Random variables

Last modified: 2026-09-28 23:45:35 (PDT)

Random variables

Definition 0.1 (Random variable)**.** A **random variable** is a function from a sample space Ω to the real numbers $\mathbb{R}$: it assigns a number to each possible outcome of a random experiment.

Remark. We use uppercase letters (X, Y, Z, ...) for random variables, and lowercase letters (x, y, z, ...) for their realized (observed) values. We write $\{X = x\}$ for the event $\{\omega \in \Omega : X(\omega) = x\}$, and similarly for $\{X \leq x\}$ and other conditions on X. A fully rigorous definition also requires the function to be *measurable*, so that every such set is an event with a probability; that requirement never binds in these notes.
See also: https://en.wikipedia.org/wiki/Random_variable

Example 0.1 (Number of heads in two coin flips)**.** Flip a fair coin twice. The sample space is $\Omega = \{HH, HT, TH, TT\}$, each outcome with probability $1/4$. Let X be the number of heads:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0$$

Then $\{X = 1\} = \{HT, TH\}$, so $\Pr(X = 1) = 1/4 + 1/4 = 1/2$.

Definition 0.2 (Range of a random variable)**.** The **range** of a random variable X, denoted $\mathcal{R}(X)$, is the set of values X can take:

$$\mathcal{R}(X) \stackrel{\text{def}}{=} \{X(\omega) : \omega \in \Omega\}$$

Remark. The range of X is the image of X as a function on Ω. For a general function, these notes say "image", because some sources use "range" for the codomain instead; for a random variable, "range" is the standard term.

Example 0.2 (Range of the number of heads)**.** In Example 0.1, $\mathcal{R}(X) = \{0, 1, 2\}$.

Definition 0.3 (Discrete random variable)**.** A random variable X is **discrete** if its range $\mathcal{R}(X)$ is countable.

Example 0.3 (Discrete random variables)**.** The number of heads in Example 0.1 is discrete, with a finite range. A count with no upper bound, such as the number of emergency-department visits in a day, is discrete with a countably infinite range, $\{0, 1, 2, \ldots\}$.

Definition 0.4 (Continuous random variable)**.** A random variable X is **continuous** if $\Pr(X = x) = 0$ for every $x \in \mathbb{R}$.

Definition 0.5 (Uniform distribution)**.** A random variable X has the **uniform distribution** on an interval $[\alpha, \beta]$, with $\alpha < \beta$, written $X \sim \mathrm{Uniform}(\alpha, \beta)$, if the probability that X falls in any subinterval of $[\alpha, \beta]$ is proportional to the length of that subinterval:

$$\Pr(a \le X \le b) = \frac{b - a}{\beta - \alpha} \quad \text{for all } \alpha \le a \le b \le \beta$$

Example 0.4 (A uniformly distributed random variable)**.** Let $X \sim \mathrm{Uniform}(0, 1)$ (Definition 0.5), so $\Pr(a \le X \le b) = b - a$ for all $0 \le a \le b \le 1$. Taking $a = b = x$ gives $\Pr(X = x) = x - x = 0$ for every $x \in [0, 1]$. Taking $a = 0$ and $b = 1$ gives $\Pr(0 \le X \le 1) = 1$, so by the complement rule X falls outside $[0, 1]$ with probability 0, and $\Pr(X = x) = 0$ for every x outside $[0, 1]$ as well. So X is continuous. Its range, $[0, 1]$, is uncountable.

Example 0.5 (A random variable that is neither discrete nor continuous)**.** Some random variables are neither discrete nor continuous: a time to event that equals exactly 0 with positive probability, and otherwise spreads over $(0, \infty)$, is one. For instance, let $U \sim \mathrm{Uniform}(0, 1)$ (Definition 0.5), and let $T = 1/(1 - U) - 2$ when $1/2 < U < 1$, and $T = 0$ otherwise. On $\{1/2 < U < 1\}$, T is increasing in U and takes values in $(0, \infty)$, and solving $t = 1/(1 - u) - 2$ for u gives $u = 1 - 1/(t + 2)$. So $\{T > 0\} = \{1/2 < U < 1\}$, and $\{T = t\} = \{U = 1 - 1/(t + 2)\}$ for each $t > 0$. Since $\Pr(U = 1/2) = \Pr(U = 1) = 0$

(Example 0.4):

$$\begin{aligned}
\Pr(T > 0) &= \Pr(1/2 < U < 1) && (\{T > 0\} = \{1/2 < U < 1\}) \\
&= \Pr(1/2 \leq U \leq 1) - \Pr(U = 1/2) - \Pr(U = 1) && \text{(additivity)} \\
&= \tfrac{1}{2} - 0 - 0 && \text{(definition of the uniform distribution)} \\
&= \tfrac{1}{2} && \text{(simplify)}
\end{aligned}$$

T is not continuous, because $\Pr(T = 0) = 1 - \Pr(T > 0) = 1/2$ by the complement rule. T is not discrete either. If it were, its range would be countable, so $\{T > 0\}$ would be the disjoint union of the countably many events $\{T = t\}$ with $t \in \mathcal{R}(T)$ and $t > 0$, and:

$$\begin{aligned}
\Pr(T > 0) &= \sum_{t \in \mathcal{R}(T),\, t > 0} \Pr(T = t) && \text{(countable additivity)} \\
&= \sum_{t \in \mathcal{R}(T),\, t > 0} \Pr\big(U = 1 - \tfrac{1}{t+2}\big) && (\{T = t\} = \{U = 1 - 1/(t+2)\}) \\
&= 0 && (U \text{ is continuous})
\end{aligned}$$

which contradicts $\Pr(T > 0) = 1/2$.

Characteristics of probability distributions

Definition 0.6 (Probability mass function (PMF))**.** If X is a discrete random variable, the **probability mass function** of X at value x, denoted $f(x)$, $f_X(x)$, $\mathrm{P}(x)$, $\mathrm{P}_X(x)$, or $\mathrm{P}(X = x)$, is the probability that X takes exactly the value x:

$$\mathrm{P}(X = x) \stackrel{\text{def}}{=} \Pr(\{X = x\})$$

Remark. See also https://en.wikipedia.org/wiki/Probability_mass_function

Example 0.6 (PMF of the number of heads)**.** For the number of heads X in two fair coin flips (Example 0.1):

$$\mathrm{P}(X = 0) = \tfrac{1}{4}, \quad \mathrm{P}(X = 1) = \tfrac{1}{2}, \quad \mathrm{P}(X = 2) = \tfrac{1}{4}$$

Definition 0.7 (Bernoulli distribution)**.** A discrete random variable X has the **Bernoulli distribution** with parameter $\pi \in [0, 1]$, written $X \sim \mathrm{Ber}(\pi)$, if:

$$
\begin{aligned}
\Pr(X = x) &\overset{\text{def}}{=} 1_{x \in \{0,1\}} \pi^x (1-\pi)^{1-x} \\
&= \begin{cases} \pi, & x = 1 \\ 1-\pi, & x = 0 \end{cases}
\end{aligned}
$$

Example 0.7 (The first of two coin flips)**.** In Example 0.1, let X_1 indicate heads on the first flip, so $X_1(HH) = X_1(HT) = 1$ and $X_1(TH) = X_1(TT) = 0$. Then $\mathrm{P}(X_1 = 1) = \Pr(\{HH, HT\}) = 1/2$ and $\mathrm{P}(X_1 = 0) = \Pr(\{TH, TT\}) = 1/2$, so $X_1 \sim \mathrm{Ber}(1/2)$.

Definition 0.8 (Probability density function (PDF))**.** If X is a continuous random variable, a **probability density function** of X, denoted $f(x)$, $f_X(x)$, $\mathrm{p}(x)$, $\mathrm{p}_X(x)$, or $\mathrm{p}(X = x)$, is a function f that satisfies:

- $f(x) \geq 0$ for every x.
- The integral of f over any interval is the probability that X falls in that interval:

$$
\Pr(a \leq X \leq b) = \int_a^b f(x)\,dx \quad \text{for all } a \leq b
$$

Remark. See also https://en.wikipedia.org/wiki/Probability_density_function#Formal_definition

Example 0.8 (Density of a uniform random variable)**.** Let $X \sim \mathrm{Uniform}(0,1)$ (Example 0.4), and let $f(x) = 1$ for $x \in [0,1]$ and $f(x) = 0$ otherwise. For $0 \leq a \leq b \leq 1$:

$$
\begin{aligned}
\int_a^b f(x)\,dx &= \int_a^b 1\,dx && (f = 1 \text{ on } [0,1]) \\
&= b - a && \text{(integrate)} \\
&= \Pr(a \leq X \leq b) && \text{(definition of the uniform distribution)}
\end{aligned}
$$

An interval reaching outside $[0,1]$ adds nothing to either side: $f = 0$ there, and X falls outside $[0,1]$ with probability 0 (Example 0.4). So f is a density of X.

Theorem 0.1 (A density is not unique)**.** *If f is a density of a continuous random variable X, and $g \geq 0$ differs from f at only finitely many points, then g is also a density of X.*

Proof

Proof. For $a \le b$, the difference $g - f$ is 0 at all but finitely many points, so:

$$\begin{aligned}
\int_a^b g(x)\,dx &= \int_a^b f(x)\,dx + \int_a^b (g(x) - f(x))\,dx && \text{(linearity of the integral)} \\
&= \int_a^b f(x)\,dx + 0 && \text{(a function that is 0 at all but finitely many points integrat} \\
&= \Pr(a \le X \le b) && \text{(}f\text{ is a density of }X\text{)}
\end{aligned}$$

Together with $g \ge 0$, this is Definition 0.8 for g. ☐

Remark. So the value of a density at a single point carries no probability by itself. These notes use the version that is continuous wherever possible.

Theorem 0.2 (Some continuous random variables have no density)**.** *There exist continuous random variables that have no probability density function.*

Remark. The proof needs measure theory beyond these notes. The standard example is the Cantor distribution, which puts no probability on any single point, but puts all of its probability on a set of total length 0, so any candidate density would integrate to 0 over a set of probability 1.
Every continuous random variable in these notes also has a probability density function. Continuous random variables without a density do not arise in applications.

Definition 0.9 (Normal distribution)**.** A random variable X has the **normal distribution** (or **Gaussian distribution**) with mean parameter $\mu \in \mathbb{R}$ and variance parameter $\sigma^2 > 0$, written $X \sim \mathrm{N}(\mu, \sigma^2)$, if X is continuous with density:

$$\mathrm{p}(X = x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

Example 0.9 (Normal densities at their centers)**.** For $X \sim \mathrm{N}(0, 1)$, the density at $x = 0$ is:

$$
\begin{aligned}
p(X=0) &= \frac{1}{1\cdot\sqrt{2\pi}} e^{-\frac{(0-0)^2}{2\cdot 1}} && \text{(normal density with } \mu=0, \sigma^2=1) \\
&= \frac{1}{\sqrt{2\pi}} && (e^0 = 1) \\
&\approx 0.399 && \text{(evaluate)}
\end{aligned}
$$

For $X \sim \mathrm{N}(0, 0.1^2)$, the same steps give $p(X=0) = 1/(0.1\sqrt{2\pi}) \approx 3.99$, a density greater than 1.

Remark. A density is not a probability: it can exceed 1, as the second density in Example 0.9 does.

Theorem 0.3 (The normal density integrates to 1, with mean μ and variance σ^2)**.** *If $X \sim \mathrm{N}(\mu, \sigma^2)$ (Definition 0.9), then:*

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1$$

and X has mean $\mathrm{E}[X] = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$.

Remark. The proof evaluates three integrals by calculus, one of them the Gaussian integral; it is omitted here (Casella and Berger 2002). These facts are where the parameters' names come from.

Definition 0.10 (Standard normal distribution)**.** The **standard normal distribution** is the normal distribution with mean parameter 0 and variance parameter 1, $\mathrm{N}(0,1)$.

Example 0.10 (Half of a standard normal lies below 0)**.** If Z has the standard normal distribution, its density $p(Z=z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ takes the same value at z and $-z$, so $\Pr(Z \le 0) = \Pr(Z \ge 0)$. Those two probabilities sum to 1, because $\Pr(Z=0) = 0$ for a continuous Z, so each is $\frac{1}{2}$.

Definition 0.11 (Exponential distribution)**.** A random variable T has the **exponential distribution** with rate $\lambda > 0$ if T is continuous with density:

$$f(t) \stackrel{\text{def}}{=} \begin{cases} \lambda e^{-\lambda t}, & t \ge 0 \\ 0, & t < 0 \end{cases}$$

Example 0.11 (Probability of an event within one time unit)**.** Let T be exponential with rate λ. The probability that T falls in $[0,1]$ is:

$$\begin{aligned}
\Pr(0 \le T \le 1) &= \int_0^1 \lambda \mathrm{e}^{-\lambda t}\, dt && \text{(definition of a density)} \\
&= \left[-\mathrm{e}^{-\lambda t}\right]_{t=0}^{1} && \text{(antiderivative of } \lambda \mathrm{e}^{-\lambda t}\text{)} \\
&= 1 - \mathrm{e}^{-\lambda} && \text{(evaluate at the bounds)}
\end{aligned}$$

With $\lambda = 0.5$, for example, $\Pr(0 \le T \le 1) = 1 - \mathrm{e}^{-0.5} \approx 0.393$.

Definition 0.12 (Cumulative distribution function (CDF))**.** For a random variable X (discrete or continuous), its **cumulative distribution function** is:

$$F(t) \overset{\text{def}}{=} \Pr(X \le t), \quad t \in \mathbb{R}.$$

Remark. The CDF is always defined, whether X is discrete or continuous.
See also https://en.wikipedia.org/wiki/Cumulative_distribution_function

Example 0.12 (CDF of the number of heads)**.** Summing the PMF in Example 0.6 over values at or below t:

$$F(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{4}, & 0 \le t < 1 \\ \frac{3}{4}, & 1 \le t < 2 \\ 1, & t \ge 2 \end{cases}$$

Theorem 0.4 (Properties of a CDF)**.** *The CDF F of any random variable X is non-decreasing, and:*

$$\lim_{t \to -\infty} F(t) = 0, \qquad \lim_{t \to \infty} F(t) = 1$$

Proof

Proof. Non-decreasing. For $s < t$, $\{X \le t\}$ is the disjoint union of $\{X \le s\}$ and $\{s < X \le t\}$, so finite additivity gives:

$$
\begin{aligned}
F(t) &= \Pr(X \le s) + \Pr(s < X \le t) && \text{(additivity, and the definition of the CDF)} \\
&\ge \Pr(X \le s) && \text{(probabilities are non-negative)} \\
&= F(s) && \text{(definition of the CDF)}
\end{aligned}
$$

Limit at $-\infty$. The events $A_n = \{X \le -n\}$, $n = 1, 2, \ldots$, are decreasing, and their intersection is empty, because every outcome ω has $X(\omega) > -n$ once $n > -X(\omega)$. By continuity of probability, $F(-n) = \Pr(A_n) \to 0$. For $t \le -n$, $0 \le F(t) \le F(-n)$, because F is non-decreasing, so $F(t) \to 0$ as $t \to -\infty$.
Limit at ∞. The events $B_n = \{X > n\}$ are decreasing, and their intersection is empty, because every outcome ω has $X(\omega) \le n$ once $n \ge X(\omega)$. By continuity of probability, $\Pr(B_n) \to 0$, so, by the complement rule:

$$
\begin{aligned}
F(n) &= 1 - \Pr(X > n) && \text{(complement rule)} \\
&\to 1 - 0 && (\Pr(B_n) \to 0)
\end{aligned}
$$

For $t \ge n$, $F(n) \le F(t) \le 1$, because F is non-decreasing, so $F(t) \to 1$ as $t \to \infty$. □

Theorem 0.5 (The CDF determines the distribution)**.** *If random variables X and Y have the same CDF, then* $\Pr(X \in B) = \Pr(Y \in B)$ *for every Borel set $B \subseteq \mathbb{R}$, which includes every interval, and every set built from intervals by countably many unions, intersections, and complements.*

Remark. The proof needs measure theory beyond these notes (Billingsley 1995); see also https://en.wikipedia.org/wiki/Cumulative_distribution_function.

Definition 0.13 (Quantile function (population inverse CDF))**.** For a random variable X with cumulative distribution function (CDF) F, its population **quantile function** (generalized inverse of F) is:

$$Q(p) \stackrel{\text{def}}{=} \inf\{t : F(t) \ge p\}, \quad 0 < p < 1.$$

Remark. The restriction to $0 < p < 1$ keeps $Q(p)$ finite: for $p = 1$, the set $\{t : F(t) \ge 1\}$ is empty whenever $F(t) < 1$ for every t, as for a normal distribution, and the infimum of the empty set is $+\infty$.

Example 0.13 (Median of the number of heads)**.** From Example 0.12, $F(t) \ge 0.5$ first holds at $t = 1$, so the median of the number of heads is $Q(0.5) = 1$.

Theorem 0.6 (A density integrates to the CDF, and is its derivative)**.** *If X is a continuous random variable with density f and CDF F, then for every t:*

$$F(t) = \int_{-\infty}^{t} f(x)\,dx$$

and at every t where f is continuous:

$$f(t) = \frac{\partial}{\partial t} F(t)$$

Proof

Proof. For any $a < t$, $\{X \leq t\}$ is the disjoint union of $\{X \leq a\}$ and $\{a < X \leq t\}$, and $\Pr(X = a) = 0$ because X is continuous. So:

$$\begin{aligned} F(t) &= \Pr(X \leq a) + \Pr(a < X \leq t) && \text{(additivity, and the definition of the CDF)} \\ &= F(a) + \Pr(a \leq X \leq t) && \text{(definition of the CDF; } \Pr(X = a) = 0) \\ &= F(a) + \int_a^t f(x)\,dx && \text{(definition of the density)} \end{aligned}$$

Letting $a \to -\infty$, $F(a) \to 0$ (Theorem 0.4), which gives $F(t) = \int_{-\infty}^{t} f(x)\,dx$. The fundamental theorem of calculus then gives $\frac{\partial}{\partial t} \int_{-\infty}^{t} f(x)\,dx = f(t)$ at every t where f is continuous. ☐

Corollary 0.1 (A CDF has its density as derivative at all but finitely many points)**.** *If X is a continuous random variable with CDF F and a density f that is continuous at all but finitely many points, then F is differentiable with $F' = f$ at all but finitely many points.*

Proof

Proof. At every point t where f is continuous, Theorem 0.6 gives $F'(t) = f(t)$, and by assumption those are all but finitely many points. ☐

Remark. Every density in these notes is continuous at all but finitely many points, so Corollary 0.1 applies to each of them.

Example 0.14 (CDF and density of a uniform random variable)**.** For X uniform on $[0, 1]$ with the density f of Example 0.8, and $t \in [0, 1]$:

$$
\begin{aligned}
F(t) &= \int_{-\infty}^{t} f(x)\,dx && \text{(the CDF is the integral of the density)} \\
&= \int_{-\infty}^{0} 0\,dx + \int_{0}^{t} 1\,dx && \text{(split at 0, and substitute } f\text{)} \\
&= t && \text{(integrate)}
\end{aligned}
$$

For $t \in (0,1)$, where f is continuous, $\frac{\partial}{\partial t}F(t) = 1 = f(t)$. At $t = 1$, f jumps from 1 to 0, and F has no derivative: its slope is 1 on the left and 0 on the right.

Theorem 0.7 (A piecewise-smooth CDF has its derivative as a density)**.** *If the CDF F of a random variable X is continuous everywhere, and has a continuous derivative at all but finitely many points, then X is continuous, and F' (given any values at the finitely many exceptional points) is a density of X.*

Proof

Proof. For events $A \subseteq B$, B is the disjoint union of A and $B \setminus A$, so additivity gives $\Pr(A) \le \Pr(B)$. For any x and $\Delta > 0$, $\{X = x\} \subseteq \{x - \Delta < X \le x\}$, so:

$$
\begin{aligned}
0 &\le \Pr(X = x) && \text{(probabilities are non-negative)} \\
&\le \Pr(x - \Delta < X \le x) && (\{X = x\} \subseteq \{x - \Delta < X \le x\}) \\
&= F(x) - F(x - \Delta) && \text{(additivity, and the definition of the CDF)}
\end{aligned}
$$

and $F(x) - F(x - \Delta) \to 0$ as $\Delta \downarrow 0$, because F is continuous. So $\Pr(X = x) = 0$ for every x, and X is continuous. Then, for $a \le b$:

$$
\begin{aligned}
\Pr(a \le X \le b) &= \Pr(a < X \le b) && (\Pr(X = a) = 0) \\
&= F(b) - F(a) && \text{(additivity, and the definition of the CDF)} \\
&= \int_{a}^{b} F'(x)\,dx && \text{(fundamental theorem of calculus)}
\end{aligned}
$$

The last step applies the fundamental theorem of calculus on each piece of $[a, b]$ between exceptional points, where F' is continuous, and joins the pieces using the continuity of F at the exceptional points. Finally, $F' \ge 0$ wherever it exists, because F is non-decreasing (Theorem 0.4). □

Example 0.15 (A density from a CDF with a corner)**.** Let X have CDF $F(t) = 0$ for $t < 0$, $F(t) = t^2$ for $0 \le t \le 1$, and $F(t) = 1$ for $t > 1$. This F is continuous everywhere, and has a continuous derivative everywhere except $t = 1$, where its slope drops from 2 to 0. By Theorem 0.7, X is continuous with density $f(t) = 2t$ for $0 \le t < 1$ and $f(t) = 0$

otherwise. As a check, $\int_0^1 2t\,dt = 1$.

Theorem 0.8 (A density as a limit of interval probabilities)**.** *If X is a continuous random variable with density f, and f is continuous at x, then $f(x)$ is the limit of the probability that X falls in an interval starting at x, divided by the width of that interval, as that width shrinks to 0:*

$$f(x) = \lim_{\Delta \downarrow 0} \frac{\Pr(x \leq X < x + \Delta)}{\Delta}$$

Proof

Proof. For $\Delta > 0$:

$$\begin{aligned} \frac{\Pr(x \leq X < x + \Delta)}{\Delta} &= \frac{\Pr(x < X \leq x + \Delta)}{\Delta} && (\Pr(X = x) = \Pr(X = x + \Delta) = 0) \\ &= \frac{F(x + \Delta) - F(x)}{\Delta} && \text{(additivity, and the definition of the CDF)} \end{aligned}$$

By Theorem 0.6, F is differentiable at x with $F'(x) = f(x)$, so this difference quotient converges to $f(x)$ as $\Delta \downarrow 0$. □

Remark. For small $\Delta > 0$, $f(x) \cdot \Delta \approx \Pr(x \leq X < x + \Delta)$. Although a density is not unique (Theorem 0.1), this limit pins down its value at every point where it is continuous. See also Rothman et al. (2021) (Chapter 22, p. 535).

Example 0.16 (The uniform density as a limit)**.** For $X \sim \text{Uniform}(0, 1)$ and $x \in [0, 1)$, $\Pr(x \leq X < x + \Delta) = \Delta$ once $\Delta \leq 1 - x$, so:

$$f(x) = \lim_{\Delta \downarrow 0} \frac{\Delta}{\Delta} = 1$$

For $x < 0$ or $x > 1$, the interval eventually misses $[0, 1]$, so the limit is 0. Both agree with the density of Example 0.8.

Proposition 0.1 (The density limit is the right derivative of the CDF)**.** *If X is a continuous random variable with CDF F, then for every x:*

$$\lim_{\Delta \downarrow 0} \frac{\Pr(x \leq X < x + \Delta)}{\Delta} = \lim_{\Delta \downarrow 0} \frac{F(x + \Delta) - F(x)}{\Delta}$$

whenever either limit exists; the right-hand side is the right derivative of F at x.

Proof

Proof. The first two lines of the proof of Theorem 0.8 use only that X is continuous, so for every $\Delta > 0$ the two quotients are equal. Equal functions of Δ have the same limit as $\Delta \downarrow 0$, or both have none. ☐

Example 0.17 (The density limit at a jump)**.** At a point where f is not continuous, the limit can differ from the chosen value $f(x)$. For $X \sim \text{Uniform}(0,1)$ with the density f of Example 0.8, which sets $f(1) = 1$ and $f(x) = 0$ for $x > 1$, and for $\Delta > 0$:

$$\begin{aligned}
\Pr(1 \le X < 1+\Delta) &= \Pr(1 \le X \le 1+\Delta) && (\Pr(X = 1+\Delta) = 0) \\
&= \int_1^{1+\Delta} f(x)\,dx && (\text{definition of the density}) \\
&= 0 && (f = 0 \text{ on } (1, 1+\Delta]\text{; one point does not change an integral})
\end{aligned}$$

So the limit of $\Pr(1 \le X < 1+\Delta)/\Delta$ at $x = 1$ is 0, although $f(1) = 1$. By Proposition 0.1, 0 is also the right derivative of F at 1, matching the slope of 0 on the right in Example 0.14.

Theorem 0.9 (Density functions integrate to 1)**.** *If X is a continuous random variable with density f:*

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Proof

Proof.

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \lim_{b\to\infty} \int_{-\infty}^{b} f(x)\,dx && (\text{definition of an improper integral}) \\
&= \lim_{b\to\infty} F(b) && (\text{the CDF is the integral of the density}) \\
&= 1 && (\text{limit of a CDF})
\end{aligned}$$

The last step is Theorem 0.4. ☐

Example 0.18 (The uniform density integrates to 1)**.** For X uniform on $[0, 1]$ (Example 0.8), $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 1\,dx = 1$.

Definition 0.14 (Jointly distributed random variables)**.** Random variables $X_1, \dots, X_n$ are **jointly distributed** when they are defined on the same sample space Ω, so that an event about several of them at once, such as $\{X_1 \le x_1\} \cap \{X_2 \le x_2\}$, has a probability.

Remark. We write $\Pr(X \in A, Y \in B)$ for $\Pr(\{X \in A\} \cap \{Y \in B\})$, and similarly for more variables.

Example 0.19 (The first flip and the total)**.** In Example 0.1, let X_1 indicate heads on the first flip (as in Example 0.7), and let X be the total number of heads. Both are functions on the same sample space, $\{HH, HT, TH, TT\}$, so they are jointly distributed, and, for example:

$$\Pr(X_1 = 1, X = 1) = \Pr(\{HH, HT\} \cap \{HT, TH\}) = \Pr(\{HT\}) = \tfrac{1}{4}$$

Definition 0.15 (Joint distribution)**.** The **joint distribution** of jointly distributed random variables $X_1, \dots, X_n$ is the function that assigns to each set $A \subseteq \mathbb{R}^n$ for which $\{(X_1, \dots, X_n) \in A\}$ is an event the probability $\Pr((X_1, \dots, X_n) \in A)$.

Example 0.20 (Joint distribution of the first flip and the total)**.** In Example 0.19, the joint distribution of X_1 and X assigns to the set $A = \{(x_1, x) : x_1 = 1,\ x \ge 1\}$ the probability

$$\Pr((X_1, X) \in A) = \Pr(\{HH, HT\}) = \tfrac{1}{2}$$

since the first flip is heads exactly on HH and HT, and each of those outcomes has at least one head.

Theorem 0.10 (A joint distribution is determined by its rectangle probabilities)**.** *If jointly distributed random variables $X_1, \dots, X_n$, and jointly distributed random variables $Y_1, \dots, Y_n$, satisfy*

$$\Pr(X_1 \in A_1, \dots, X_n \in A_n) = \Pr(Y_1 \in A_1, \dots, Y_n \in A_n)$$

for all sets $A_1, \dots, A_n$ of real numbers for which these are events, then $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ have the same joint distribution.

Remark. The proof needs measure theory beyond these notes: the sets $A_1 \times \cdots \times A_n$ form a π-system, and two probability measures that agree on a π-system agree on the σ-algebra it generates (Billingsley 1995).

Definition 0.16 (Joint probability mass function)**.** For jointly distributed discrete random variables X and Y, the **joint probability mass function** (joint PMF) of X and Y is the probability that X takes the value x and Y takes the value y:

$$\mathrm{P}(X = x,\, Y = y) \stackrel{\text{def}}{=} \Pr(\{X = x\} \cap \{Y = y\})$$

Example 0.21 (Joint PMF of the first flip and the total)**.** Continuing Example 0.19, each outcome of the two flips fixes both X_1 and X, so the joint PMF puts probability 1/4 on each outcome's pair of values:

	$X = 0$	$X = 1$	$X = 2$
$X_1 = 0$	$1/4$ (TT)	$1/4$ (TH)	0
$X_1 = 1$	0	$1/4$ (HT)	$1/4$ (HH)

Definition 0.17 (Marginal distribution)**.** When a random variable X is one of several jointly distributed random variables, the distribution of X by itself (its CDF, and its PMF or density) is called the **marginal distribution** of X.

Remark. The word "marginal" only says that the other variables are being set aside; the marginal distribution of X is the same object as the distribution of X.

Example 0.22 (Marginal distribution of the total)**.** In Example 0.21, the marginal distribution of the total X is the PMF of Example 0.6: $\mathrm{P}(X = 0) = 1/4$, $\mathrm{P}(X = 1) = 1/2$, $\mathrm{P}(X = 2) = 1/4$.

Theorem 0.11 (Marginal PMF from a joint PMF)**.** *If X and Y are jointly distributed discrete random variables, then for every x:*

$$\mathrm{P}(X = x) = \sum_{y \in \mathcal{R}(Y)} \mathrm{P}(X = x,\, Y = y)$$

Proof

Proof. The event $\{X = x\}$ is the disjoint union of the events $\{X = x\} \cap \{Y = y\}$ over the countably many values $y \in \mathcal{R}(Y)$, so:

$$\begin{aligned}
P(X = x) &= \Pr\left(\bigcup_{y \in \mathcal{R}(Y)} (\{X = x\} \cap \{Y = y\})\right) && \text{(the events partition } \{X = x\}\text{)} \\
&= \sum_{y \in \mathcal{R}(Y)} \Pr(\{X = x\} \cap \{Y = y\}) && \text{(countable additivity)} \\
&= \sum_{y \in \mathcal{R}(Y)} P(X = x,\, Y = y) && \text{(definition of the joint PMF)}
\end{aligned}$$

□

Example 0.23 (Summing a row of the joint PMF)**.** In Example 0.21, summing the $X_1 = 1$ row:

$$\begin{aligned}
P(X_1 = 1) &= P(X_1 = 1, X = 0) + P(X_1 = 1, X = 1) + P(X_1 = 1, X = 2) && \text{(marginal PMF from the joint PM} \\
&= 0 + \tfrac{1}{4} + \tfrac{1}{4} && \text{(read the table)} \\
&= \tfrac{1}{2} && \text{(add)}
\end{aligned}$$

which matches $X_1 \sim \text{Ber}(1/2)$ from Example 0.7.

Definition 0.18 (Joint probability density function)**.** For jointly distributed continuous random variables X and Y, a **joint probability density function** (joint density) of X and Y, denoted $f_{X,Y}(x, y)$ or $p(X = x,\, Y = y)$, is a function $f_{X,Y}$ on $\mathbb{R}^2$ that satisfies:

- $f_{X,Y}(x, y) \geq 0$ for every (x, y).
- The integral of $f_{X,Y}$ over any region $A \subseteq \mathbb{R}^2$ is the probability that the pair (X, Y) falls in A:
$$\Pr((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\, dx\, dy$$

Remark. As with events, a fully rigorous version restricts A to a designated collection of regions; every region that arises in these notes is in it.

Example 0.24 (A joint density on a triangle)**.** Let $f_{X,Y}(x, y) = 2$ for $0 \leq x \leq y \leq 1$, and 0 otherwise. The triangle $\{(x, y) : 0 \leq x \leq y \leq 1\}$ has area $1/2$, so $f_{X,Y}$ gives the whole plane probability $2 \cdot \frac{1}{2} = 1$, and it is the joint density of a pair (X, Y) with $X \le Y$ always. For example, the probability that both are at most $1/2$ is 2 times the area of the smaller triangle $\{0 \le x \le y \le 1/2\}$:

$$\Pr(X \le \tfrac{1}{2}, Y \le \tfrac{1}{2}) = 2 \cdot \tfrac{1}{8} = \tfrac{1}{4}$$

Theorem 0.12 (Not every pair of continuous random variables has a joint density)**.** *If X is a continuous random variable, then the pair (X, X) has no joint density.*

Proof

Proof. Suppose f were a joint density of (X, X), and let $L = \{(x, y) : y = x\}$ be the diagonal line. Every outcome ω has $X(\omega) = X(\omega)$, so $\{(X, X) \in L\} = \Omega$. For each x, the function $y \mapsto 1_{y=x} f(x, y)$ is 0 except at the single point $y = x$, so its integral over y is 0. Then:

$$\begin{aligned}
1 &= \Pr((X, X) \in L) && (\{(X, X) \in L\} = \Omega, \text{ and } \Pr(\Omega) = 1) \\
&= \iint_L f(x, y)\, dx\, dy && \text{(definition of a joint density)} \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} 1_{y=x} f(x, y)\, dy \right) dx && \text{(iterate the integral; Tonelli's theorem)} \\
&= \int_{-\infty}^{\infty} 0\, dx && \text{(the inner integrand is 0 except at } y = x\text{)} \\
&= 0 && \text{(integrate)}
\end{aligned}$$

which is a contradiction. Tonelli's theorem allows the iterated integral because $f \ge 0$ (Fubini–Tonelli theorem; Billingsley (1995), Theorem 18.3). □

Theorem 0.13 (Marginal density from a joint density)**.** *If X and Y have joint density $f_{X,Y}$, then X has the density:*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

Proof

Proof. For $a \le b$, the event $\{a \le X \le b\}$ is the event that (X, Y) falls in the strip $[a, b] \times \mathbb{R}$, so:

$$
\begin{aligned}
\Pr(a \le X \le b) &= \Pr((X,Y) \in [a,b] \times \mathbb{R}) && \text{(same event)} \\
&= \iint_{[a,b]\times\mathbb{R}} f_{X,Y}(x,y)\,dx\,dy && \text{(definition of a joint density)} \\
&= \int_a^b \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \right) dx && \text{(iterate the integral; Tonelli's theorem)} \\
&= \int_a^b f_X(x)\,dx && \text{(definition of } f_X\text{)}
\end{aligned}
$$

Tonelli's theorem allows the iterated integral because $f_{X,Y} \ge 0$ (Fubini–Tonelli theorem; Billingsley (1995), Theorem 18.3). So f_X satisfies Definition 0.8. □

Example 0.25 (Marginal densities on the triangle)**.** For the joint density of Example 0.24 and $x \in [0,1]$, $f_{X,Y}(x,y) = 2$ exactly when $x \le y \le 1$, so:

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy && \text{(marginal density from a joint density)} \\
&= \int_x^1 2\,dy && (f_{X,Y}(x,y) = 2 \text{ for } x \le y \le 1, \text{ else } 0) \\
&= 2(1-x) && \text{(integrate)}
\end{aligned}
$$

The same steps with the roles swapped give $f_Y(y) = \int_0^y 2\,dx = 2y$ for $y \in [0,1]$.

Definition 0.19 (Joint density-mass function)**.** Let X be a discrete random variable and Y a continuous random variable, jointly distributed. A **joint density-mass function** of X and Y, denoted $\mathrm{p}(X=x,\, Y=y)$, is a function $\mathrm{p}(X=x,\, Y=y) \ge 0$ that is a probability mass in x and a probability density in y: for every x and every set $B \subseteq \mathbb{R}$,

$$
\Pr(X = x,\, Y \in B) = \int_B \mathrm{p}(X = x,\, Y = y)\,dy
$$

Remark. When X is continuous and Y is discrete, the roles swap: $\Pr(X \in B,\, Y = y) = \int_B \mathrm{p}(X = x,\, Y = y)\,dx$.

Corollary 0.2 (Marginal PMF from a joint density-mass function)**.** *If X and Y have joint density-mass function $\mathrm{p}(X=x,\, Y=y)$, then for every x:*

$$P(X = x) = \int_{-\infty}^{\infty} p(X = x, Y = y)\, dy$$

Proof

Proof. Every outcome has $Y(\omega) \in \mathbb{R}$, so $\{Y \in \mathbb{R}\} = \Omega$, and:

$$\begin{aligned} P(X = x) &= \Pr(\{X = x\} \cap \Omega) && (\{X = x\} \subseteq \Omega) \\ &= \Pr(X = x, Y \in \mathbb{R}) && (\{Y \in \mathbb{R}\} = \Omega) \\ &= \int_{-\infty}^{\infty} p(X = x, Y = y)\, dy && (\text{definition of a joint density-mass function, with } B = \mathbb{R}) \end{aligned}$$

□

Example 0.26 (A coin flip and a waiting time)**.** Let $p(X = 0, Y = y) = 1/2$ for $y \in [0, 1]$, $p(X = 1, Y = y) = 1/4$ for $y \in [0, 2]$, and 0 otherwise. Integrating over y gives $P(X = 0) = 1/2$ and $P(X = 1) = 1/2$, which add to 1. For example, the probability that $X = 1$ and $Y \leq 1$ is:

$$\Pr(X = 1, Y \leq 1) = \int_0^1 \tfrac{1}{4}\, dy = \tfrac{1}{4}$$

Video lecture

Hutchinson's Probability Refresher (27 min) covers probability mass functions and probability density functions (Hutchinson, n.d.). The login for the video site is posted on Canvas.

Survival, hazard, and cumulative hazard functions

Definition 0.20 (Survival function)**.** The **survival function** (or **survivor function**) of a random variable T, denoted $S(t)$, is the probability that T exceeds t:

$$S(t) \stackrel{\text{def}}{=} \Pr(T > t)$$

Remark. The name comes from time-to-event analysis: if T is the time at which a participant dies, then $S(t)$ is the probability that the participant is still alive at time t.

The definition itself applies to any random variable.

Theorem 0.14 (Survival function and CDF)**.** *For any random variable T with CDF $F(t)$:*

$$\mathrm{S}(t) = 1 - F(t)$$

If T is continuous with density $f(t)$, then also:

$$\mathrm{S}(t) = \int_{u=t}^{\infty} f(u)\, du$$

Proof

Proof. The event $\{T > t\}$ is the complement of the event $\{T \le t\}$, so the complement rule gives:

$$\begin{aligned} \mathrm{S}(t) &\stackrel{\text{def}}{=} \Pr(T > t) && \text{(definition of the survival function)} \\ &= 1 - \Pr(T \le t) && \text{(complement rule)} \\ &= 1 - F(t) && \text{(definition of the CDF)} \end{aligned}$$

For continuous T, the density integrates to 1 over the real line (Theorem 0.9), and F is the integral of the density up to t (Theorem 0.6), so:

$$\begin{aligned} 1 - F(t) &= \int_{u=-\infty}^{\infty} f(u)\, du - \int_{u=-\infty}^{t} f(u)\, du && \text{(total integral is 1; } F \text{ is the integral of } f\text{)} \\ &= \int_{u=t}^{\infty} f(u)\, du && \text{(split the first integral at } t \text{ and cancel)} \end{aligned}$$

□

Example 0.27 (Survival function of an exponential distribution)**.** Let T be exponential with rate $\lambda > 0$ (Definition 0.11), so T has density $f(t) = \lambda e^{-\lambda t}$ for $t \ge 0$. For $t \ge 0$, by Theorem 0.14:

$$
\begin{aligned}
\mathrm{S}(t) &= \int_{u=t}^{\infty} \lambda \mathrm{e}^{-\lambda u}\, du && \text{(integral form of the survival function)} \\
&= \left[-\mathrm{e}^{-\lambda u}\right]_{u=t}^{\infty} && \text{(antiderivative of } \lambda \mathrm{e}^{-\lambda u}\text{)} \\
&= 0 - (-\mathrm{e}^{-\lambda t}) && \text{(evaluate at the bounds; } \mathrm{e}^{-\lambda u} \to 0 \text{ as } u \to \infty\text{)} \\
&= \mathrm{e}^{-\lambda t} && \text{(simplify)}
\end{aligned}
$$

For $t < 0$, $\mathrm{S}(t) = \Pr(T > t) = 1$, since $T \geq 0$. With $\lambda = 0.5$, for example, $\mathrm{S}(2) = \mathrm{e}^{-1} \approx 0.368$.

Definition 0.21 (Hazard function)**.** The **hazard function** (also called the **hazard rate** or **hazard rate function**) of a continuous random variable T at a value t with $\Pr(T \geq t) > 0$, typically denoted $\lambda(t)$ or $\mathrm{h}(t)$, is the limit of the conditional probability that T falls in an interval starting at t, given $T \geq t$, divided by the width of that interval, as that width shrinks to 0:

$$
\lambda(t) \stackrel{\text{def}}{=} \lim_{\Delta \downarrow 0} \frac{\Pr(t \leq T < t + \Delta \mid T \geq t)}{\Delta}
$$

Remark. Sources differ on the symbol: $\mathrm{h}(t)$ appears in Dobson and Barnett (2018), Vittinghoff et al. (2012), Klein and Moeschberger (2003), and Kleinbaum and Klein (2012), while $\lambda(t)$ appears in Rothman et al. (2021) and Kalbfleisch and Prentice (2011). If T is the time at which an event occurs, then $\lambda(t)$ is a rate, not a probability: for a small interval width $\Delta > 0$, $\lambda(t) \cdot \Delta$ is approximately the probability that the event occurs in $[t, t + \Delta)$, given that it has not occurred before t. Many sources write the hazard as $\lambda(t) = \mathrm{p}(T = t \mid T \geq t)$, reading it as the density of T at t, conditional on the event $T \geq t$; that notation abbreviates the limit in Definition 0.21. For a discrete T, the analogous quantity is the discrete-time hazard.

The name "hazard" carries a connotation that the event is undesirable — death, relapse, equipment failure, and so on. When the event in question is neutral or desirable (recovery, conception, graduation, response to treatment), the same quantity $\lambda(t)$ is often called the **event incidence rate** instead. This terminology parallels the convention that conditional probabilities of undesirable events are called **risks**, while the same conditional probabilities for neutral or desirable events are simply called **probabilities**. The math is identical; only the name changes with the valence of the event.

Example 0.28 (A hazard greater than 1)**.** A hazard can exceed 1, just as a density can (Example 0.9). Let T be exponential with rate $\lambda = 2$, so $\mathrm{S}(t) = \mathrm{e}^{-2t}$ for $t \geq 0$ (Example 0.27). Since T is continuous, $\Pr(T = s) = 0$ for every s, so $\Pr(T \geq s) = \Pr(T > s) = \mathrm{S}(s)$. For $t \geq 0$ and $\Delta > 0$, $\{T \geq t\}$ is the disjoint union of $\{t \leq T < t + \Delta\}$ and

$\{T \geq t + \Delta\}$, so:

$$\begin{aligned}
\Pr(t \leq T < t + \Delta \mid T \geq t) &= \frac{\Pr(\{t \leq T < t + \Delta\} \cap \{T \geq t\})}{\Pr(T \geq t)} && \text{(definition of conditional probability)} \\
&= \frac{\Pr(t \leq T < t + \Delta)}{\Pr(T \geq t)} && \text{(subset property)} \\
&= \frac{\Pr(T \geq t) - \Pr(T \geq t + \Delta)}{\Pr(T \geq t)} && \text{(additivity)} \\
&= \frac{\mathrm{S}(t) - \mathrm{S}(t + \Delta)}{\mathrm{S}(t)} && (\Pr(T \geq s) = \mathrm{S}(s)) \\
&= \frac{\mathrm{e}^{-2t} - \mathrm{e}^{-2(t+\Delta)}}{\mathrm{e}^{-2t}} && \text{(exponential survival function)} \\
&= 1 - \mathrm{e}^{-2\Delta} && \text{(divide by } \mathrm{e}^{-2t}\text{)}
\end{aligned}$$

The subset property applies because $\{t \leq T < t + \Delta\} \subseteq \{T \geq t\}$. Then:

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta \downarrow 0} \frac{1 - \mathrm{e}^{-2\Delta}}{\Delta} && \text{(definition of the hazard function)} \\
&= \frac{\partial}{\partial \Delta}(1 - \mathrm{e}^{-2\Delta})\Big|_{\Delta=0} && \text{(definition of the derivative; } 1 - \mathrm{e}^0 = 0\text{)} \\
&= 2\mathrm{e}^0 && \text{(chain rule)} \\
&= 2 && (\mathrm{e}^0 = 1)
\end{aligned}$$

So $\lambda(t) = 2 > 1$ at every $t \geq 0$: a hazard is a rate, not a probability.

Definition 0.22 (Discrete-time hazard)**.** The **discrete-time hazard** of a discrete random variable T at a value t with $\Pr(T \geq t) > 0$ is the conditional probability that T equals t, given $T \geq t$:

$$\Pr(T = t \mid T \geq t)$$

Example 0.29 (Rolling until the first six)**.** Roll a fair die repeatedly, and let T be the number of the roll that first shows a six. Then $T \geq t$ exactly when the first $t - 1$ rolls are not sixes, which has probability $(5/6)^{t-1}$, and $T = t$ when, in addition, roll t is a six. So, for each $t = 1, 2, \ldots$:

$$
\begin{aligned}
\Pr(T = t \mid T \geq t) &= \frac{\Pr(T = t,\, T \geq t)}{\Pr(T \geq t)} && \text{(definition of conditional probability)} \\
&= \frac{\Pr(T = t)}{\Pr(T \geq t)} && (T = t \text{ implies } T \geq t) \\
&= \frac{(5/6)^{t-1} \cdot (1/6)}{(5/6)^{t-1}} && \text{(rolls are independent)} \\
&= \frac{1}{6} && \text{(cancel)}
\end{aligned}
$$

The discrete-time hazard is the same at every roll: having gone without a six so far does not change the chance of a six on the next roll.

Corollary 0.3 (A discrete-time hazard lies in $[0, 1]$)**.** *The discrete-time hazard of a discrete random variable T at a value t with $\Pr(T \geq t) > 0$ satisfies:*

$$0 \leq \Pr(T = t \mid T \geq t) \leq 1$$

Proof

Proof. $\{T \geq t\}$ is the disjoint union of $\{T = t\}$ and $\{T > t\}$, so, by additivity, $\Pr(T \geq t) = \Pr(T = t) + \Pr(T > t) \geq \Pr(T = t)$. Since $\{T = t\} \subseteq \{T \geq t\}$:

$$
\begin{aligned}
\Pr(T = t \mid T \geq t) &= \frac{\Pr(\{T = t\} \cap \{T \geq t\})}{\Pr(T \geq t)} && \text{(definition of conditional probability)} \\
&= \frac{\Pr(T = t)}{\Pr(T \geq t)} && \text{(subset property)}
\end{aligned}
$$

The numerator is non-negative and at most the positive denominator, so the ratio lies in $[0, 1]$. □

Remark. Unlike the continuous-time hazard function, which is a rate and can exceed 1 (Example 0.28), the discrete-time hazard is a probability.

Theorem 0.15 (Hazard equals density over survival)**.** *If T is a continuous random variable with density $f(t)$ and survival function $\mathrm{S}(t)$, then for every t with $\mathrm{S}(t) > 0$ at which f is continuous:*

$$\lambda(t) = \frac{f(t)}{\mathrm{S}(t)}$$

ℹ Proof

Proof. The proof uses three facts. First, for $\Delta > 0$ the event $\{t \le T < t+\Delta\}$ is a subset of the event $\{T \ge t\}$, so intersecting them leaves $\{t \le T < t+\Delta\}$ (the subset property). Second, $\Pr(T = t) = 0$ for a continuous T, so $\Pr(T \ge t) = \Pr(T > t) = \mathrm{S}(t)$, which is positive. Third, because f is continuous at t, Theorem 0.8 gives $f(t)$ as the limit of $\Pr(t \le T < t+\Delta)/\Delta$.

$$
\begin{aligned}
\lambda(t) &\overset{\text{def}}{=} \lim_{\Delta \downarrow 0} \frac{\Pr(t \le T < t+\Delta \mid T \ge t)}{\Delta} && \text{(definition of the hazard function)}\\
&= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \cdot \frac{\Pr(\{t \le T < t+\Delta\} \cap \{T \ge t\})}{\Pr(T \ge t)} && \text{(definition of conditional probability)}\\
&= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \cdot \frac{\Pr(t \le T < t+\Delta)}{\Pr(T \ge t)} && \text{(subset property)}\\
&= \frac{1}{\Pr(T \ge t)} \cdot \lim_{\Delta \downarrow 0} \frac{\Pr(t \le T < t+\Delta)}{\Delta} && \text{(}\Pr(T \ge t)\text{ does not depend on }\Delta\text{)}\\
&= \frac{f(t)}{\Pr(T \ge t)} && \text{(density as a limit; }f\text{ is continuous at }t\text{)}\\
&= \frac{f(t)}{\mathrm{S}(t)} && \text{(}\Pr(T = t) = 0\text{ for continuous }T\text{)}
\end{aligned}
$$

□

Example 0.30 (Hazard function of an exponential distribution)**.** Continuing Example 0.27, for $t > 0$, where f is continuous:

$$
\begin{aligned}
\lambda(t) &= \frac{f(t)}{\mathrm{S}(t)} && \text{(hazard equals density over survival)}\\
&= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} && \text{(substitute the exponential density and survival function)}\\
&= \lambda && \text{(cancel } e^{-\lambda t}\text{)}
\end{aligned}
$$

At $t = 0$, where f jumps from 0 to λ, Definition 0.21 gives the same value directly: $\Pr(0 \le T < \Delta \mid T \ge 0)/\Delta = \left(1 - e^{-\lambda\Delta}\right)/\Delta \to \lambda$ as $\Delta \downarrow 0$. So the exponential distribution has a constant hazard for $t \ge 0$, equal to its rate parameter; for $t < 0$, $f(t) = 0$, so $\lambda(t) = 0$.

Definition 0.23 (Cumulative hazard function)**.** The **cumulative hazard function** of a continuous random variable T, often denoted $\Lambda(t)$ or $\mathrm{H}(t)$, is the integral of its hazard function up to t:

$$\Lambda(t) \overset{\text{def}}{=} \int_{u=-\infty}^{t} \lambda(u)\,du$$

Example 0.31 (Cumulative hazard function of an exponential distribution)**.** Continuing Example 0.30, the hazard is $\lambda(u) = \lambda$ for $u \geq 0$ and 0 for $u < 0$, so for $t \geq 0$:

$$\begin{aligned}
\Lambda(t) &= \int_{u=-\infty}^{0} 0\,du + \int_{u=0}^{t} \lambda\,du && \text{(split the integral at 0)} \\
&= 0 + \lambda t && \text{(integrate each piece)} \\
&= \lambda t && \text{(simplify)}
\end{aligned}$$

and $\Lambda(t) = 0$ for $t < 0$.

Corollary 0.4 (Cumulative hazard of a non-negative random variable)**.** *If T is a continuous random variable with $T \geq 0$, such as a time to event, then $\lambda(u) = 0$ for every $u < 0$, and for every $t \geq 0$:*

$$\Lambda(t) = \int_{u=0}^{t} \lambda(u)\,du$$

Proof

Proof. Let $u < 0$. Every outcome has $T(\omega) \geq 0 > u$, so $\{T \geq u\} = \Omega$, which has probability $1 > 0$, and $\lambda(u)$ is defined. For $0 < \Delta \leq -u$, the event $\{u \leq T < u + \Delta\}$ is empty, because $u + \Delta \leq 0$, so:

$$\begin{aligned}
\lambda(u) &= \lim_{\Delta \downarrow 0} \frac{\Pr(u \leq T < u + \Delta \mid T \geq u)}{\Delta} && \text{(definition of the hazard function)} \\
&= \lim_{\Delta \downarrow 0} \frac{\Pr(\{u \leq T < u + \Delta\} \cap \{T \geq u\})}{\Delta \cdot \Pr(T \geq u)} && \text{(definition of conditional probability)} \\
&= \lim_{\Delta \downarrow 0} \frac{\Pr(\emptyset \cap \Omega)}{\Delta \cdot \Pr(\Omega)} && \text{(the two events found above)} \\
&= \lim_{\Delta \downarrow 0} \frac{0}{\Delta \cdot 1} && (\emptyset \cap \Omega = \emptyset;\ \Pr(\emptyset) = 0;\ \Pr(\Omega) = 1) \\
&= 0 && \text{(simplify)}
\end{aligned}$$

Then, for $t \geq 0$:

$$
\begin{aligned}
\Lambda(t) &= \int_{u=-\infty}^{t} \lambda(u)\, du && \text{(definition of the cumulative hazard)} \\
&= \int_{u=-\infty}^{0} \lambda(u)\, du + \int_{u=0}^{t} \lambda(u)\, du && \text{(split the integral at 0)} \\
&= \int_{u=-\infty}^{0} 0\, du + \int_{u=0}^{t} \lambda(u)\, du && (\lambda(u) = 0 \text{ for } u < 0) \\
&= \int_{u=0}^{t} \lambda(u)\, du && \text{(integrate)}
\end{aligned}
$$

□

Remark. The form in Corollary 0.4, with lower limit 0, is the form most survival-analysis texts use.

Corollary 0.5 (Survival function from the cumulative hazard)**.** *If T is a continuous random variable whose density f is continuous at all but finitely many points, then for every t with $\mathrm{S}(t) > 0$:*

$$\Lambda(t) = -\log\{\mathrm{S}(t)\}$$

and therefore:

$$\mathrm{S}(t) = \exp\{-\Lambda(t)\} \tag{1}$$

Proof

Proof. Since $\mathrm{S}(t) = 1 - F(t)$ (Theorem 0.14) and F has derivative f wherever f is continuous (Theorem 0.6), S has derivative $-f$ at those points. At each such u with $\mathrm{S}(u) > 0$:

$$
\begin{aligned}
\frac{d}{du}(-\log\{\mathrm{S}(u)\}) &= -\frac{1}{\mathrm{S}(u)} \cdot \frac{d}{du}\mathrm{S}(u) && \text{(chain rule)} \\
&= -\frac{1}{\mathrm{S}(u)} \cdot (-f(u)) && (\mathrm{S}' = -f) \\
&= \frac{f(u)}{\mathrm{S}(u)} && \text{(simplify)} \\
&= \lambda(u) && \text{(hazard equals density over survival)}
\end{aligned}
$$

The function $-\log\{\mathrm{S}(u)\}$ is continuous, because S is continuous for a continuous T, and it has derivative $\lambda(u)$ at all but finitely many points, so it is an antiderivative of λ for the fundamental theorem of calculus. As $u \to -\infty$, $\mathrm{S}(u) \to 1$, so $-\log\{\mathrm{S}(u)\} \to 0$. Therefore:

$$\begin{aligned}
\Lambda(t) &\stackrel{\text{def}}{=} \int_{u=-\infty}^{t} \lambda(u)\, du && \text{(definition of the cumulative hazard)} \\
&= [-\log\{\mathrm{S}(u)\}]_{u=-\infty}^{t} && (-\log \mathrm{S} \text{ is an antiderivative of } \lambda) \\
&= -\log\{\mathrm{S}(t)\} - 0 && \text{(evaluate at the bounds)} \\
&= -\log\{\mathrm{S}(t)\} && \text{(simplify)}
\end{aligned}$$

Exponentiating both sides of $-\Lambda(t) = \log\{\mathrm{S}(t)\}$ gives Equation 1. □

Example 0.32 (Recovering the exponential survival function from its cumulative hazard). Continuing Example 0.31, for $t \geq 0$:

$$\begin{aligned}
\mathrm{S}(t) &= \exp\{-\Lambda(t)\} && \text{(survival function from the cumulative hazard)} \\
&= \exp\{-\lambda t\} && (\text{substitute } \Lambda(t) = \lambda t)
\end{aligned}$$

which matches the survival function computed directly in Example 0.27.

Table 1: Probability distribution functions of a continuous random variable T

Name	Symbols	Definition
Probability density function (PDF)	$f(t), \mathrm{p}(t)$	$f \geq 0$ with $\int_a^b f(u)\, du = \Pr(a \leq T \leq b)$
Cumulative distribution function (CDF)	$F(t)$	$\Pr(T \leq t)$
Survival function	$\mathrm{S}(t), \bar{F}(t)$	$\Pr(T > t)$
Hazard function	$\lambda(t), \mathrm{h}(t)$	$\lim_{\Delta \downarrow 0} \Pr(t \leq T < t + \Delta \mid T \geq t)/\Delta$
Cumulative hazard function	$\Lambda(t), \mathrm{H}(t)$	$\int_{u=-\infty}^{t} \lambda(u)\, du$
Log-hazard function	$\eta(t)$	$\log\{\lambda(t)\}$

Remark. For a continuous random variable $T \geq 0$ whose density is continuous at all but finitely many points, the following results connect the functions in Table 1:

- Theorem 0.14
- Theorem 0.15
- Corollary 0.4
- Corollary 0.5

Each arrow in the following diagram converts one function into the next:

$$f(t) \xleftarrow[\mathrm{S}(t)\lambda(t)]{-\mathrm{S}'(t)} \mathrm{S}(t) \xleftarrow{\exp\{-\Lambda(t)\}} \Lambda(t) \xleftarrow{\int_{u=0}^{t} \lambda(u)\,du} \lambda(t) \xleftarrow{\exp\{\eta(t)\}} \eta(t)$$

$$f(t) \xrightarrow[\int_{u=t}^{\infty} f(u)\,du]{f(t)/\lambda(t)} \mathrm{S}(t) \xrightarrow[-\log\{\mathrm{S}(t)\}]{} \Lambda(t) \xrightarrow[\Lambda'(t)]{} \lambda(t) \xrightarrow[\log\{\lambda(t)\}]{} \eta(t)$$

References

Billingsley, Patrick. 1995. *Probability and Measure*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. Wiley.

Casella, George, and Roger Berger. 2002. *Statistical Inference*. 2nd ed. Cengage Learning. https://www.cengage.com/c/statistical-inference-2e-casella-berger/9780534243128/.

Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models*. 4th ed. CRC press. https://doi.org/10.1201/9781315182780.

Hutchinson, Brian. n.d. *DATA 471/571 (Machine Learning) and CSCI 481/581 (Deep Learning) Video Lectures*. Western Washington University. Accessed September 28, 2026. https://facultyweb.cs.wwu.edu/~hutchib2/video_lectures/data371/.

Kalbfleisch, John D, and Ross L Prentice. 2011. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Klein, John P, and Melvin L Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. Springer. https://link.springer.com/book/10.1007/b97377.

Kleinbaum, David G, and Mitchel Klein. 2012. *Survival Analysis: A Self-Learning Text*. 3rd ed. Springer. https://link.springer.com/book/10.1007/978-1-4419-6646-9.

Rothman, Kenneth J., Timothy L. Lash, Tyler J. VanderWeele, and Sebastien Haneuse. 2021. *Modern Epidemiology*. Fourth edition. Wolters Kluwer.

Vittinghoff, Eric, David V Glidden, Stephen C Shiboski, and Charles E McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. Springer. https://doi.org/10.1007/978-1-4614-1353-0.