## 5 Continuous Random Variables

### 5.1 Definition and Basic Examples

Recall that discrete random variable may be defined in terms of a probability mass function $p(x)$; a non-negative function for which $\sum p(x)=1$. To describe a continuous variable, we need only replace the sum by its continuous analogue, the integral.

Definition 5.1. Suppose $S$ is a sample space and $X: S \rightarrow \mathbb{R}$ a random variable. We say that $X$ is continuous if there exists a function $f: \mathbb{R} \rightarrow \mathbb{R}$ with the following properties:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$;
- For all (measureable) subsets $K \subseteq \mathbb{R}$, we have $\mathbb{P}\{X \in K\}=\int_{K} f(x) \mathrm{d} x$

We call $f$ the probability density function (p.d.f.) of $X$.

Example 5.2. Find $k$ such that the following is the p.d.f. of a continuous r.v. $X$ and find its cumulative distribution function $F(x)=\mathbb{P}\{X \leq x\}$.

$$
f(x)= \begin{cases}k x^{3} & \text { if } 0<x<2 \\ 0 & \text { otherwise }\end{cases}
$$

We require

$$
1=\int_{\mathbb{R}} f(x) \mathrm{d} x=\int_{0}^{2} k x^{3} \mathrm{~d} x=\frac{1}{4} k \cdot 2^{4} \Longrightarrow k=\frac{1}{4}
$$

from which

$$
\begin{aligned}
F(x) & =\mathbb{P}\{X \leq x\}=\int_{-\infty}^{2} f(x) \mathrm{d} x \\
& = \begin{cases}0 & \text { if } x \leq 0 \\
\int_{0}^{x} \frac{1}{4} x^{3} \mathrm{~d} x=\frac{1}{16} x^{4} & \text { if } 0<x<2 \\
1 & \text { if } x \geq 2\end{cases}
\end{aligned}
$$




Basic Observations and Conventions Let $X$ have density function $f(x)$.

- For $K$ to be measurable loosely means that $\int_{K} \mathrm{~d} x$ exists. We're mostly concerned with intervals, which are certainly measurable!
- More generally, it is permissible for $X: S \rightarrow \mathbb{R}^{n}$ to be a continuous random vector, in which case we'd use a multiple integral.
- As in the example, we often have $f(x)=0$ outside some bounded set. In such cases, it is common to define $f(x)$ only on that set.
- The singleton event $\{a\}$ almost never occurs;

$$
\mathbb{P}\{X=a\}=\int_{a}^{a} f(x) \mathrm{d} x=0
$$

- The choice of whether to use strict or non-strict inequalities is irrelevant:

$$
\mathbb{P}\{X<a\}=\mathbb{P}\{X \leq a\}=\int_{-\infty}^{a} f(x) \mathrm{d} x
$$

- The cumulative distribution function $F(x)$ is useful, though it might be difficult or impossible to compute explicitly. Appealing to the fundamental theorem of calculus, $F(x)$ is continuous on $\mathbb{R}$ and, at any $x$ where $f(x)$ is continuous, $F(x)$ is differentiable with derivative $F^{\prime}(x)=f(x)$.
- If we change the value of $f(x)$ at isolated points $x$, then its integral is unchanged. For instance, the function

$$
g(x)= \begin{cases}\frac{1}{4} x^{3} & \text { if } 0<x<1 \text { or } 1<x \leq 2 \\ 0 & \text { otherwise }\end{cases}
$$

differs from the density $f(x)$ of Example 5.2 at two values $(x=1,2)$, yet it describes the same variable $X$, since $\int_{a}^{b} g(x) \mathrm{d} x=\int_{a}^{b} f(x) \mathrm{d} x$ for all $a, b$. Since a distribution has many density functions, the definition should really refer to $a$ density, not the density. It is conventional to assume that $f(x)$ is continuous wherever possible but there will still likely be endpoints of intervals where a choice must be made. The point is that this choice is irrelevant! By contrast, the cumulative distribution function $F(x)$ is uniquely defined.

Example 5.3. Suppose $X$ has density $f_{X}(x)$. We find the cumulative distribution function and density for the random variable $Y=\frac{1}{2} X+1$.

$$
F_{Y}(y)=\mathbb{P}\left\{\frac{1}{2} X+1 \leq y\right\}=\mathbb{P}\{X \leq 2(y-1)\}=F_{X}(2 y-2)
$$

Whenever $f_{X}(x)$ is continuous, we can differentiate to find the density

$$
f_{Y}(y)=F_{Y}^{\prime}(y)=2 F_{X}^{\prime}(2 y-2)=2 f_{X}(2 y-2)
$$

If $X$ is the random variable described in Example 5.2 with $F_{X}(x)=\frac{1}{16} x^{4}$ and $f_{X}(x)=\frac{1}{4} x^{3}$ on [0,2], then

$$
\begin{aligned}
& F_{Y}(y)= \begin{cases}0 & \text { if } y \leq 1 \\
(y-1)^{4} & \text { if } 1<y<2 \\
1 & \text { if } y \geq 2\end{cases} \\
& f_{Y}(y)= \begin{cases}4(y-1)^{3} & \text { if } 1<y<2 \\
0 & \text { otherwise }\end{cases}
\end{aligned}
$$




The above example shows that you can sometimes get away with a simple substitution when changing random variables, but only if you work with the cumulative function $F(x)$ ! Since the density requires a derivative, direct substitution is almost guaranteed to fail. Remember also that if $F(x)$ is only stated on the interesting part of its domain (i.e. the interval on which $0<F(x)<1$ ), then you'll need to be careful translating this to your new variable.

Example 5.4. For a trickier example, we repeat the last with the variable $Z=X^{2}$.

$$
F_{Z}(z)=\mathbb{P}\left\{X^{2} \leq z\right\}=\mathbb{P}\{-\sqrt{z} \leq X \leq \sqrt{z}\}=F_{X}(\sqrt{z})-F_{X}(-\sqrt{z})
$$

Note how this is is not $F_{X}(\sqrt{z})$ ! Whenever $f_{X}(x)$ is continuous, we can differentiate:

$$
f_{Z}(z)=F_{Z}^{\prime}(z)=\frac{1}{2 \sqrt{z}}\left(f_{X}(\sqrt{z})+f_{X}(-\sqrt{z})\right)
$$

If $X$ is the explicit distribution from Example 5.2 then the substitutions are easy since $X$ is always positive. Indeed, provided $0<z=x^{2}<4$, we have

$$
\begin{aligned}
& F_{Z}(z)=F_{X}(\sqrt{z})-F_{X}(-\sqrt{z})=F_{X}(\sqrt{z})=\frac{1}{16} z^{2} \\
& f_{Z}(z)=\frac{1}{2 \sqrt{z}} \cdot \frac{1}{4}(\sqrt{z})^{3}=\frac{1}{8} z=F_{Z}^{\prime}(z)
\end{aligned}
$$




We finish by considering the simplest family of continuous distributions, which model a random number generator where any value in a finite interval is equally likely.

Definition 5.5. The random variable $X \sim U(a, b)$ has a uniform distribution on the interval $[a, b]$ if it has density and cumulative distribution functions

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



Of course each individual outcome has probability zero!

### 5.2 Expectation and Variance

Definition 5.6. The expectation of a continuous random variable $X$ with density $f(x)$ is

$$
\mathbb{E}[X]:=\int_{-\infty}^{\infty} x f(x) \mathrm{d} x
$$

The intuition for this is twofold.

1. Recall how the density plays a similar role for continuous variables as the mass did for a discrete variable. We therefore replace the sum in the expression $\mathbb{E}[X]=\sum_{x} x p(x)$ with an integral.
2. Let $\Delta x$ be very small, consider the sequence $x_{n}=n \Delta x$ where $n$ is any integer, and define a discrete random variable $Y$ with mass function

$$
p\left(x_{n}\right)=\mathbb{P}\left\{x_{n}<X \leq x_{n+1}\right\}=F\left(x_{n+1}\right)-F\left(x_{n}\right) \approx F^{\prime}\left(x_{n}\right) \Delta x=f\left(x_{n}\right) \Delta x
$$

where we used the mean value theorem from calculus. But then

$$
\mathbb{E}[Y]=\sum_{n} x_{n} p\left(x_{n}\right) \approx \sum_{n} x_{n} f\left(x_{n}\right) \Delta x \approx \int_{\mathbb{R}} x f(x) \mathrm{d} x
$$

where the approximations should become equalities as $n \rightarrow \infty$.
Examples 5.7. 1. If $f(x)=\frac{1}{4} x^{3}$ on $[0,2]$, then the expectation of $X$ is

$$
\mathbb{E}[X]=\int_{0}^{2} \frac{1}{4} x^{4} \mathrm{~d} x=\frac{1}{20} 2^{5}=\frac{8}{5}
$$

2. The expectation of the uniform distribution on $[a, b]$ is the midpoint of the interval

$$
\mathbb{E}[X]=\int_{a}^{b} \frac{x}{b-a} \mathrm{~d} x=\left.\frac{1}{2(b-a)} x^{2}\right|_{a} ^{b}=\frac{a+b}{2}
$$

3. The function $f(x)=\frac{2}{x^{3}}$ when $x \geq 1$ and $f(x)=0$ otherwise defines a continuous random variable $X$, since

$$
\int_{\mathbb{R}} f(x) \mathrm{d} x=\int_{1}^{\infty} \frac{2}{x^{3}} \mathrm{~d} x=-\left.\frac{1}{x^{2}}\right|_{1} ^{\rightarrow \infty}=1
$$

with cumulative distribution function

$$
F(x)= \begin{cases}0 & \text { if } x \leq 1 \\ 1-\frac{1}{x^{2}} & \text { if } x>1\end{cases}
$$

This has expectation

$$
\mathbb{E}[X]=\int_{1}^{\infty} \frac{2}{x^{2}} \mathrm{~d} x=-\left.\frac{2}{x}\right|_{1} ^{\rightarrow \infty}=2
$$



4. Suppose $X$ has density $f_{X}(x)=\frac{3}{2} \sqrt{x}$ on $[0,1]$, and thus $F_{X}(x)=x^{3 / 2}$. To compute the expectation of the random variable $Y=X^{3}$, we first need its cumulative function

$$
F_{Y}(y)=\mathbb{P}\left\{X^{3} \leq y\right\}=\mathbb{P}\{X \leq \sqrt[3]{y}\}=F_{X}(\sqrt[3]{y})=\sqrt{y}
$$

valid whenever $0<y<1$. We can differentiate to find the density of $Y$ and thus calculate the expected value

$$
\begin{aligned}
\mathbb{E}\left[X^{3}\right] & =\mathbb{E}[Y]=\int_{-\infty}^{\infty} y F_{Y}^{\prime}(y) \mathrm{d} y=\int_{0}^{1} y \cdot \frac{1}{2} y^{-1 / 2} \mathrm{~d} y \\
& =\left.\frac{1}{3} y^{3 / 2}\right|_{0} ^{1}=\frac{1}{3}
\end{aligned}
$$

Compare this to the expectation of $X$ itself:

$$
\mathbb{E}[X]=\int_{0}^{1} \frac{3}{2} x^{3 / 2} \mathrm{~d} x=\frac{3}{5}
$$

This shouldn't be surprising: since the only interesting values

 of $X$ lie in the interval $(0,1)$ we have $Y<X$ !
Just as for discrete variables, this can be computed more easily without first finding $F_{Y}(y)$ :

$$
\mathbb{E}\left[X^{3}\right]=\int_{0}^{1} x^{3} f_{X}(x) \mathrm{d} x=\int_{0}^{1} \frac{3}{2} x^{7 / 2} \mathrm{~d} x=\frac{1}{3}
$$

The last observation comes from the general result:
Theorem 5.8. If $X$ has density $f(x)$, and $g$ is a function, then

$$
\mathbb{E}[g(X)]=\int_{\mathbb{R}} g(x) f(x) \mathrm{d} x
$$

The argument is a little tricky since we need to break it into two pieces and we need to change the order of a double-integral twice.

Proof. Let $Y=g(X)$. We break the expectation into two integrals:

$$
\mathbb{E}[Y]=\int_{0}^{\infty} t f_{Y}(t) \mathrm{d} t+\int_{-\infty}^{0} t f_{Y}(t) \mathrm{d} t
$$

The first may be evaluated by twice changing the order of integration,

$$
\begin{aligned}
\int_{0}^{\infty} t f_{Y}(t) \mathrm{d} t & =\int_{0}^{\infty} \int_{0}^{t} f_{Y}(t) \mathrm{d} y \mathrm{~d} t=\int_{0}^{\infty} \int_{y}^{\infty} f_{Y}(t) \mathrm{d} t \mathrm{~d} y \\
& =\int_{0}^{\infty} \mathbb{P}\{g(X)>y\} \mathrm{d} y=\int_{0}^{\infty} \int_{x: g(x)>y} f(x) \mathrm{d} x \mathrm{~d} y \\
& =\int_{x: g(x)>0} \int_{0}^{g(x)} f(x) \mathrm{d} y \mathrm{~d} x=\int_{x: g(x)>0} g(x) f(x) \mathrm{d} x
\end{aligned}
$$

A similar calculation evaluates the second integral; to paraphrase slightly ${ }^{1}$

$$
\begin{aligned}
\int_{-\infty}^{0} t f_{Y}(t) \mathrm{d} t & =-\int_{0}^{\infty} \int_{t}^{0} f_{Y}(t) \mathrm{d} y \mathrm{~d} t=-\int_{-\infty}^{0} \mathbb{P}\{g(X)<y\} \mathrm{d} y \\
& =-\int_{x: g(x)<0} \int_{g(x)}^{0} f(x) \mathrm{d} y \mathrm{~d} x=\int_{x: g(x)<0} g(x) f(x) \mathrm{d} x
\end{aligned}
$$

Summing these gives the result.

Corollary 5.9. 1. Expectations are linear $\mathbb{E}[a X+b]=a \mathbb{E}[X]+b$
2. If $g(X)$ is always non-negative, then

$$
\mathbb{E}[g(X)]=\int_{0}^{\infty} \mathbb{P}\{g(X)>y\} \mathrm{d} y
$$

Definition 5.10. The variance of a continuous distribution $X$ with expectation $\mathbb{E}[X]=\mu$, is

$$
\begin{aligned}
\operatorname{Var} X & =\mathbb{E}\left[(X-\mu)^{2}\right]=\int_{\mathbb{R}}(x-\mu)^{2} f(x) \mathrm{d} x=\int_{\mathbb{R}} x^{2} f(x) \mathrm{d} x-2 \mu \int_{\mathbb{R}} x f(x) \mathrm{d} x+\mu^{2} \int_{\mathbb{R}} f(x) \mathrm{d} x \\
& =\mathbb{E}\left[X^{2}\right]-\mu^{2}
\end{aligned}
$$

Examples 5.11. 1. The uniform distribution on $[a, b]$ has variance

$$
\begin{aligned}
\operatorname{Var} X & =\int_{a}^{b} \frac{1}{b-a}\left(x-\frac{a+b}{2}\right)^{2} \mathrm{~d} x=\left.\frac{1}{3(b-a)}\left(x-\frac{a+b}{2}\right)^{3}\right|_{a} ^{b}=\frac{2}{3(b-a)}\left(\frac{b-a}{2}\right)^{3} \\
& =\frac{1}{12}(b-a)^{2}
\end{aligned}
$$

and standard deviation $\sigma=\frac{1}{2 \sqrt{3}}(b-a)$. Otherwise said $\frac{1}{\sqrt{3}} \approx 57.7 \%$ of outcomes lie within one standard deviation of the mean.
2. As with discrete variables, the linearity of the expectation proves that

$$
\operatorname{Var}(a X+b)=a^{2} \operatorname{Var} X
$$

Observe that if $Y \sim U(a, b)$, then $Y=(b-a) X+a$ where $X \sim U(0,1)$ follows a standardized normal distribution. Indeed

$$
\mathbb{E}[Y]=\frac{a+b}{2}=(b-a) \frac{1}{2}+a=\mathbb{E}[(b-a) X+a]
$$

and

$$
\operatorname{Var} Y=\frac{1}{12}(b-a)^{2}=(b-a)^{2} \operatorname{Var} X
$$

[^0]3. Players throw darts at a circular target of radius 10 cm . Suppose that the probability of landing a distance $r \geq 0$ from the center of the target has density
$$
f(r)=\frac{1}{(r+1)^{2}}
$$
(a) The probability of hitting the target at all is
$$
\int_{0}^{10} \frac{1}{(r+1)^{2}} \mathrm{~d} r=-\left.\frac{1}{r+1}\right|_{0} ^{10}=1-\frac{1}{11}=\frac{10}{11} \approx 90.9 \%
$$
(b) If they hit the target, the player receives a score $X(r)=10-r$. The expected score from one dart is then
\[

$$
\begin{aligned}
\mu & =\mathbb{E}[X]=\int_{0}^{10} \frac{10-r}{(r+1)^{2}} \mathrm{~d} r=\int_{0}^{10} \frac{-(r+1)+11}{(r+1)^{2}} \mathrm{~d} r=-\ln (r+1)-\left.\frac{11}{r+1}\right|_{0} ^{10} \\
& =-\ln 11-1+11=10-\ln 11 \approx 7.6
\end{aligned}
$$
\]

Moreover

$$
\begin{aligned}
\mathbb{E}\left[X^{2}\right] & =\int_{0}^{10} \frac{(10-r)^{2}}{(r+1)^{2}} \mathrm{~d} r=\int_{0}^{10} 1-\frac{22}{r+1}+\frac{121}{(r+1)^{2}} \mathrm{~d} r \\
& =10-22 \ln 11+121 \cdot \frac{10}{11}=120-22 \ln 11 \\
\Longrightarrow \operatorname{Var} X & =\mathbb{E}\left[X^{2}\right]-\mu^{2}=120-22 \ln 11-(10-\ln 11)^{2} \\
& =20-2 \ln 11-(\ln 11)^{2} \approx 9.45
\end{aligned}
$$

Note that the standard deviation is $\sigma \approx 3.07$.
(c) Alternatively, suppose a player scores $Y=10$ points if they land within 5 cm of the center and $Y=5$ points if they land between 5 and 10 cm from the center. This time

$$
\mathbb{E}[Y]=\int_{0}^{5} \frac{10}{(r+1)^{2}} \mathrm{~d} r+\int_{5}^{10} \frac{5}{(r+1)^{2}} \mathrm{~d} r=\frac{575}{66} \approx 8.71
$$

Note in this case that $Y$ is a discrete random variable!
4. A stick of length 1 is broken at a random point a distance $X$ along the stick where $X \sim U(0,1)$. If $p \in(0,1)$ represents a point on the original stick, find the expectation and standard deviation of the length of the piece containing $p$.
The length of the stick containing $p$ is a function of $X$ (and $p$ ):

$$
Y=g(X)= \begin{cases}X & \text { if } X>p \\ 1-X & \text { if } X<p\end{cases}
$$

Note that $X=p$ is irrelevant since the probability of breaking the stick precisely at $p$ is zero! Since the density function for $X$ is $f(x)=1$, we therefore have

$$
\mathbb{E}[Y]=\int_{0^{p}}(1-x) \mathrm{d} x+\int_{p}^{1} x \mathrm{~d} x=\frac{1}{2}-\frac{1}{2} p^{2}+p-\frac{1}{2} p^{2}=\frac{1}{2}+p-p^{2}=\frac{1}{2}+p q
$$

and

$$
\begin{aligned}
\operatorname{Var} Y & =\mathbb{E}\left[Y^{2}\right]-\mu_{Y}^{2}=\int_{0}^{p}(1-x)^{2} \mathrm{~d} x+\int_{p}^{1} x^{2} \mathrm{~d} x-\left(\frac{1}{2}+p q\right)^{2} \\
& =\frac{1}{3}\left(1-(1-p)^{3}\right)+\frac{1}{3}\left(1-p^{3}\right)-\left(\frac{1}{2}+p q\right)^{2} \\
& =\frac{1}{3}\left(1-1+3 p-3 p^{2}+p^{3}+1-p^{3}\right)-\frac{1}{4}-p q-p^{2} q^{2} \\
& =p q+\frac{1}{3}-\frac{1}{4}-p q-p^{2} q^{2}=\frac{1}{12}-p^{2} q^{2}
\end{aligned}
$$

### 5.3 The Normal Distribution

Even non-mathematicians have heard of the normal distribution, even if not by that name; more familiar is the characteristic bell-shaped curve exhibited by its density function. The function $a e^{-k t^{2}}$ produces the classic shape, though not all such functions can be densities. If we use polar co-ordinates,

$$
\begin{aligned}
\int_{0}^{\infty} a e^{-k x^{2}} \mathrm{~d} x \int_{0}^{\infty} a e^{-k y^{2}} \mathrm{~d} y & =a^{2} \int_{0}^{\infty} \int_{0}^{\infty} e^{-k x^{2}-k y^{2}} \mathrm{~d} x \mathrm{~d} y=a^{2} \int_{0}^{\frac{\pi}{2}} \int_{0}^{\infty} e^{-k r^{2}} r \mathrm{~d} r \mathrm{~d} \theta \\
& =\left.a^{2} \frac{\pi}{2}\left(-\frac{1}{2 k} e^{-k r^{2}}\right)\right|_{0} ^{\infty}=\frac{a^{2} \pi}{4 k}
\end{aligned}
$$

from which

$$
\int_{-\infty}^{\infty} a e^{-k x^{2}} \mathrm{~d} x=2 \sqrt{\frac{a^{2} \pi}{4 k}}=a \sqrt{\frac{\pi}{k}}=1 \Longleftrightarrow a=\sqrt{\frac{k}{\pi}}
$$

The standard approach, for reasons to be seen shortly, is to relabel $k=\frac{1}{2 \sigma^{2}}$, whence $a=\frac{1}{\sqrt{2 \pi} \sigma}$.
Definition 5.12. A random variable $X \sim N\left(\mu, \sigma^{2}\right)$ has normal or Gaussian distribution with parameters $\mu$ and $\sigma^{2}$ if its density is

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

where $\exp t=e^{t}$ is the exponential function.
The density and the cumulative distribution function

$$
F(x)=\int_{-\infty}^{x} \frac{1}{\sqrt{2 \pi} \sigma} \exp \left(-\frac{(t-\mu)^{2}}{2 \sigma^{2}}\right) d t
$$

are drawn.
The cumulative distribution function of the standard normal distribution $N(0,1)$ is typically denoted


$$
\Phi(x)=\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{x} e^{-\frac{t^{2}}{2}} d t
$$

Before discussing some nice facts about the normal distribution, we confront its major irritant; the cumulative distribution function $F(x)$ cannot be explicitly evaluated since the density does not have a closed-form anti-derivative. We must therefore rely on a computer or on a table of values.

Lemma 5.13. If $X \sim N\left(\mu, \sigma^{2}\right)$, then $Z=\frac{X-\mu}{\sigma} \sim N(0,1)$ is the standard normal distribution. In particular,

$$
\mathbb{P}\{a \leq X \leq b\}=\mathbb{P}\left\{\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)
$$

By virtue of the Lemma, we need only have a table of values for the standard normal distribution in order to compute with any such.

Proof. Simply compute the distribution function for Z and differentiate:

$$
\begin{aligned}
& F_{Z}(z)=\mathbb{P}\left\{\frac{X-\mu}{\sigma} \leq z\right\}=\mathbb{P}\{X \leq \mu+\sigma z\}=F_{X}(\mu+\sigma z) \\
& \Longrightarrow f_{Z}(z)=\sigma f_{X}(\mu+\sigma z)=\frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{z^{2}}{2}\right)
\end{aligned}
$$

Example 5.14. Suppose $X \sim N(4,9)$ has parameters $\mu=4$ and $\sigma^{2}=9$. Then:

1. $\mathbb{P}\{X \leq 2\}=\Phi\left(\frac{2-4}{3}\right)=\Phi\left(-\frac{2}{3}\right)=1-\Phi\left(\frac{2}{3}\right) \approx 0.2525$. As we did here, it is common to convert $\Phi(-x)=1-\Phi(x)$ since normal distribution tables are often given only when $x \geq 0$.
2. $\mathbb{P}\{3 \leq X \leq 5\}=\Phi\left(\frac{5-4}{3}\right)-\Phi\left(\frac{3-4}{3}\right)=\Phi\left(\frac{1}{3}\right)-\Phi\left(-\frac{1}{3}\right)=2 \Phi\left(\frac{1}{3}\right)-1 \approx 0.2611$
3. $\mathbb{P}\{|X-4| \geq 6\}=1-\mathbb{P}\{-2 \leq X \leq 10\}=1-\Phi\left(\frac{10-4}{3}\right)+\Phi\left(\frac{-2-4}{3}\right)=1-\Phi(2)+\Phi(-2)=$ $2(1-\Phi(2)) \approx 0.0455$

Since we use the symbols $\mu$ and $\sigma^{2}$ for the parameters of a normal distribution, the expectation and variance are precisely what you think they are!

Lemma 5.15. If $X \sim N\left(\mu, \sigma^{2}\right)$, then $\mathbb{E}[X]=\mu$ and $\operatorname{Var} X=\sigma^{2}$.

Proof. Both required integrals may be evaluated by making the substitution $z=\frac{x-\mu}{\sigma} \Longrightarrow \mathrm{d} z=\frac{1}{\sigma} \mathrm{~d} x$.

$$
\mathbb{E}[X]=\int_{-\infty}^{\infty} \frac{x}{\sqrt{2 \pi} \sigma} \exp \left(-\frac{(x-\mu)^{2}}{2 \sigma^{2}}\right) \mathrm{d} x=\int_{-\infty}^{\infty} \frac{\mu+\sigma z}{\sqrt{2 \pi}} \exp \left(-\frac{z^{2}}{2}\right) \mathrm{d} z=\mu
$$

since $z e^{-z^{2} / 2}$ is an odd function. To find the variance,

$$
\begin{aligned}
\operatorname{Var} X & =\int_{-\infty}^{\infty} \frac{(x-\mu)^{2}}{\sqrt{2 \pi} \sigma} \exp \left(-\frac{(x-\mu)^{2}}{2 \sigma^{2}}\right) \mathrm{d} x=\int_{-\infty}^{\infty} \frac{\sigma^{2} z^{2}}{\sqrt{2 \pi}} \exp \left(-\frac{z^{2}}{2}\right) \mathrm{d} z \\
& =\sigma^{2}\left(-\left.\frac{z}{\sqrt{2 \pi}} \exp \left(-\frac{z^{2}}{2}\right)\right|_{-\infty} ^{\infty}+\int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{z^{2}}{2}\right) \mathrm{d} z\right)=\sigma^{2}
\end{aligned}
$$

where we used integration by parts ${ }^{a}$

$$
{ }^{a} \text { Specifically } \int z \cdot z e^{-z^{2} / 2} \mathrm{~d} z \text { where } u=z \text { and } \mathrm{d} v=z e^{-z^{2} / 2} \mathrm{~d} z
$$

Example 5.16. A runner takes exactly 11 seconds to complete a 100 m race. Their personal best is 10.7 seconds. If the time recorded on a (human-operated!) stopwatch is a normally distributed random variable $T \sim N(11,0.04)$, find the probability that the runner records a new personal best.
Convert to a standard normal distribution:

$$
\mathbb{P}\{T \leq 10.7\}=\Phi\left(\frac{10.7-11}{\sqrt{0.04}}\right)=\Phi(-1.5)=1-\Phi(1.5) \approx 0.0668
$$

The example should give you some pause. According to the model, there is a positive probability that the error is so large as to record a negative time for the race! Indeed

$$
\mathbb{P}\{T \leq 0\}=\Phi\left(\frac{0-11}{\sqrt{0.04}}\right)=\Phi(-55)
$$

This number is so absurdly small that it isn't worth thinking about. Indeed even the chance of an error of at least 1 second either way is minuscule

$$
\mathbb{P}\{|T-11|>1\}=2 \Phi(-5) \approx 0.0000005
$$

This does however take us to why the normal distribution is so important: many natural phenomena turn out to be approximately normally distributed. For example, the height of a random member of a large population, the time a person requires to complete a task, experimental measurement error, etc. Indeed the term Gaussian is in honor of Gauss' analysis of measurement error. The formal reason for the normal distribution seeming so common is the Central Limit Theorem, one of the most important results in probability. Loosely stated, it says that the sum

$$
X_{1}+\cdots+X_{n}
$$

of a large number of independent random variables is approximately normally distributed, regardless of the distributions $X_{i}$ ! While its discussion is a core topic for a future course, it is worth stating the simplest and oldest version explicitly.

Theorem 5.17 (DeMoivre-Laplace). Suppose that $X \sim B(n, p)$. That is $X=X_{1}+\cdots+X_{n}$ where the $X_{i}$ are independent Bernoulli trials with probability of success $p$. Then $X$ is approximately normally distributed $N(n p, n p q)$. That is,

$$
\mathbb{P}\{a \leq X \leq b\} \approx \Phi\left(\frac{b-n p}{\sqrt{n p q}}\right)-\Phi\left(\frac{a-n p}{\sqrt{n p q}}\right)
$$

Example 5.18. A fair die is rolled 10000 times and the number of sixes are counted; thus $X \sim$ $B\left(10000, \frac{1}{6}\right)$ is approximately normally distributed $N\left(\frac{10000}{6}, \frac{50000}{36}\right)=N\left(\frac{5000}{3}, \frac{12500}{9}\right) \approx N(1667,1389)$. The probability of obtaining at least 1600 sixes is approximately

$$
\mathbb{P}\{X \geq 1600\} \approx 1-\Phi\left(\frac{1600-\frac{5000}{3}}{\frac{50 \sqrt{5}}{3}}\right)=1-\Phi\left(\frac{-200}{50 \sqrt{5}}\right)=\Phi\left(\frac{4}{\sqrt{5}}\right) \approx 96.32 \%
$$

A more accurate estimate using the original binomial distribution yields $96.48 \%$, though this took a computer several seconds!

The Lognormal Distribution A related distribution commonly encountered in financial mathematics and engineering is $Y=e^{X}$, where $X \sim N\left(\mu, \sigma^{2}\right)$. This is often used to model the return on a asset $\left.\right|^{2}$ The expected return is therefore

$$
\begin{aligned}
\mathbb{E}[Y] & =\int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \pi} \sigma} e^{x} \exp \left(-\frac{(x-\mu)^{2}}{2 \sigma^{2}}\right) \mathrm{d} x=\int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \pi} \sigma} \exp \left(-\frac{x^{2}-2 \mu x+\mu^{2}-2 \sigma^{2} x}{2 \sigma^{2}}\right) \mathrm{d} x \\
& =\int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \pi} \sigma} \exp \left(-\frac{\left(x-\mu-\sigma^{2}\right)^{2}-2 \mu \sigma^{2}-\sigma^{4}}{2 \sigma^{2}}\right) \mathrm{d} x \\
& =e^{\mu+\frac{1}{2} \sigma^{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \pi} \sigma} \exp \left(-\frac{\left(x-\mu-\sigma^{2}\right)^{2}}{2 \sigma^{2}}\right) \mathrm{d} x=e^{\mu+\frac{1}{2} \sigma^{2}}
\end{aligned}
$$

According to this model, the expected return on a riskier, high volatility, investment is larger than on a safe investment. Of course, a high-risk investment means you're also more likely to lose money!

Example 5.19. The price of corn is currently $\$ 4$ per bushel. Its price in 1 year is modelled by a lognormal distribution

$$
Y=4 e^{X} \quad \text { where } \quad X=N(-0.05,0.1) \Longrightarrow \mathbb{E}[Y]=4 e^{0}=\$ 4
$$

A tortilla manufacturer wants to guarantee that they can purchase 1000 bushels of corn in 1 year at a price of no more than $\$ 5$ per bushel, and decides to purchase an insurance policy ${ }^{3}$ to cover this risk. What would be a fair premium for this policy?
In 1 year, the payout of the insurance policy is a function of $Y$, and thus $X$ :

$$
P(X)= \begin{cases}1000(Y-5)=1000\left(4 e^{X}-5\right) & \text { if } Y>5(X>\ln 1.25) \\ 0 & \text { if } Y<5\end{cases}
$$

The expected payout, and thus the fair premium, is then

$$
\begin{aligned}
\mathbb{E}[P(X)] & =1000 \int_{\ln 1.25}^{\infty} \frac{4 e^{x}-5}{\sqrt{0.2 \pi}} e^{-\frac{(x+0.05)^{2}}{0.2}} \mathrm{~d} x \\
& =\frac{4000}{\sqrt{0.2 \pi}} \int_{\ln 1.25}^{\infty} e^{-\frac{x^{2}+0.1 x+0.05^{2}-0.2 x}{0.2}} \mathrm{~d} x-5000 \mathbb{P}\{X>\ln 1.25\} \\
& =\frac{4000}{\sqrt{0.2 \pi}} \int_{\ln 1.25}^{\infty} e^{-\frac{(x-0.05)^{2}}{0.2}} \mathrm{~d} x-5000 \mathbb{P}\left\{Z>\frac{\ln 1.25+0.05}{\sqrt{0.1}}\right\} \\
& =1000\left[4 \Phi\left(-\frac{\ln 1.25-0.05}{\sqrt{0.1}}\right)-5 \Phi\left(-\frac{\ln 1.25+0.05}{\sqrt{0.1}}\right)\right]=1000(4 \Phi(-0.5475)-5 \Phi(-0.8638)) \\
& =1000(1.168-0.9695)=\$ 198.50
\end{aligned}
$$

[^1]$$
C=S \Phi\left(\frac{\ln \frac{S}{K}+r+\frac{1}{2} \sigma^{2}}{\sigma}\right)-e^{-r} K \Phi\left(\frac{\ln \frac{S}{K}+r-\frac{1}{2} \sigma^{2}}{\sigma}\right)
$$

### 5.4 The Exponential Distribution

Definition 5.20. A random variable $X \sim \operatorname{Exponential}(\lambda)$ is exponentially distributed with parameter $\lambda>0$ if its density function is

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

The cumulative distribution function is

$$
F(x)=\mathbb{P}\{X \leq x\}=1-e^{-\lambda x} \Longrightarrow \mathbb{P}\{X>x\}=e^{-\lambda x}
$$

The mean and variance are easy to memorise.
Theorem 5.21. If $X \sim \operatorname{Exponential}(\lambda)$, then $\mathbb{E}[X]=\frac{1}{\lambda}$ and $\operatorname{Var} X=\frac{1}{\lambda^{2}}$

Proof. This is easiest if we use a recurrence following integration by parts. If $\mathbb{E}\left[X^{n}\right]=\int_{0}^{\infty} \lambda x^{n} e^{-\lambda x} \mathrm{~d} x$, then for $n \geq 1$ we have

$$
\mathbb{E}\left[X^{n}\right]=-\left.x^{n} e^{-\lambda x}\right|_{0} ^{\rightarrow \infty}+n \int_{0}^{\infty} x^{n-1} e^{-\lambda x} \mathrm{~d} x=\frac{n}{\lambda} \mathbb{E}\left[X^{n-1}\right]
$$

from which

$$
\mathbb{E}\left[X^{n}\right]=\frac{n!}{\lambda^{n}}
$$

Plainly $\mathbb{E}[X]=\frac{1}{\lambda}$, while

$$
\operatorname{Var} X=\frac{2}{\lambda^{2}}-\left(\frac{1}{\lambda}\right)^{2}=\frac{1}{\lambda^{2}}
$$

We previously met the exponential variable when we considered the time to the next event generated by a Poisson process. Its application is typically in such contexts.

Example 5.22. Metro trains arrive at a stop according to a Poisson process with parameter $\lambda$ per minute. Otherwise said, the number of trains arriving in an interval of $t$ minutes is a Poisson random variable

$$
N(t) \sim \operatorname{Poisson}(\lambda t) \Longrightarrow \mathbb{P}\{N(t)=k\}=\frac{(\lambda t)^{k}}{k!} e^{-\lambda t}
$$

If $X$ is the time until the next train arrives, we see that

$$
\mathbb{P}\{X \leq t\}=\mathbb{P}\{N(t) \geq 1\}=1-\mathbb{P}\{N(t)=0\}=1-e^{-\lambda t}
$$

whence $X \sim \operatorname{Exponential}(\lambda)$.
To put some flesh on this, if the parameter is $\lambda=\frac{1}{5}$ of a train per minute, then:
(a) The probability of having to wait at least 10 minutes is $\mathbb{P}\{X>10\}=e^{-2} \approx 13.5 \%$.
(b) The probability of a train arriving within 1 minute is $\mathbb{P}\{X \leq 1\}=1-e^{-0.2} \approx 18.1 \%$.
(c) The expected wait for a train is $\frac{1}{\lambda}=5$ minutes.

Suppose you are waiting for an event to occur and the time required $X$ is exponentially distributed. If you've already been waiting for time $t$, then the probability that you will have to wait at least $s$ longer is

$$
\mathbb{P}\{X>s+t \mid X>t\}=\frac{\mathbb{P}\{X>s+t\}}{\mathbb{P}\{X>t\}}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda t}}=e^{-\lambda s}=\mathbb{P}\{X>s\}
$$

This property, equivalently

$$
\mathbb{P}\{X>s+t\}=\mathbb{P}\{X>s\} \mathbb{P}\{X>t\}
$$

says that the distribution is memoryless; how long you've already been waiting has no bearing on how long you still have to wait. This is related to the assumption made in defining the Poisson process; that events happening in adjacent small intervals are independent.
It is perhaps surprising that the exponential distribution is the only memoryless distribution! Here is a sketch argument. Suppose $X$ is memoryless and consider the function

$$
G(x)=\mathbb{P}\{X>x\}
$$

To be memoryless is equivalent to

$$
G(s+t)=G(s) G(t)
$$

Observe first that setting $s=0$ results in

$$
G(t)=G(0) G(t) \Longrightarrow G(0)=1=\mathbb{P}\{X \geq 0\}
$$

It follows that the distribution only takes non-negative values. Now observe that

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \ln G(t)=\frac{G^{\prime}(t)}{G(t)}=\lim _{t \rightarrow 0} \frac{G(t+s)-G(t)}{s G(t)}=\lim _{s \rightarrow 0} \frac{G(s)-G(0)}{s}=G^{\prime}(0)
$$

is constant! If we define $\lambda:=-G^{\prime}(0)$, we quickly see that $G(t)=e^{-\lambda t}$, as required.
Of course this runs very much counter to our intuition, and often to reality! To labor the point...
Example 5.23. Two passengers are waiting on opposite platforms for trains. The time until each train arrives is a random variable $X \sim \operatorname{Exponential}\left(\frac{1}{20}\right)$. The first train arrives after 10 minutes. How long do we expect to have to wait for the second train to arrive?
Since the second person's wait has no memory of the first 10 minutes, the remaining time is still exponentially distributed with the same parameter $\lambda=\frac{1}{20}$. We therefore expect them to wait an additional 20 minutes, for a total of 30 minutes.
This seems unreasonable. Suppose the wait time for the second person's train was merely some, not necessarily exponential, distribution $X$ with the same mean 20. Let $Y$ be the additional time required for the next train. Then

$$
F_{Y}(y)=\mathbb{P}\{Y \leq y\}=\mathbb{P}\{X \leq y+10 \mid X>10\}=\frac{\mathbb{P}\{10<X \leq y+10\}}{\mathbb{P}\{X>10\}}=\frac{F_{X}(y+10)-F_{X}(10)}{1-F_{X}(10)}
$$

[^2]where $F_{X}(x)$ is the cumulative distribution function for $X$. It follows that the density for $Y$ is
$$
f_{Y}(y)=\frac{f_{X}(y+10)}{1-F_{X}(10)} \Longrightarrow \mathbb{E}[Y]=\frac{1}{1-F_{X}(10)} \int_{0}^{\infty} y f_{X}(y+10) \mathrm{d} y
$$

To evaluate this explicitly, we need more information about $X$. Let's perform two sanity checks:

- If $X \sim \operatorname{Exponential}\left(\frac{1}{20}\right)$, then

$$
\mathbb{E}[Y]=\frac{1}{e^{-\frac{10}{20}}} \int_{0}^{\infty} \frac{y}{20} e^{-\frac{y+10}{20}} \mathrm{~d} y=\int_{0}^{\infty} \frac{y}{20} e^{-\frac{y}{20}} \mathrm{~d} y=\mathbb{E}[X]=20 \text { minutes }
$$

as before.

- If $X \sim U(0,40)$ were uniformly distributed with the same mean 20, then

$$
\mathbb{E}[Y]=\frac{1}{1-\frac{1}{4}} \int_{0}^{30} \frac{y}{40} \mathrm{~d} y=\frac{4}{3} \cdot \frac{30^{2}}{80}=15 \text { minutes }
$$

which results in the second passenger waiting 25 minutes in total. In line with our intuition, the fact that the passenger had already waited 10 minutes means that the expected time to the next train is now less. Indeed, this fact should be obvious without the integral: after the first 10 minutes, the second train's arrival would have to be uniformly distributed $Y \sim U(0,30)$.

Hazard Rates Variations on the above are common when analyzing the lifetime of an object. Suppose the life of an object is measured by a random variable $T$. If the object has survived until time $t$, we can ask the probability that it will die in the interval $(t, t+\Delta t)$.

$$
\mathbb{P}\{t<T<t+\Delta t \mid T>t\}=\frac{\mathbb{P}\{t<T<t+\Delta t\}}{\mathbb{P}\{X>t\}}=\frac{F(t+\Delta t)-F(t)}{1-F(t)} \approx \frac{f(t)}{1-F(t)} \Delta t
$$

The failure or hazard rate of $T$ is the expression

$$
\lambda(t)=\frac{f(t)}{1-F(t)}
$$

Example 5.24. If $T$ is the human lifespan measured in years, then the hazard rate $\lambda(t)$ represents (approximately) the probability that a person aged $t$ will die within the next year. Data tables such as these are used by actuaries to price life insurance policies and by the Social Security Administration to analyze expected payouts and therefore assess the health of the program. Some snippets according to Social Security:

- The safest age, for both sexes, is roughly $9-10$, where the hazard rate is roughly 1 in 10,000 ; that is, roughly one in every 10,000 nine year olds will die every year.
- By the early 20 's, the hazard rate for men has increased tenfold to over $\frac{1}{1000}$, more than double that for young women. It stays significantly higher for men thereafter.
- A nine year old has the same risk of death in a random 20 minute period as one ticket has of winning the jackpot on the Mega Millions (300 million to 1).

The hazard rate in fact determines the distribution: since $F^{\prime}(t)=f(t)$, we see that

$$
\begin{aligned}
\lambda(t)=\frac{F^{\prime}(t)}{1-F(t)} & \Longrightarrow \int_{0}^{t} \lambda(x) \mathrm{d} x=-\ln (1-F(t)) \\
& \Longrightarrow F(t)=1-\exp \left(-\int_{0}^{t} \lambda(x) \mathrm{d} x\right) \\
& \Longrightarrow f(t)=\lambda(t) \exp \left(-\int_{0}^{t} \lambda(x) \mathrm{d} x\right)
\end{aligned}
$$

The exponential distribution is the unique distribution with a constant hazard rate, reflecting the fact that it is memoryless. This is also why we use the symbol $\lambda$ in both cases.

Examples 5.25. 1. Suppose a distribution $T$ taking positive values has hazard rate $\lambda(t)=\frac{1}{t+1}$. The cumulative distribution function is then

$$
F(t)=1-\exp \left(-\int_{0}^{t}(x+1)^{-1} \mathrm{~d} x\right)=1-\exp (-\ln (t+1))=1-\frac{1}{t+1}
$$

2. The mileage obtained from a set of tires on used on normal roads is given by a distribution $X$ with hazard rate $\lambda(x)$. A set of tires with 20,000 miles has probability

$$
\mathbb{P}\{X>30,000 \mid X>20,000\}=0.9
$$

of lasting another 10,000 miles. The manufacturer states that the risk of tire failure is 10 times higher when used for off-road driving. If the tires were used off-road, what is the probability that a set with 20,000 miles will need replaced by 30,000 miles?

Let $Y$ be the lifetime of the tires when used off-road. We have

$$
\begin{aligned}
\mathbb{P}\{Y>30,000 \mid Y>20,000\} & =\frac{\mathbb{P}\{Y>30,000\}}{\mathbb{P}\{Y>20,000\}}=\exp \left(-\int_{20000}^{30000} 10 \lambda(x) \mathrm{d} x\right) \\
& =\left[\exp \left(-\int_{20000}^{30000} \lambda(x) \mathrm{d} x\right)\right]^{10} \\
& =(\mathbb{P}\{X>30,000 \mid X>20,000\})^{10} \\
& =0.9^{10}=34.9 \%
\end{aligned}
$$

The off-road tires will need replaced with probability $65.1 \%$.


[^0]:    ${ }^{1}$ The negative here is crucial when you change the order of integration! It is also intuitive if you think about the signs of the integrals: $\int_{-\infty}^{0} t f_{Y}(t) \mathrm{d} t \leq 0$ and $\int_{-\infty}^{0} \mathbb{P}\{g(X)<y\} \mathrm{d} y \geq 0$.

[^1]:    ${ }^{2}$ If you have $\$ 1000$ at the start of an investment period, then at the end you will have $\$ 1000 Y$.
    ${ }^{3}$ In financial math parlance this insurance policy amounts to a call option with asset price $S=\$ 4000$, strike price $K=\$ 5000$, volatility $\sigma=\sqrt{0.1}=0.316$, and risk-free rate of return $r=0$ (what you'd obtain putting money in the bank or what you'd pay to borrow money). If $r>0$, then $X \sim N\left(r-\frac{1}{2} \sigma^{2}, \sigma^{2}\right)$. The idea is that the expected return from investing in corn should equal the risk-free rate, otherwise everyone would invest in corn! We'd also need to discount the premium back to the start time using the risk-free rate, thus recovering the standard Black-Scholes price for a call

[^2]:    ${ }^{4}$ This is why our argument is only a sketch. We assume that $G(x)=\mathbb{P}\{X>x\}=1-\mathbb{P}\{X \leq x\}$ is differentiable, when it need only be right-continuous. It isn't too hard to fix this flaw.

