac{(b-a)^2}{12}. \end{aligned}$$

Recall the definition of the discrete Uniform distribution: $P\{X = k\} = 1/n$ for $k = 1, \ldots, n$. In this case, $\mathbb{E}X = (n+1)/2$ and $\text{Var}(X) = (n^2-1)/12$. Let N_n have the discrete uniform distribution on $[n] := \{1, \ldots, n\}$, for each $n = 1, 2, \ldots$. Then we say $(N_n/n,\ n \ge 1)$ converges in distribution to the Uniform distribution on $[0,1]$, in the following sense. If $k_n \to \infty$ in such a way that $k_n/n \to u \in [0,1]$, then

$$P\{N_n \le k_n\} = k_n/n = P\{N_n/n \le k_n/n\}$$

and $k_n/n \to u$ and $P\{N_n/n \le k_n/n\} \to u = P\{U \le u\}$, for all $u \in (0,1)$. Note also that

$$\begin{aligned} \mathbb{E}N_n/n = n^{-1}\mathbb{E}N_n &= \frac{n+1}{2n} \to \frac{1}{2} = \mathbb{E}U \quad \text{and} \\ \text{Var}(N_n/n) = n^{-2}\,\text{Var}(N_n) &= \frac{n^2-1}{12n^2} \to \frac{1}{12} = \text{Var}(U). \end{aligned}$$

Therefore, one way to generate an *approximately* uniform random variable U is to sample x uniformly from $\{1, \ldots, n\}$, for some large value of $n = 1, 2, \ldots$, and then put $U = x/n$. This procedure gives a uniform random variable up to precision $1/n$. See Section 13.3 for a much better way to do this.

13.2. **Expectation of continuous random variables.** Let X be a random variable with density f. Then, for $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}g(X) := \int_{-\infty}^{\infty} g(x)f(x)dx = \int_{-\infty}^{\infty} g^+(x)f(x)dx - \int_{-\infty}^{\infty} g^-(x)f(x)dx,$$

where $g^\pm(x) := \max(\pm g(x), 0)$. The same rules apply for computing expectations of continuous random variables, with the modification that we now compute the integral instead of a sum. In particular, for random variables $X, Y \in \mathbb{R}$ and $a, b \in \mathbb{R}$, we have

- $\mathbb{E}(aX + bY + c) = a\mathbb{E}X + b\mathbb{E}Y + c$,
- $\mathrm{Var}(X) = \mathbb{E}X^2 - [\mathbb{E}X]^2$,
- $\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$, where
- $\mathrm{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$.

Now, suppose $X \geq 0$ is a continuous random variable with density f. Then

$$\begin{aligned}
\mathbb{E}X &= \int_0^\infty xf(x)dx \\
&= \int_0^\infty \left[\int_0^x dy\right] f(x)dx \\
&= \int_0^\infty \int_0^\infty I_{\{0\leq y\leq x\}} f(x)dydx \\
&= \int_0^\infty \left[\int_y^\infty f(x)dx\right] dy \\
&= \int_0^\infty P\{X \geq y\}dy.
\end{aligned}$$

Also,

$$\begin{aligned}
\mathbb{E}X^2 &= \int_0^\infty P\{X^2 \geq y\}dy \\
&= \int_0^\infty P\{X \geq \sqrt{y}\}dy \\
&= \int_0^\infty 2xP\{X \geq x\}dx.
\end{aligned}$$

Example 13.2 (Order statistics)**.** *Let U_1, U_2, U_3, U_4 be i.i.d. Uniform$(0, 1)$ random variables and write $U_{(1)} < U_{(2)} < U_{(3)} < U_{(4)}$ to denote the* order statistics*. Then*

$$\begin{aligned}
P\{U_{(1)} \leq u\} &= P\{\min U_i \leq u\} \\
&= 1 - P\{U_i > u \textit{ for all } i = 1, \dots, 4\} \\
&= 1 - (1 - u)^4 \\
&=: F_{(1)}(u).
\end{aligned}$$

The density of $U_{(1)}$ is

$$f_{(1)}(u) = F'_{(1)}(u) = \begin{cases} 4(1-u)^3, & 0 < u < 1 \\ 0, & \textit{otherwise.} \end{cases}$$

We can compute the expectation of $U_{(1)}$ in a couple of different ways. By definition,

$$\mathbb{E}U_{(1)} = \int_0^1 4u(1-u)^3 du = \int_0^1 4(1-y)y^3 dy = \int_0^1 4y^3 - 4y^4 dy = 1 - 4/5 = 1/5.$$

Alternatively,

$$\mathbb{E}U = \int_0^1 P\{U_{(1)} \geq u\} du = \int_0^1 (1-u)^4 du = 1/5.$$

13.3. **Constructing Uniform (0,1) random variables.** Let $I_1, I_2, \ldots$ be independent Bernoulli(1/2) random variables (based on repeated flips of a fair coin). Given $I_1, I_2, \ldots$, we define

$$Y := \sum_{j=1}^{\infty} I_j 2^{-j}.$$

Since $\{I_j\}_{j\geq 1}$ are random, so is Y. Y is a random variable with dyadic (binary) expansion $0.I_1I_2I_3\ldots$. The set

$$\mathcal{D} := \left\{0.\epsilon_1\epsilon_2\ldots : \epsilon_j \in \{0,1\}\right\}$$

of *dyadic rationals* is *dense* in [0,1], i.e., for every $u \in [0,1]$ there exists a $\delta > 0$ such that some $d \in (u-\delta, u+\delta)$ is in $\mathcal{D}$.

Suppose we terminate our expansion above at $j = 5$. Then we obtain an expansion $Y_5 := 0.i_1i_2i_3i_4i_5$, where $i_1, \ldots, i_5$ are realizations of independent Bernoulli trials. What is the distribution of Y_5?

$$P\{Y_5 = u\} = \begin{cases} 2^{-5}, & u = 0.\epsilon_1\ldots\epsilon_5,\ \epsilon_j \in \{0,1\} \\ 0, & \text{otherwise.} \end{cases}$$

In general, if we stop after $n \geq 1$ trials and define $Y_n := i_1 \ldots i_n$, then

$$P\{Y_n = u\} = \begin{cases} 2^{-n}, & u = 0.\epsilon_1\ldots\epsilon_n,\ \epsilon_j \in \{0,1\} \\ 0, & \text{otherwise.} \end{cases}$$

In other words, for $n \geq 1$, Y_n is *uniformly distributed* on

$$\mathcal{D}_n := \{0.\epsilon_1\ldots\epsilon_n : \epsilon_j \in \{0,1\},\ j = 1,\ldots,n\}.$$

Now, take any $u \in [0,1]$. Since $\mathcal{D}$ is dense in u, there exists a sequence $d_1, d_2, \ldots$, with $d_n \in \mathcal{D}_n$ for each $n \geq 1$ such that $d_n \to u$. In fact, we can choose d_n so that $|u - d_n| \leq 2^{-n}$ for each $n \geq 1$. Now, generate a sequence $I_1, I_2, \ldots$ i.i.d. Bernoulli(1/2) and, for each $n \geq 1$, define $Y_n := I_1 \ldots I_n$ as above. For the d_n we have chosen, we have

$$d_n - 2^{-n} \leq P\{Y_n \leq u\} \leq d_n + 2^{-n} \quad \text{for all } n \geq 1.$$

By the sandwich lemma, we have $\lim_{n\to\infty} P\{Y_n \leq u\} = \lim_{n\to\infty} d_n = u$, for all $u \in (0,1)$. This is the notion of *convergence in distribution*. This method tells us that we can approximate a uniform random variable to precision 2^{-n} by flipping a coin only n times. This approach is more efficient than that described above, where we must generate a number between 1 and n to obtain precision only up to n^{-1}.

13.4. **The Normal (Gaussian) distribution.** A random variable Z has the *standard normal distribution* if it is continuous with density

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

The Normal distribution is the most important distribution in statistics, as we will see later in the course. The expectation and variance can be computed as follows:

$$\begin{aligned} \mathbb{E}Z &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0 \quad \text{and} \\ \mathrm{Var}(Z) = \mathbb{E}Z^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \times z e^{-z^2/2} dz \\ &= \left[z(-e^{-z^2/2}/\sqrt{2\pi}) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\frac{e^{-z^2/2}}{\sqrt{2\pi}} = 1. \end{aligned}$$

The density of Z is commonly denoted ϕ and the distribution function is denoted Φ. For $\alpha \in (0,1)$, z_α is the number such that $\Phi(z_\alpha) = \alpha$, e.g., $z_{0.5} = 0, z_{0.84} \approx 1, z_{0.025} \approx -2$, and so on. These are called the α-quantiles of the standard normal distribution and are used repeatedly in statistical inference.

A random variable X is said to obey the *Normal distribution* with parameter (μ, σ^2), for $-\infty < \mu < \infty$ and $\sigma^2 > 0$, written $X \sim N(\mu, \sigma^2)$, if

$$Z := \frac{X - \mu}{\sigma} \quad \text{is a standard Normal random variable.}$$

We denote the distribution of X by $\Phi_{\mu,\sigma}$. The above transformation of X is called a change of *location* of μ and change of *scale* by σ. Consequently, the family $\{\Phi_{\mu,\sigma} : -\infty < \mu < \infty, \sigma^2 > 0\}$ of Normal distributions is called a *location-scale family* of distributions, which will be covered in another course on statistical inference.

Since $X = \mu + \sigma Z$, where $Z \sim N(0,1)$, we have

$$\begin{aligned} \mathbb{E}X &= \mu + \sigma \mathbb{E}Z = \mu \quad \text{and} \\ \mathrm{Var}(X) &= \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \,\mathrm{Var}(Z) = \sigma^2. \end{aligned}$$

Furthermore, X has a density

$$\begin{aligned} f_X(x)dx &= P\{x \le X \le x + dx\} \\ &= P\left\{ \frac{x-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma} + \frac{dx}{\sigma} \right\} \\ &= P\left\{ \frac{x-\mu}{\sigma} \le Z \le \frac{x-\mu}{\sigma} + \frac{dx}{\sigma} \right\} \\ &= f_Z\left(\frac{x-\mu}{\sigma} \right) \frac{dx}{\sigma}. \end{aligned}$$

Therefore, we denote the density of X by $\phi_{\mu,\sigma}$, which satisfies

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < \infty. \tag{22}$$

Theorem 13.3 (Demoivre–Laplace Theorem)**.** *Let X be a Binomial random variable with parameter (n, p). Then X is approximately Normally distributed with parameter (np, npq).*

Remark 13.4. *This approximation requires $npq \geq 10$ (or so) in order to be accurate.*

Example 13.5 (Normal approximation to Binomial)**.** *Spin a pointer 10 times with 4 possible outcomes, A,B,C,D. Let N_A denote the number of times the pointer points toward A. Then N_A follows the Binomial distribution with parameter* $(10, 1/4)$*. We have*

$$P\{N_A \geq 4\} = \sum_{k=4}^{10} \binom{10}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{10-k} \approx 0.224.$$

By Demoivre–Laplace, N_A is approximately Normally distributed with mean $\mu = np = 2.5$ and standard deviation $\sigma = \sqrt{npq} \approx 1.37$. The Normal approximation (with continuity correction) gives

$$P\{D \geq 4\} \approx P\left\{\frac{D - 2.5}{1.37} \geq \frac{3.5 - 2.5}{1.37}\right\} = 1 - \Phi(0.73) = 0.233.$$

Example 13.6 (Marbles)**.** *A large collection of marbles has diameters that are approximately Normally distributed with mean 1 cm. One third of the marbles have diameters greater than 1.1 cm.*

(a) **What is the standard deviation?** *From the normal table, $z_{0.667} = 0.43$; hence, $\frac{X-\mu}{\sigma} \sim N(0,1)$ and*

$$\frac{1.1 - 1}{\sigma} = 0.43 \quad \Longrightarrow \quad \sigma = 0.1cm/0.43 = 0.23cm.$$

(b) **What fraction of marbles have diameters within 0.2 cm of the mean?**

$$\begin{aligned} P\{0.8 \leq D \leq 1.2\} &= P\left\{\frac{-0.2}{0.23} \leq Z \leq \frac{1.2 - 1}{0.23}\right\} \\ &= P\{-0.86 \leq Z \leq 0.86\} \\ &= \Phi(0.86) - \Phi(-0.86) \\ &= \Phi(0.86) - [1 - \Phi(0.86)] \\ &= 2\Phi(0.86) - 1 \\ &= 0.61. \end{aligned}$$

(c) **What diameter is exceeded by 75% of marbles?** *The 0.25 quantile of Z is $z_{0.25} = -0.675$. Hence, the diameter exceeded by 75%, denoted d_{75}, is*

$$d_{75} = \mu - 0.675\sigma = 1 - 0.675 \times 0.23 = 0.84\ cm.$$

13.5. **The Exponential distribution.** Consider modeling arrivals of a bus as follows. Fix some $\lambda > 0$ and assume that within each time interval, the number of bus arrivals in that interval follows a Poisson distribution with parameter $\lambda \times$ (length of the interval). For example, if $\lambda = 2$/hour, then the number of arrivals between 12:00PM and 2:00PM is Poisson with parameter $2 \times 2 = 4$. Further, we assume that the number of arrivals in non-overlapping intervals are independent Poisson random variables with the appropriate means. Therefore, the number of arrivals between 12:00 and 2:00 is Poisson(4) and is independent of the number of arrivals between 2:30 and 3:00, which is Poisson(1). The above model is called a *time-homogeneous Poisson process* with intensity parameter $\lambda > 0$. Writing $N_{(s,t)}$ to denote the number of arrivals in the interval (s, t), the Poisson process is determined by the marginal distributions

$$N_{(s,t)} \sim \text{Poisson}(\lambda(t - s)).$$

Let $T := \inf\{t > 0 : N_{[0,t]} = 1\}$ denote the time of the first arrival. What is the distribution of T?

- First method: Note that the event $\{T > t\}$ is the same as the event that there are no arrivals in the interval $[0, t]$ for our Poisson process. Therefore,

$$P\{T > t\} = P\{N_{[0,t]} = 0\} = e^{-\lambda t},$$

by the assumption that counts follow the Poisson distribution; therefore, $F_T(t) = P\{T \leq t\} = 1 - P\{T > t\} = 1 - e^{-\lambda t}$ and $f_T(t) = F_T'(t) = \lambda e^{-\lambda t}$.
- Second method: For $t \geq 0$ and $dt > 0$,

$$\begin{aligned} f_T(t)dt &= P\{t \leq T \leq t + dt\} \\ &= P\{\text{no arrival in } [0,t), \text{ one arrival in } [t, t+dt]\} \\ &= P\{N_{[0,t)} = 0\} \times P\{N_{[t,t+dt)} = 1\} \\ &= e^{-\lambda t} \times e^{-\lambda dt}(\lambda dt)^1/1! \\ &= \lambda e^{-\lambda t} \times e^{-\lambda dt} dt. \end{aligned}$$

For the density of T, we have

$$f_T(t) = \lim_{dt \downarrow 0} \frac{P\{t \leq T \leq t + dt\}}{dt} = \lim_{dt \downarrow 0} \frac{\lambda e^{-\lambda t} \times e^{-\lambda dt} dt}{dt} = \lambda e^{-\lambda t}.$$

You can check that f_T above integrates to one and so is the proper probability density. The random variable T is said to have the *Exponential distribution* with intensity parameter $\lambda > 0$, denoted $T \sim \text{Exp}(\lambda)$.

Now, put $X = \lambda T$, where $T \sim \text{Exp}(\lambda)$. Then X is called *standard Exponential* random variable:

$$\begin{aligned} &F_X(x) = P\{X \leq x\} = P\{T \leq x/\lambda\} = 1 - e^{-x}, \quad x > 0 \quad \text{and} \\ &f_X(x) = e^{-x}. \end{aligned}$$

To compute the moments of the Exponential distribution, we introduce the *Gamma function*, which for real numbers $r > 0$ is defined by

$$\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} dx.$$

The Gamma function takes some special values. In particular, $\Gamma(1) = 1$ and, for any $r > 0$,

$$\begin{aligned} \Gamma(r+1) &= \int_0^\infty x^r e^{-x} dx \\ &= [x^r(-e^{-x})]_0^\infty - \int_0^\infty r x^{r-1}(-e^{-x}) dx \\ &= r\Gamma(r). \end{aligned}$$

Therefore, if r is a positive integer, then $\Gamma(r+1) = r(r-1)\cdots 2 \cdot 1 \cdot \Gamma(1) = r!$.

Now, let X be a standard Exponential random variable. Using the Gamma function, we have

$$\begin{aligned}\mathbb{E}X &= \int_0^\infty xe^{-x}dx = \Gamma(2) = 1,\\ \mathbb{E}X^2 &= \int_0^\infty x^2e^{-x}dx = \Gamma(3) = 2, \quad \text{and}\\ \mathbb{E}X^n &= \int_0^\infty x^ne^{-x}dx = \Gamma(n+1) = n!.\end{aligned}$$

We also have $\mathrm{Var}(X) = \mathbb{E}X^2 - [\mathbb{E}X]^2 = 2 - 1 = 1$. Now, let $T \sim \mathrm{Exp}(\lambda)$. Then T has the same distribution as X/λ and so

$$\mathbb{E}T = 1/\lambda; \quad \mathrm{Var}(T) = 1/\lambda^2; \quad \mathrm{SD}(T) = 1/\lambda.$$

Example 13.7 (Electrical systems)**.** *Consider two components C_1, C_2 of an electrical system. The system functions as long as both components are functioning. Suppose lifetimes T_1, T_2 of C_1 and C_2 are modeled by independent Exponential random variables with parameters λ_1, λ_2 and let T be the lifetime of the system. What is the distribution of $T := \min(T_1, T_2)$?*

- *Method 1:* **CDF approach**
  Let $t \geq 0$, then

$$\begin{aligned}P\{T > t\} &= P\{T_1 > t,\, T_2 > t\}\\ &= P\{T_1 > t\}P\{T_2 > t\}\\ &= e^{-\lambda_1 t}e^{-\lambda_2 t}\\ &= e^{-(\lambda_1+\lambda_2)t}\\ &= e^{-\lambda t},\end{aligned}$$

  where $\lambda = \lambda_1 + \lambda_2$.
- *Method 2:* **Infinitesimals approach**
  Let $t \geq 0$, then

$$\begin{aligned}f(t)dt &= P\{t - dt < T \leq t\}\\ &= P\{t - dt < T_1 \leq t,\, T_2 > t\} + P\{t - dt < T_2 \leq t\, T > t\}\\ &= P\{t - dt < T_1 \leq t\}P\{T_2 > t\} + P\{t - dt < T_2 \leq t\}P\{T_1 > t\}\\ &= \lambda_1 e^{-\lambda_1 t}dte^{-\lambda_2 t} + \lambda_2 e^{-\lambda_2 t}dte^{-\lambda_1 t}\\ &= \lambda e^{-\lambda t}dt.\end{aligned}$$

We conclude that $T = \min(T_1, T_2) \sim Exp(\lambda_1 + \lambda_2)$ and $\mathbb{E}T = 1/(\lambda_1 + \lambda_2)$.

What is the probability that component 1 fails before component 2?

$$\begin{aligned}P\{T_2 > T_2\} &= \int_0^\infty P\{T_2 > T_1 \mid T_1 = t\}f_{T_1}(t)dt\\ &= \int_0^\infty P\{T_2 > t \mid T_1 = t\}\lambda_1 e^{-\lambda_1 t}dt\\ &= \int_0^\infty e^{-\lambda_2 t}\lambda_1 e^{-\lambda_1 t}dt\\ &= \lambda_1/(\lambda_1 + \lambda_2).\end{aligned}$$

For $s \geq 0$, what is $P\{T > s$ and $T_2 > T_1\}$? (Exercise.)

The following is a (very good) exercise.

Proposition 13.8. *Let $T_1, T_2, \ldots, T_n$ be independent Exponential random variables with intensities $\lambda_1, \ldots, \lambda_n$, respectively. Then the minimum of $T_1, \ldots, T_n$ follows the Exponential distribution with parameter $\lambda_\bullet := \lambda_1 + \cdots + \lambda_n$.*

When studying discrete distributions, we mentioned that the Geometric distribution is the unique discrete distribution with the memoryless property. In fact, the Exponential distribution is the unique continuous distribution with the memoryless property: Let $T \sim \mathrm{Exp}(\lambda)$, $s \neq t \geq 0$, then

$$\begin{aligned} P\{T \geq t+s \mid T \geq t\} &= \frac{P\{T \geq t+s,\, T \geq t\}}{P\{T \geq t\}} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s}. \end{aligned}$$

Given $T \geq t$, the *residual lifetime* $T - t$ has the Exponential distribution with parameter λ. The Exponential distribution is the only memoryless non-negative continuous distribution. The memoryless property is useful for quick probability calculations. For example, suppose the lifetime of a light bulb is Exponentially distributed with parameter 1/200. Then, given the bulb is still working after 800 hours, how much longer do you expect it to work? By the memoryless property, the residual lifetime is still Exponential with parameter 1/200, and so we expect it to work for another 200 hours.

13.6. **The Gamma distribution.** Consider a Poisson process with intensity λ. Let W_r be the waiting time of the rth arrival. Then

$$P\{W_r > t\} = P\{N_{[0,t]} \leq r-1\} = \sum_{j=0}^{r-1} e^{-\lambda t}(\lambda t)^j / j!$$

and

$$\begin{aligned} f_r(t)dt &= P\{t \leq W_r \leq t + dt\} \\ &= P\{N_{[0,t)} = r-1,\, N_{[t,t+dt)} = 1\} \\ &= \frac{e^{-\lambda t}(\lambda t)^{r-1}}{(r-1)!} \times \lambda dt e^{-\lambda dt} \\ &= \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} dt e^{-\lambda dt}. \end{aligned}$$

A random variable W with density

$$f(t) = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} I_{(0,\infty)}(t)$$

is said to have the *Gamma distribution* with shape parameter r and intensity λ, written $W \sim \mathrm{Gamma}(r, \lambda)$.

If $W \sim \text{Gamma}(r, \lambda)$, then $X = \lambda W \sim \text{Gamma}(r, 1)$. Now, suppose $X \sim \text{Gamma}(r, 1)$, then

$$\begin{aligned}
\mathbb{E}X &= \frac{1}{\Gamma(r)}\int_0^\infty x \times x^{r-1}e^{-x}dx = \frac{\Gamma(r+1)}{\Gamma(r)} = \frac{r\Gamma(r)}{\Gamma(r)} = r, \\
\mathbb{E}X^2 &= \Gamma(r+2)/\Gamma(r) = r(r+1), \quad \text{and} \\
\text{Var}(X) &= r^2 + r - r^2 = r.
\end{aligned}$$

Therefore, if $W \sim \text{Gamma}(r, \lambda)$, then $\mathbb{E}W = r/\lambda$ and $\text{Var}(W) = r/\lambda^2$.

13.7. **Hazard functions.** Let T be a nonnegative continuous random variable with density f and distribution function F. We interpret T as the lifetime of *something*. Define

$$\lambda(t) := \frac{P\{t < T \leq t + dt \mid T > t\}}{dt}$$

to be the ratio of (the probability of dying in the next dt time units, given that it has lived at least $t \geq 0$) to dt. We call $\lambda(t)$ the *death rate*, *failure rate*, or *hazard rate* at time t. The *survival probability* is given by

$$S(t) := P\{T > t\} = 1 - F(t), \quad t > 0.$$

From the hazard rate, we obtain the distribution of T:

$$\begin{aligned}
\lambda(t) &= \frac{P\{t \leq T \leq t + dt \mid T > t\}}{dt} \\
&= \frac{P\{t \leq T \leq t + dt\}}{dtP\{T > t\}} \\
&= \frac{f(t)}{1 - F(t)} \\
&= -\frac{d}{dt}\log(1 - (F(t))).
\end{aligned}$$

As an example, for T following the Exponential distribution with rate λ, we have

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda,$$

which is called the *constant force of mortality* in actuarial modeling.

In general, let $\lambda(t)$ be a hazard function for some random variable T. Then

$$\int_0^t \lambda(s)ds = [-\log(1 - F(s))]_0^t = -\log(1 - F(t)) + \log(1 - F(0)).$$

Therefore,

$$S(t) = 1 - F(t) = \exp\left\{-\int_0^t \lambda(s)ds\right\}$$

and

$$f(t) = -\frac{d}{dt}S(t) = -\exp\left\{-\int_0^t \lambda(s)ds\right\}(-\lambda(t)) = \lambda(t)S(t).$$

This shows that the Exponential distribution is the only memoryless distribution since, if T is memoryless, then $\lambda(t) = f(0)$ for all $t \geq 0$. Thus, T has constant hazard rate function and must follow the Exponential distribution with parameter $\lambda = f(0)$.

Example 13.9 (The Rayleigh distribution)**.** *Suppose T has hazard function $\lambda(t) = t$. Then*

$$\begin{aligned} S(t) &= \exp\left\{-\int_0^t \lambda(s)ds\right\} \\ &= \exp\left\{-\int_0^t sds\right\} \\ &= e^{-t^2/2} \end{aligned}$$

and $f(t) = \lambda(t)S(t) = te^{-t^2/2}$. The mean of T is

$$\mathbb{E}T = \int_0^\infty t^2 e^{-t^2/2}dt = \frac{\sqrt{2\pi}}{2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}t^2 e^{-t^2/2}dt = \sqrt{\pi/2}.$$

Also,

$$\mathbb{E}T = \int_0^\infty S(t)dt = \int_0^\infty e^{-t^2/2}dt = \frac{\sqrt{2}}{2}\int_0^\infty u^{-1/2}e^{-u}du = \frac{1}{\sqrt{2}}\Gamma(1/2).$$

Therefore,

$$\begin{aligned} \Gamma(1/2) &= \sqrt{\pi} \\ \Gamma(3/2) &= \frac{1}{2}\Gamma(1/2) = \sqrt{\pi}/2 \\ \Gamma(5/2) &= \frac{3}{2}\Gamma(3/2) = 3\sqrt{\pi}/4. \end{aligned}$$

We have

$$\mathbb{E}T^2 = \int_0^\infty 2tS(t)dt = 2\int_0^\infty te^{-t^2/2}dt = 2.$$

and $Var(T) = 2 - \pi/2$.

We say that T has the (standard) Rayleigh distribution*.*

13.8. **Some extreme value theory.** Suppose we want to analyze extreme (possibly catastrophic) events, such as the occurrence of a "storm of the century" or an economic collapse. Then we consider the distribution of the maximum of a sequence of random variables. For example, we might be interested in the largest daily drop of the Dow Jones Industrial Average, the maximum insurance claim submitted over a certain period, maximum weekly rainfall, etc.

Let $X_1, X_2, \ldots$ be i.i.d. standard exponential random variables and, for $n \geq 1$, let $M_n := \max(X_1, \ldots, X_n)$. Then, for $x \geq 0$,

$$P\{M_n \leq x\} = P\{X_1 \leq x, \ldots, X_n \leq x\} = (1 - e^{-x})^n,$$

which has density

$$f_{M_n}(x) = ne^{-x}(1 - e^{-x})^{n-1} = \binom{n}{1}e^{-x} \times (1 - e^{-x})^{n-1}.$$

Now, put $Y_n := M_n - \log n$, let $y \in \mathbb{R}$ and $x_n := y + \log n$. Then, for all large n, $x_n \geq 0$ and

$$\begin{aligned} P\{Y_n \leq y\} &= P\{M_n \leq x_n\} \\ &= (1 - e^{-x_n})^n \\ &= (1 - e^{-y-\log n})^n \\ &= (1 - e^{-y}/n)^n \\ &= P\{N = 0\}, \end{aligned}$$

where N has the Binomial distribution with parameter $(n, e^{-y}/n)$. Consequently, as $n \to \infty$, $p_n = e^{-y}/n \to 0$, $np_n = ne^{-y}/n \to e^{-y}$ and $B_n \to B_\infty \sim \text{Poisson}(e^{-y})$, by the law of small numbers. We therefore have

$$P\{Y_n \leq y\} = (1 - e^{-y}/n)^n \to e^{-e^{-y}} = P\{B_\infty = 0\}.$$

A random variable Y with distribution function $F(y) = e^{-e^{-y}}$ has the *(standard) Double Exponential Extreme Value* distribution, denoted $Y \sim \mathcal{EE}(1)$.

Now, suppose $T \sim \text{Exp}(1)$. Then for all $y \in \mathbb{R}$,

$$P\{-\log T \leq y\} = P\{\log T \geq -y\} = P\{T \geq e^{-y}\} = e^{-e^{-y}},$$

and so $-\log T \sim \mathcal{EE}(1)$.

Some exercises on the Extreme value distribution:

- If $Y \sim \mathcal{EE}(1)$, what is its density?
- Fix k and let $M_{n:k}$ be the kth largest of $X_1, \ldots, X_n$. For $y \in \mathbb{R}$, what is $\lim_{n\to\infty} P\{M_{n:k} - \log n \leq y\}$?