# Probability Theory

## Probability Spaces and Events

Consider a random experiment with several possible outcomes. For example, we might roll a pair of dice, flip a coin three times, or choose a random real number between 0 and 1. The **sample space** for such an experiment is the set of all possible outcomes. For example:

- The sample space for a pair of die rolls is the set

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}.$$

- The sample space for three coin flips is the set $\{0, 1\}^3$, where 0 represents heads and 1 represents tails.

- The sample space for a random number between 0 and 1 is the interval $[0, 1]$.

An **event** is any statement about the outcome of an experiment to which we can assign a probability. For example, if we roll a pair of dice, possible events include:

- "Both dice show even numbers."

- "The first die shows a 5 or 6."

- "The sum of the values shown by the dice is greater than or equal to 7."

From a formal point of view, events are usually defined to be certain subsets of the sample space. Thus the event "both dice show even numbers" refers to the subset $\{2, 4, 6\} \times \{2, 4, 6\}$. Despite this, it is more common to write statements than subsets when referring to a specific event.

In the special case of an experiment with finitely many outcomes, we can define the probability of any subset of the sample space, and therefore every subset is an event. In the general case, however, probability is a measure on the sample space, and only measurable subsets of the sample space are events.

The following definition serves as the foundation of modern probability theory:

**Definition: Probability Space**
A **probability space** is an ordered triple $(\Omega, \mathcal{E}, P)$, where:

- $\Omega$ is a set (the **sample space**). Elements of $\Omega$ are called **outcomes**.
- $\mathcal{E}$ is a $\sigma$-algebra over $\Omega$. Elements of $\mathcal{E}$ are called **events**.
- $P\colon \mathcal{E} \to [0,1]$ is a measure satisfying $P(\Omega) = 1$. This is the **probability measure** on $\Omega$.

In general, a measure $\mu\colon \mathcal{M} \to [0,\infty]$ on a set $S$ is called a **probability measure** if $\mu(S) = 1$, in which case the triple $\big(S, \mathcal{M}, \mu\big)$ forms a probability space.

**EXAMPLE 1** Finite Probability Spaces
Consider an experiment with finitely many outcomes $\omega_1, \ldots, \omega_n$, with corresponding probabilities $p_1, \ldots, p_n$. Such an experiment corresponds to a finite probability space $(\Omega, \mathcal{E}, P)$, where:

- $\Omega$ is the set $\{\omega_1, \ldots, \omega_n\}$,
- $\mathcal{E}$ is the collection of all subsets of $\Omega$, and
- $P\colon \mathcal{E} \to [0,1]$ is the probability measure on $\Omega$ defined by the formula

$$P(\{\omega_{i_1}, \ldots, \omega_{i_k}\}) \;=\; p_{i_1} + \cdots + p_{i_k}. \qquad \blacksquare$$

**EXAMPLE 2** The Interval $[0,1]$
Consider an experiment whose outcome is a random number between 0 and 1. We can model such an experiment by the probability space $(\Omega, \mathcal{E}, P)$, where:

- $\Omega$ is the interval $[0,1]$,
- $\mathcal{E}$ is the collection of all Lebesgue measurable subsets of $[0,1]$, and
- $P\colon \mathcal{E} \to [0,1]$ is Lebesgue measure on $[0,1]$.

Using this model, the probability that the outcome lies in a given set $E \subset [0,1]$ is equal to the Lebesgue measure of $E$. For example, the probability that the outcome is rational is 0, and the probability that the outcome lies between 0.3 and 0.4 is $1/10$. $\blacksquare$

**EXAMPLE 3** Product Spaces
Let $(\Omega_1, \mathcal{E}_1, P_1)$ and $(\Omega_2, \mathcal{E}_2, P_2)$ be probability spaces corresponding to two possible

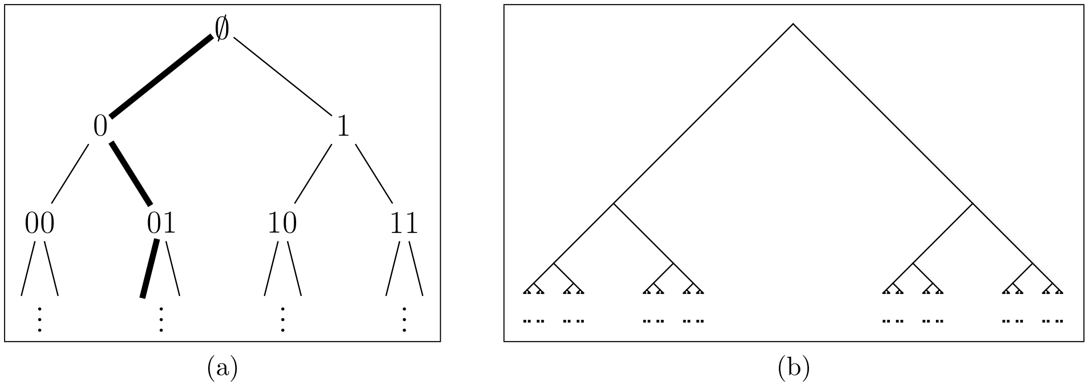



Figure 1: (a) Each infinite sequence of coin flips corresponds to a path down an infinite binary tree. In this case, the sequence begins with 010. (b) The leaves of an infinite binary tree form a Cantor set.

experiments. Now, imagine that we perform both experiments, recording the outcome for each. The combined outcome for this experiment is an ordered pair $(\omega_1, \omega_2)$, where $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. In fact, the combined experiment corresponds to a probability space $(\Omega, \mathcal{E}, P)$, where:

- $\Omega$ is the Cartesian product $\Omega_1 \times \Omega_2$.
- $\mathcal{E}$ is the $\sigma$-algebra generated by all events of the form $E_1 \times E_2$, where $E_1 \in \mathcal{E}_1$ and $E_2 \in \mathcal{E}_2$.
- $P\colon \mathcal{E} \to [0,1]$ is the product of the measures $P_1$ and $P_2$. That is, $P$ is the unique measure with domain $\mathcal{E}$ satisfying $P(E_1 \times E_2) = P_1(E_1)\, P_2(E_2)$ for all $E_1 \in \mathcal{E}_1$ and $E_2 \in \mathcal{E}_2$.

For example, if we pick two random numbers between 0 and 1, the corresponding sample space is the square $[0,1] \times [0,1]$, with the probability measure being two-dimensional Lebesgue measure. ■

**EXAMPLE 4** The Cantor Set

Suppose we flip an infinite sequence of coins, recording 0 for heads and 1 for tails. The outcome of this experiment will be an infinite sequence $(1, 0, 1, 1, 0, 1, 0, 0, 1, \ldots)$ of 0's and 1's, so the sample space $\Omega$ is the infinite product $\{0,1\}^{\mathbb{N}}$.

Although the sample space $\{0,1\}^{\mathbb{N}}$ is infinite-dimensional, we can visualize each outcome as a path down an infinite binary tree, as shown in Figure 1a. If we imagine that this tree has "leaves" at infinity, then each outcome corresponds to one such leaf. Indeed, it is possible to draw this tree so that the leaves are visible, as shown in Figure 1b. As you can see, the leaves of the tree form a Cantor set. Indeed, under the product topology, the sample space $\Omega = \{0,1\}^{\mathbb{N}}$ is homeomorphic to the standard middle-thirds Cantor set.

It is not too hard to put a measure on $\Omega$. Given a finite sequence $b_1, \ldots, b_n$ of 0's and 1's, let $B(b_1, \ldots, b_n)$ be the set of outcomes whose first $n$ flips are $b_1, \ldots, b_n$, and define

$$P_0\big(B(b_1, \ldots, b_n)\big) \;=\; \frac{1}{2^n}.$$

Let $\mathcal{B}$ be the collection of all such sets, and let

$$P^*(E) \;=\; \inf\left\{\textstyle\sum P_0(B_n) \;\middle|\; B_1, B_2, \ldots \in \mathcal{B} \text{ and } E \subset \bigcup B_n\right\}$$

for every $E \subset \Omega$. Then $P^*$ is an outer measure on $\Omega$, and the resulting measure $P$ is a probability measure. ■

The mechanism described above for putting a measure on $\{0,1\}^{\mathbb{N}}$ can be modified to put a measure on $\Omega^{\mathbb{N}}$ for any probability space $(\Omega, \mathcal{E}, P)$. For example, it is possible to talk about an experiment in which we roll an infinite sequence of dice, or pick an infinite sequence of random numbers between 0 and 1, and for each of these there is a corresponding probability space.

## Random Variables

A **random variable** is a quantity whose value is determined by the results of a random experiment. For example, if we roll two dice, then the sum of the values of the dice is a random variable.

In general, a random variable may take values from any set $S$.

> **Definition: Random Variable**
> A **random variable** on a probability space $(\Omega, \mathcal{E}, P)$ is a function $X\colon \Omega \to S$, where $S$ is any set.

In the case where $S = \mathbb{R}$ (or more generally if $S$ is a topological space), we usually require a random variable to be a *measurable* function, i.e. $X^{-1}(U)$ should be measurable for every open set $U \subseteq S$.

**EXAMPLE 5** Here are some basic examples of random variables:

- If we roll two dice, then the values $X$ and $Y$ that show on the dice are random variables $\Omega \to \{1,2,3,4,5,6\}$. Expressions involving $X$ and $Y$, such as the sum $X + Y$ or the quantity $X^2 + Y^3$, are also random variables.

- If we pick a random number between 0 and 1, then the number $X$ that we pick is a random variable $\Omega \to [0,1]$.
- For an infinite number of coin flips, we can define a sequence $C_1, C_2, \ldots$ of random variables $\Omega \to \{0,1\}$ by

$$C_n = \begin{cases} 0 & \text{if the } n\text{th flip is heads,} \\ 1 & \text{if the } n\text{th flip is tails.} \end{cases}$$ ■

Although a random variable is defined as a function, we usually think of it as a variable that depends on the outcome $\omega \in \Omega$. In particular, we will often write $X$ when we really mean $X(\omega)$. For example, if $X$ is a real-valued random variable, then

$$P(X \leq 3)$$

would refer to

$$P\big(\{\omega \in \Omega \mid X(\omega) \leq 3\}\big).$$

**Definition: Distribution of a Random Variable**
The **distribution** of a random variable $X\colon \Omega \to S$ is the probability measure $P_X$ on $S$ defined by

$$P_X(T) \;=\; P(X \in T).$$

for any measurable set $T \subset S$.

The expression $P(X \in T)$ in the definition above refers to the probability that the value of $X$ is an element of $T$, i.e. the probability of the event $X^{-1}(T)$. Thus the distribution of $X$ is defined by the equation

$$P_X(T) \;=\; P\big(X^{-1}(T)\big)$$

Note that the set $X^{-1}(T)$ is automatically measurable since $X$ is a measurable function. In measure-theoretic terms, $P_X$ is the pushforward of the probability measure $P$ by to the set $S$ via $X$.

**EXAMPLE 6** Die Roll
Let $X\colon \Omega \to \{1,2,3,4,5,6\}$ be the value of a die roll. Then for any subset $T$ of $\{1,2,3,4,5,6\}$, we have

$$P_X(T) \;=\; P(X \in T) \;=\; \frac{|T|}{6}.$$

That is, $P_X$ is the probability measure on $\{1,2,3,4,5,6\}$ where each point has probability $1/6$. ■

**EXAMPLE 7** Random Real Number
Let $X\colon \Omega \to [0,1]$ be a random real number between 0 and 1. Assuming $X$ is equally likely to lie in any portion of the interval $[0,1]$, the distribution $P_X$ is just Lebesgue measure on $[0,1]$. This is known as the **uniform distribution** on $[0,1]$. ■

The most basic type of random variable is a discrete variable:

**Definition: Discrete Random Variable**
A random variable $X\colon \Omega \to S$ is said to be **discrete** if $S$ is finite or countable.

If $X\colon \Omega \to S$ is discrete, then the probability distribution $P_X$ for $X$ is completely determined by the probability of each element of $S$. In particular:

$$P_X(T) \;=\; \sum_{x\in T} P_X(\{x\})$$

for any subset $T$ of $S$.

**EXAMPLE 8** Difference of Dice
Let $X$ and $Y$ be random die rolls, and let $Z = |X - Y|$. Then $Z$ is a random variable $\Omega \to \{0,1,2,3,4,5\}$, with the following probability distribution:

$$P_Z(\{0\}) \;=\; \frac{1}{6}, \qquad P_Z(\{1\}) \;=\; \frac{10}{36}, \qquad P_Z(\{2\}) \;=\; \frac{8}{36},$$

$$P_Z(\{3\}) \;=\; \frac{6}{36}, \qquad P_Z(\{4\}) \;=\; \frac{4}{36}, \qquad P_Z(\{5\}) \;=\; \frac{2}{36}.$$ ■

We end with a useful formula for integrating with respect to a probability distribution. This is essentially just a restatement of the formula for the Lebesgue integral with respect to a pushforward measure.

## Proposition 1 Integration Formula for Distributions

*Let $X\colon \Omega \to S$ be a random variable, and let $g\colon S \to \mathbb{R}$ be a measurable function. Then*

$$\int_S g\, dP_X \;=\; \int_\Omega g(X)\, dP,$$

*where $g(X)$ denotes the random variable $g \circ X\colon \Omega \to \mathbb{R}$.*

## Continuous Random Variables

One particularly important type of random variable is a continuous variable on $\mathbb{R}$.

**Definition: Continuous Random Variable**
Let $X\colon \Omega \to \mathbb{R}$ be a random variable with probability distribution $P_X$. We say that $X$ is **continuous** if there exists a measurable function $f_X\colon \mathbb{R} \to [0,\infty]$ such that

$$P_X(T) \;=\; \int_T f_X\,dm$$

for every measurable set $T \subset \mathbb{R}$, where $m$ denotes Lebesgue measure. In this case, the function $f_X$ is called a **probability density function** for $X$.

That is, $X$ is continuous if $P_X$ is absolutely continuous with respect to Lebesgue measure, i.e. if

$$dP_X \;=\; f_X\,dm$$

for some non-negative measurable function $f_X$. Recall the following formula for integrals with respect to such measures.

### Proposition 2 Weighted Integration

*Let $X\colon \Omega \to \mathbb{R}$ be a continuous random variable with probability density $f_X$, and let $g\colon \mathbb{R} \to \mathbb{R}$ be a measurable function. Then*

$$\int_{\mathbb{R}} g\,dP_X \;=\; \int_{\mathbb{R}} g f_X\,dm.$$

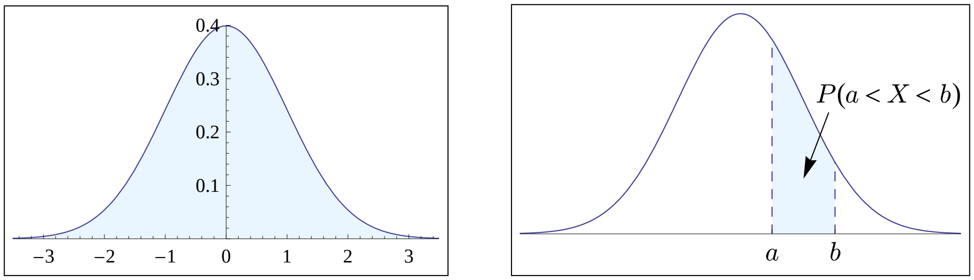
**EXAMPLE 9** Standard Normal Distribution
Consider a random variable $X\colon \Omega \to \mathbb{R}$ with probability density function defined by

$$f_X(x) \;=\; \frac{1}{\sqrt{2\pi}}\exp\!\left(-\frac{1}{2}x^2\right).$$

In this case, $X$ is said to have the **standard normal distribution**. A graph of the function $f_X$ is shown in Figure 2a.

For such an $X$, the probability $P_X(T)$ that the value $X$ lies in any set $T$ is given by the formula

$$P_X(T) \;=\; \int_T f_X\,dm.$$




(a) (b)

Figure 2: (a) The probability density $f_X$ for a standard normal distribution. (b) Area under the graph of $f_X$ on the interval $(a, b)$.

For example, if $(a, b)$ is an open interval, then

$$P(a < X < b) \;=\; \int_{(a,b)} f_X \, dm \;=\; \int_a^b \frac{1}{\sqrt{2\pi}} \exp\!\left(-\frac{1}{2}x^2\right) dx.$$

This is illustrated in Figure 2b. ■

The probability density function $f_X$ can be thought of as describing the probability per unit length of $P_X$. The following theorem gives us a formula for this function:

## **Theorem 3** Density Formula

*Let $X : \Omega \to \mathbb{R}$ be a random variable, and define a function $f_X : \mathbb{R} \to [0, \infty]$ by*

$$f_X(x) \;=\; \lim_{h \to 0^+} \frac{P_X\big([x-h, x+h]\big)}{2h}.$$

*Then $f_X(x)$ is defined for almost all $x \in \mathbb{R}$ and $f_X$ is measurable. Moreover:*

**1.** *If $\displaystyle\int_{\mathbb{R}} f_X \, dm = 1$, then $X$ is a continuous random variable.*

**2.** *If $X$ is continuous, then $f_X$ is a probability density function for $X$.*

**PROOF** This is related the Lebesgue differentiation theorem, though it does not follows from it immediately. ■

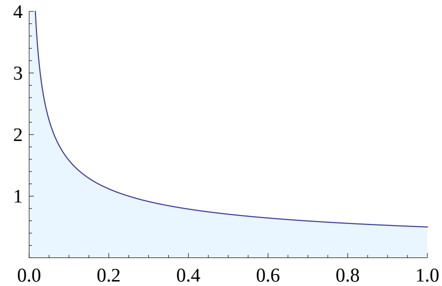



Figure 3: Probability density for $X^2$, where $X$ is chosen uniformly from $[0,1]$.

**EXAMPLE 10** Density of $X^2$

Let $X\colon \Omega \to [0,1]$ be uniformly distributed, and let $Y = X^2$. Then for any interval $[a,b] \subset [0,1]$, we have

$$Y \in [a,b] \qquad \Leftrightarrow \qquad X \in \left[\sqrt{a}, \sqrt{b}\,\right]$$

so

$$P_Y\big([a,b]\big) \;=\; P_X\big([\sqrt{a}, \sqrt{b}\,]\big) \;=\; \sqrt{b} - \sqrt{a}.$$

Therefore,

$$f_Y(x) \;=\; \lim_{h\to 0^+} \frac{P_Y\big([x-h, x+h]\big)}{2h} \;=\; \lim_{h\to 0^+} \frac{\sqrt{x+h} - \sqrt{x-h}}{2\epsilon} \;=\; \frac{1}{2\sqrt{x}}.$$

A plot of this function is shown in Figure 3. Note that the total area under the graph of the function is 1, which proves that $Y$ is continuous, and that $f_Y$ is a probability density function for $Y$. ■

Finally, it should be mentioned that most of the discussion of continuous random variables generalizes to random variables $X\colon \Omega \to \mathbb{R}^n$. Such a variable is called continuous if there exists a function $f_X\colon \mathbb{R}^n \to [0,\infty]$ so that

$$P_X(T) \;=\; \int_T f_X \, d\mu$$

for every measurable $T \subset \mathbb{R}^n$, where $\mu$ denotes Lebesgue measure on $\mathbb{R}^n$. The probability density for such a variable is determined by the formula

$$f_X(p) \;=\; \lim_{r\to 0^+} \frac{P_X\big(B(p,r)\big)}{\mu\big(B(p,r)\big)},$$

where $B(p,r)$ denotes a ball of radius $r$ in $\mathbb{R}^n$ centered at the point $p$.

# Expected Value

**Definition: Expected Value**
Let $X\colon \Omega \to \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{E}, P)$. The **expected value** of $X$ is defined as follows:

$$EX \;=\; \int_\Omega X\,dP.$$

The expected value is sometimes called the **average value** or **mean** of $X$, and is also denoted $\overline{X}$ or $\langle X \rangle$.

It is possible to calculate the expected value of a random variable directly from the distribution:

## Proposition 4 Expected Value for a Distribution

*Let $X\colon \Omega \to \mathbb{R}$ be a random variable with distribution $P_X$. Then*

$$EX \;=\; \int_{\mathbb{R}} x\,dP_X(x).$$

**PROOF** Let $g\colon \mathbb{R} \to \mathbb{R}$ be the function $g(x) = x$. By the integration formula for distributions (Proposition 1), we have

$$\int_{\mathbb{R}} x\,dP_X(x) \;=\; \int_{\mathbb{R}} g\,dP_X \;=\; \int_\Omega g(X)\,dP \;=\; \int_\Omega X\,dP \;=\; EX. \qquad \blacksquare$$

**EXAMPLE 11** Expected Value of a Discrete Variable
Let $X\colon \Omega \to S$ be a discrete random variable, where $S$ is a finite subset of $\mathbb{R}$. Then it follows from the above theorem that

$$EX \;=\; \sum_{x \in S} x\,P(\{x\}).$$

For example, if $X$ is the value of a die roll, then

$$EX \;=\; 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) \;=\; 3.5. \qquad \blacksquare$$

**EXAMPLE 12** Expected Value of a Continuous Variable
Let $X\colon \Omega \to \mathbb{R}$ be a continuous random variable with probability density function $f_X$. Then it follows from the weighted integration formula (Proposition 2) that

$$EX \;=\; \int_{\mathbb{R}} x\, f_X(x)\, dm(x).$$

For example, if $X$ is a random number from $[0,1]$ with uniform distribution, then

$$EX \;=\; \int_{[0,1]} x\, dm(x) \;=\; \frac{1}{2}.$$

For another example, consider the random variable $Y = X^2$. The probability density function for $Y$ is $f_Y(x) = 1/(2\sqrt{x})$ (see Example 10), so

$$EY \;=\; \int_{[0,1]} x\, \frac{1}{2\sqrt{x}}\, dm(x) \;=\; \frac{1}{3}.$$ ■

**EXAMPLE 13** Infinite Expected Value
It is possible for a random variable to have infinite expected value. For example, let $X\colon \Omega \to [1,\infty)$ be a continuous random variable with

$$f_X(x) \;=\; \frac{1}{x^2}.$$

Note that $\int_{[1,\infty)} f_X\, dm = 1$, so this is indeed a legitimate probability density function. Then the expected value of $X$ is

$$EX \;=\; \int_{[1,\infty)} x\, f_X(x)\, dm(x) \;=\; \int_1^\infty \frac{1}{x}\, dx \;=\; \infty.$$

The same phenomenon can occur for a discrete random variable. For example, if $X\colon \Omega \to \mathbb{N}$ satisfies

$$P_X(\{n\}) \;=\; \frac{1}{Cn^2},$$

where $C = \sum_{n=1}^\infty 1/n^2 = \pi^2/6$, then

$$EX \;=\; \sum_{n=1}^\infty n\, P_X(\{n\}) \;=\; \sum_{n=1}^\infty \frac{1}{Cn} \;=\; \infty.$$

Note that both of these examples involve positive random variables. For a general variable $X\colon \Omega \to \mathbb{R}$, it is also possible for $EX$ to be entirely undefined. ■

The following proposition lists some basic properties of expected value. These follow directly from the corresponding properties for the Lebesgue integral:

**Proposition 5** Properties of Expected Value

*Let $X\colon \Omega \to \mathbb{R}$ and $Y\colon \Omega \to \mathbb{R}$ be random variables with finite expected value. Then:*

1. *If $C \in \mathbb{R}$ is constant, then $EC = C$.*
2. *$E[\alpha X] = \alpha\, EX$ for any $\alpha \in \mathbb{R}$.*
3. *$E[X+Y] = EX + EY$.*
4. *$|EX| \leq E|X|$.*
5. *If $X \geq 0$, then $EX \geq 0$, with equality if and only if $P(X=0) = 1$.*

## Variance and Standard Deviation

There is much more to the distribution of a variable than its expected value. For example, the average length of an adult blue whale is about 80 feet, but this tells us very little about whale lengths: are most blue whales about 75–85 feet long, or is the range more variable, say 20–140 feet? The tendency of a random variable to take values far away from the mean is called **dispersion**. Two common measures of dispersion are the variance and standard deviation:

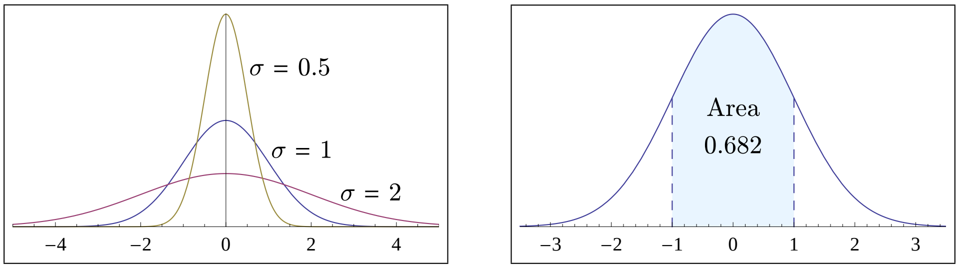
**Definition: Variance and Standard Deviation**
Let $X\colon \Omega \to \mathbb{R}$ be a random variable with finite expected value $\mu$. The **variance** of $X$ is the quantity

$$\mathrm{Var}(X) \;=\; E\big[(X-\mu)^2\big].$$

The square root of $\mathrm{Var}(X)$ is called the **standard deviation** of $X$.

Though the variance has nicer theoretical properties, the standard deviation has more meaning, since it is measured in the same units as $X$. For example, if $X$ is a random variable with units of length, then the standard deviation of $X$ is measured in feet, while the variance is measured in square feet. The standard deviation in the length of an adult blue whale is about 10 feet, meaning that most whales are about 70–90 feet long.

To give you a feel for these quantities, Figure 4a shows several normal distributions with different standard deviations. The standard normal distribution has standard deviation 1, and normal distributions with standard deviations of 0.5 and 2 have also been plotted. As a general rule, a normally distributed random variable has roughly




(a) (b)

Figure 4: (a) Normal distributions with standard deviations of 0.5, 1, and 2. (b) A normally distributed variable has a 68.2% chance of being within one standard deviation of the mean.

a 68.2% probability of being within one standard deviation of the mean, as shown in Figure 4b. In particular,

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \;\approx\; 0.682689.$$

For example, if we assume that the length of a random adult blue whale is normally distributed with a mean of 80 feet and a standard deviation of 10 feet, then approximately 68% of blue whales will be between 70 and 90 feet long.

As with expected value, we can compute the variance of a random variable directly from the distribution:

## Proposition 6 Variance of a Distribution

*Let $X\colon \Omega \to \mathbb{R}$ be a random variable with finite mean $\mu$ and distribution $P_X$. Then*

$$\mathrm{Var}(X) \;=\; \int_{\mathbb{R}} (x-\mu)^2 \, dP_X(x).$$

**PROOF** Let $g\colon \mathbb{R} \to \mathbb{R}$ be the function $g(x) = (x-\mu)^2$. By the integration formula for distributions (Proposition 1), we have

$$\int_R g \, dP_X \;=\; \int_\Omega g(X)\, dP \;=\; \int_\Omega (X-\mu)^2 \, dP \;=\; E\big[(X-\mu)^2\big] \;=\; \mathrm{Var}(X). \qquad \blacksquare$$

**EXAMPLE 14** Variance of a Discrete Variable
Let $X : \Omega \to S$ be a discrete random variable, where $S \subset \mathbb{R}$, and let $\mu = EX$. Then

$$\mathrm{Var}(X) \;=\; \sum_{x \in S} (x-\mu)^2 \, P(\{x\}).$$

For example, if $X$ is the value of a die roll, then $EX = 3.5$, so

$$\mathrm{Var}(X) \;=\; \sum_{n=1}^{6} \frac{(n-3.5)^2}{6} \;\approx\; 2.92.$$

Taking the square root yields a standard deviation of approximately 1.71. ■

**EXAMPLE 15** Variance of a Continuous Variable
Let $X : \Omega \to \mathbb{R}$ be a continuous random variable with probability density function $f_X$. Then it follows from the weighted integration formula (Proposition 2) that

$$\mathrm{Var}(X) \;=\; \int_{\mathbb{R}} (x-\mu)^2 f_X(x)\, dm(x).$$

For example, if $X$ is a random number from $[0,1]$ with uniform distribution, then $EX = 1/2$, so

$$\mathrm{Var}(X) \;=\; \int_{[0,1]} \left(x - \frac{1}{2}\right)^2 dm(x) \;=\; \frac{1}{12}.$$

Taking the square root yields a standard deviation of approximately 0.29. ■

**EXAMPLE 16** Normal Distributions
A random variable $X : \Omega \to \mathbb{R}$ is said to be **normally distributed** if it has a probability density function of the form.

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

Such a variable has mean $\mu$ and standard deviation $\sigma$, as can be seen from the following integral formulas:

$$\int_{\mathbb{R}} \frac{x}{\sigma\sqrt{2\pi}} \exp\!\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dm(x) \;=\; \mu.$$

$$\int_{\mathbb{R}} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} \exp\!\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dm(x) \;=\; \sigma^2.$$ ■

**EXAMPLE 17** Infinite Variance
As with the expected value, it is possible for the variance of a random variable to be infinite or undefined. For example, let $X\colon \Omega \to \mathbb{R}$ be a random variable with distribution

$$f_X \;=\; \frac{C}{1+|x|^3}$$

where $C = 3\sqrt{3}/(4\pi)$. Then $X$ has finite expected value, with

$$EX \;=\; \int_{\mathbb{R}} \frac{Cx}{1+|x|^3}\, dm(x) \;=\; 0.$$

But

$$\operatorname{Var}(X) \;=\; \int_{\mathbb{R}} \frac{Cx^2}{1+|x|^3}\, dm(x) \;=\; \infty.$$

From a measure-theoretic point of view, this random variable is an $L^1$ function on the probability space $\Omega$, but it is not $L^2$. ■

We end by stating another standard formula for variance.

### Proposition 7 Alternative Formula for Variance

*Let $X\colon \Omega \to \mathbb{R}$ be a random variable with finite expected value. Then*

$$\operatorname{Var}(X) \;=\; E[X^2] - (EX)^2.$$

**PROOF** Let $\mu = EX$. Then

$$\begin{aligned}\operatorname{Var}(X) \;&=\; E\big[(X-\mu)^2\big] \;=\; E\big[X^2 - 2\mu X + \mu^2\big] \\ &=\; E[X^2] - 2\mu(EX) + \mu^2 \;=\; E[X^2] - 2\mu^2 + \mu^2 \;=\; E[X^2] - \mu^2.\end{aligned}$$

■

It follows from this formula that $\operatorname{Var}(X)$ is finite if and only if $E[X^2] < \infty$, i.e. if and only if $X \in L^2(\Omega)$.

## Exercises

1. Let $X\colon \Omega \to (0,1]$ have the uniform distribution, and let $Y = 1/X$.
   a) Find the probability density function $f_Y\colon [1,\infty) \to [0,\infty]$ for $Y$.

b) What is the expected value of $Y$?

2. Let $X\colon \Omega \to [0,1]$ have the uniform distribution, and let $Y = \sin(8X)$. Use the integration formula for distributions (Proposition 1) to compute $EY$.

3. Let $X$ and $Y$ be the values of two die rolls, and let $Z = \max(X, Y)$.

   a) Find $P_Z(\{n\})$ for $n \in \{1, 2, 3, 4, 5, 6\}$.

   b) Determine $EZ$, and find the standard deviation for $Z$.

4. Let $X\colon \Omega \to [1, \infty)$ be a continuous random variable with probability density function
$$f_X(x) = \frac{3}{x^4}.$$
Compute $EX$ and $\mathrm{Var}(X)$.