Chapter 2
Random Variables

2.1 Discrete Random Variables

We start with an example.

Example 2.1. Toss two fair coins. Let $X$ be the number of heads. $X$ is a function from the sample space $\Omega = \{HH, HT, TH, TT\}$ into the set $\{0, 1, 2\}$. The distribution of $X$ is given by the following table.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P(X = k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

More generally, we have the following definition.

Discrete Random Variables

A discrete random variable is a function from a sample space $\Omega$ into a countable set (usually the positive integers). The distribution of a random variable $X$ is the sequence of probabilities $P(X = k)$ for all $k$ in the range of $X$. We must have

$$P(X = k) \geq 0 \text{ for every } k \text{ and } \sum_k P(X = k) = 1.$$ 

The term discrete refers to the fact that the random variables, in this section, take values in countable sets. Next section deals with continuous random variables: random variables whose range include intervals of the real numbers. We now give several examples of the important discrete random variables.
2.1.1 Bernoulli Random Variables

These are the simplest possible random variables. Perform a random experiment with two possible outcomes: success or failure. Set $X = 1$ if the experiment is a success and $X = 0$ if the experiment is a failure. Such a 0–1 random variable is called a Bernoulli random variable. The usual notation is $P(X = 1) = p$ and $P(X = 0) = q = 1 - p$.

Example 2.2. Roll a fair die. We say that we have a success if we roll a 6. Thus, the probability of success is $P(X = 1) = 1/6$. We have $p = 1/6$ and $q = 5/6$.

2.1.2 Discrete Uniform Random Variables

Example 2.3. Roll a fair die. Let $X$ be the face shown. The distribution of $X$ is given by the following table.

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = k)$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
</tbody>
</table>

Below we graph this distribution.

This is called a uniform random variable. Uniform refers to the fact that all possible values of $X$ are equally likely.

2.1.3 Geometric Random Variable

Example 2.4. Roll a fair die until you get a 6. Let $X$ be the number of rolls to get the first 6. The possible values of $X$ are all strictly positive integers. Note that $X = 1$ if and only if the first roll is a 6. So $P(X = 1) = 1/6$. In order to have $X = 2$ the first roll must be anything but 6 and the second one must be 6. By independence of
2.1 Discrete Random Variables

the different rolls we get $P(X = 2) = \frac{5}{6} \times \frac{1}{6}$. More generally, in order to have $X = k$ the first $k - 1$ rolls cannot yield any 6 and the $k$th roll must be a 6. Thus,

$$P(X = k) = \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6} \text{ for all } k \geq 1.$$ 

Next, we graph this distribution

![Graph of geometric distribution](image)

Such a random variable is called geometric. More generally, we have the following.

**Geometric Random Variables**

Consider a sequence of independent identical trials. Assume that each trial can result in a success or a failure. Each trial has a probability $p$ of success and $q = 1 - p$ of failure. Let $X$ be the number of trials up to and including the first success. Then $X$ is called a geometric random variable. The distribution of $X$ is given by

$$P(X = k) = q^{k-1} p \text{ for all } k \geq 1.$$ 

Note that a geometric random variable may be arbitrarily large since the above probabilities are never 0. In order to check that the sum of these probabilities is 1 we need the following fact about geometric series:

**Geometric Series**

$$\sum_{k \geq 0} x^k = \frac{1}{1 - x} \text{ for all } x \in (-1, 1).$$
We have

\[
\sum_{k \geq 1} P(X = k) = \sum_{k \geq 1} q^{k-1} p = p \sum_{k \geq 0} q^k = \frac{p}{1-q} = 1.
\]

**Example 2.5.** Toss a fair coin until you get tails. What is the probability that exactly three tosses were necessary?

In this example we have \( p = q = 1/2 \). So

\[ P(X = 3) = q^2 p = \frac{1}{8}. \]

What is the probability that three or more tosses were necessary?

Note that the event “three or more tosses are necessary” is the same as the event “the first two tosses are heads.” Thus,

\[ P(X \geq 3) = q^2 = \frac{1}{4}. \]

**Example 2.6.** Consider \( X \) a geometric random variable. What is the probability that \( X \) is strictly larger than \( r \)?

The event “\( X > r \)” is the same as the event “the first \( r \) trials are failures.” Thus,

\[ P(X > r) = q^r. \]

**Example 2.7.** Let \( X \) be a geometric random variable. Given that \( X > r \) what is the probability that \( X > r + s \)?

We want

\[ P(X > r + s | X > r) = \frac{P(\{X > r + s\} \cap \{X > r\})}{P(X > r)}. \]

where the equality comes from the definition of a conditional probability. Note that the intersection \( \{X > r + s\} \cap \{X > r\} \) is simply \( \{X > r + s\} \). Thus,

\[ P(X > r + s | X > r) = \frac{P(X > r + s)}{P(X > r)}. \]

From Example 2.6, we know that \( P(X > r) = q^r \). So

\[ P(X > r + s | X > r) = \frac{q^{r+s}}{q^r} = q^s = P(X > s). \]

That is, given that we had \( r \) failures the probability of getting an additional \( s \) failures is the same as getting \( s \) failures to start with. In this sense, the geometric distribution is said to have the memoryless property.
Example 2.8. Two players roll a die. If the die shows 6 then \( A \) wins if the die shows 1 or 2 then \( B \) wins. The die is rolled until \( A \) or \( B \) wins. What is the probability that \( A \) wins?

Let \( T \) be the number of times the die is rolled. Note that the events \( \{ T = n \} \) are disjoint. We have

\[
P(A) = \sum_{n \geq 1} P(A \cap \{ T = n \}).
\]

The event “\( A \) wins in \( n \) rolls” is the same as the event “the first \( n - 1 \) rolls are draws and the \( n \)th roll is a 6.” Note the probability that a roll results in a draw is \( \frac{3}{6} \). Then

\[
P(A \cap \{ T = n \}) = \left( \frac{1}{2} \right)^{n-1} \times \frac{1}{6}.
\]

Summing the geometric series we get

\[
P(A) = \sum_{n \geq 1} \left( \frac{1}{2} \right)^{n-1} \times \frac{1}{6} = \frac{1}{3}.
\]

Note that the probability that \( A \) wins is

\[
P(A) = \frac{1}{3} = \frac{1/6}{1/6 + 2/6},
\]

where 1/6 is the probability of \( A \) winning in 1 roll and 2/6 is the probability of \( B \) winning in 1 roll.

Exercises 2.1

2.1. Toss three fair coins. Let \( X \) be the number of heads.

(a) Find the distribution of \( X \).

(b) Compute \( P(X \geq 2) \).

2.2. Roll two dice. Let \( X \) be the sum of the faces. Find the distribution of \( X \).

2.3. Recall that there are 38 pockets in a roulette and that 18 are red. I bet on red until I win. Let \( X \) be the number of bets I make.

(a) What is the probability that \( X \) is 2 or more?

(b) What is the probability that \( X \) is exactly 2?

2.4. I roll four dice. I win if I get at least one 6. What is the probability of winning?
2.5. Roll two fair dice. Let $X$ be the largest of the two faces. What is the distribution of $X$?

2.6. I draw two cards from a deck of 52. Let $X$ be the number of aces I draw. Find the distribution of $X$.

2.7. How many times should I toss a fair coin in order to get tails at least once with probability 90%?

2.8. In a lottery there are 100 tickets numbered from 1 to 100. Let $X$ be the ticket drawn at random. What is the distribution of $X$?

2.9. I roll a die until I get a 6. Given that the first two rolls were not 6’s, what is the probability I need 5 rolls or more in order to get a 6?

2.10. $A$ and $B$ roll a die. $A$ wins if the die shows a 6 and $B$ wins if the die shows a 1. The die is rolled until someone wins.

(a) What is the probability that $A$ wins?

(b) What is the probability that $B$ wins?

(c) Let $T$ be the number of times the die is rolled. Find the distribution of $T$.

2.11. Let $X$ be a discrete random variable.

(a) Show that

$$P(X = k) = P(X > k - 1) - P(X > k).$$

(b) Assume that for all $k \geq 1$ we have $P(X > k) = q^k$. Use (a) to show that $X$ is a geometric random variable.

2.2 Continuous Random Variables

We start with the following definition.

Continuous Random Variables

A continuous random variable is a function from a sample space $\Omega$ to an interval of the real numbers. The distribution of a continuous random variable $X$ is determined by its density function $f$ as follows. For all $a < b$ we have that

$$P(a < X < b) = \int_a^b f(x)dx.$$
The function \( f \) is positive, continuous (except possibly at finitely many points) and

\[
\int f(x) \, dx = 1,
\]

where the integral is taken on the largest interval on which \( f \) is strictly positive.

The shaded area below represents the probability that the random variable be between 2 and 4.

Note that for a continuous random variable \( X \) the following probabilities are all equal.

\[
P(a \leq X < b) = P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b).
\]

This is so because integrals of the type \( \int_a^b f(x) \, dx \) are always 0. In general, the above equalities do not hold for discrete random variables.
Example 2.1. Let $X$ have density $f(x) = cx^2$ for $x$ in $[-1,1]$ and $f(x) = 0$ elsewhere. Find $c$.

We must have

$$\int_{-1}^{1} cx^2 \, dx = 1.$$  

After integrating we get

$$c \left( \frac{2}{3} \right) = 1$$

and therefore $c = 3/2$.

What is the probability that $X$ is larger than 1/2?

$$P(X > 1/2) = \int_{1/2}^{1} f(x) \, dx = \int_{1/2}^{1} \left( \frac{3}{2} \right) x^2 \, dx = \frac{7}{16}.$$  

We next give two examples of important continuous random variables.

2.2.1 Continuous Uniform Random Variables

Example 2.2. Let $X$ be a random variable with density $f(x) = 1$ for $x$ in $[0,1]$ and $f(x) = 0$ elsewhere. Since the density of $X$ is flat on $[0,1]$, $X$ is said to be uniform on $[0,1]$. Next we graph the density of $X$.

Note that

$$\int_{0}^{1} f(x) \, dx = \int_{0}^{1} 1 \, dx = 1.$$  

What is the probability of $X$ to be in the interval $(1/2, 3/4)$?

We have that

$$P(1/2 < X < 3/4) = \int_{1/2}^{3/4} f(x) \, dx = \frac{1}{4}.$$
2.2 Continuous Random Variables

What is the probability that $X$ is larger than $1/2$?

$$P(X > 1/2) = \int_{1/2}^{1} f(x)dx = \frac{1}{2}.$$ 

More generally, we have the following.

**Continuous Uniform Random Variables**

A continuous random variable $X$ is uniform on the interval $[a, b]$ if the density of $X$ is

$$f(x) = \frac{1}{b-a} \text{ for } x \in [a, b].$$

Note that the density of a uniform is always a constant on some interval and the constant must be picked so that the area under the density is 1.

2.2.2 *Exponential Random Variables*

*Example 2.3.* Let $T$ be a random variable with density $f(t) = e^{-t}$ for $t \geq 0$. Below is the graph of $f$. 

![Graph of exponential density function](image-url)
We first check that the area under the curve is 1.
\[ \int_{0}^{A} e^{-t} dt = 1 - e^{-A}. \]

By letting \( A \) go to infinity we get that the improper integral converges and
\[ \int_{0}^{\infty} e^{-t} dt = 1. \]

What is the probability that \( T \) is larger than 1?
\[ P(T > 1) = \int_{1}^{\infty} e^{-t} dt = e^{-1}. \]

What is the probability that \( T \) is less than 1?
\[ P(T \leq 1) = 1 - P(T > 1) = 1 - e^{-1}. \]

We next state the definition of an exponential random variable.

**Exponential Random Variables**

A random variable \( X \) with density \( f(x) = ae^{-ax} \) for \( x \geq 0 \) is said to be an exponential random variable with parameter (or rate) \( a > 0 \).

**Example 2.4.** Let \( T \) be an exponential random variable with parameter \( a \). What is the probability that \( T \) is larger than \( s \)?
\[ P(T > s) = \int_{s}^{\infty} ae^{-at} dt = e^{-as}. \]

**Example 2.5.** Let \( T \) be an exponential random variable with parameter \( a \). Given that \( T \) is larger than \( s \), what is the probability that \( T \) is larger than \( t + s \)?

We want the conditional probability
\[ P(T > t + s | T > s) = \frac{P(T > t + s \cap T > s)}{P(T > s)}. \]

Note that the intersection of the events \( T > t + s \) and \( T > s \) is the event \( T > t + s \).

Thus,
\[ P(T > t + s | T > s) = \frac{P(T > t + s)}{P(T > s)}. \]
Exercises 2.2

By using the computation in Example 4, we get

\[
P(T > t + s | T > s) = \frac{e^{-\alpha(t+s)}}{e^{-\alpha s}} = e^{-\alpha t} = P(T > t).
\]

So exactly as for the geometric distribution of the preceding section we say that the exponential distribution has the memoryless property.

Exercises 2.2

2.1. Let \( f(x) = cx(1-x) \) for \( x \) in \([0,1]\) and \( f(x) = 0 \) elsewhere. Find \( c \) so that \( f \) is a density function.

2.2. Let the graph of the density \( f \) be a triangle for \( x \) in \([-1,1]\). Find \( f \).

2.3. Let \( X \) be the density of an uniform random variable on \([-2,4]\). Find the density of \( X \).

2.4. Let \( T \) be the waiting time for a bus. Assume that \( T \) has an exponential density with rate 3 per hour.
   (a) What is the probability of waiting at least 20 min for the bus?
   (b) Given that we have waited 20 min, what is the probability of waiting an additional 20 min for the bus?
   (c) Under which conditions is the exponential model appropriate for this problem?

2.5. Let \( T \) be a waiting time for a bus. Assume that \( T \) has a uniform distribution on \([0,40]\).
   (a) What is the probability of waiting at least 20 min for the bus?
   (b) Given that we have waited 20 min, what is the probability of waiting an additional 10 min for the bus?

2.6. Let \( Y \) have a density \( g(y) = cye^{-2y} \) for \( y \geq 0 \). Find \( c \).

2.7. Let \( X \) have density \( f(x) = xe^{-x} \) for \( x \geq 0 \). What is the probability that \( X \) is larger than 3?

2.8. Let \( T \) have density \( g(t) = 4t^3 \) for \( t \) in \([0,1]\).
   (a) What is the probability that \( T \) is between 1/4 and 3/4?
   (b) What is the probability that \( T \) is larger than 1/2?

2.9. (a) Show that for any random variable \( X \) we have

\[
P(a < X < b) = P(X < b) - P(X \leq a).
\]
   (b) Assume that the random variable \( X \) is continuous and is such that \( P(X > s) = e^{-2s} \). Use (a) to compute \( P(a < X < b) \).
   (c) Find the density of \( X \).
2.3 Expectation

As we have seen in the preceding two sections knowing the distribution of a random variable entails knowing a lot of information. For a discrete random variable \( X \) the distribution is given by the sequence \( P(X = k) \) for every \( k \) in the range of \( X \). For a continuous random variable \( X \) the distribution is given by the density function \( f \).

For many problems it is enough to have a rough idea of the distribution and one tries to summarize the distribution by using a few numbers. The most important of these numbers is the expectation or the average value of the distribution. We first deal with discrete random variables.

**Expectation of a Discrete Random Variable**

The expectation (or mean) of the discrete random variable \( X \) is denoted by \( E(X) \) and is given by

\[
E(X) = \sum_{k} k P(X = k),
\]

where the sum is taken over all the values in the range of \( X \).

If a random variable may take infinitely many values then the computation of its expectation involves an infinite series. The expectation is defined only if the infinite series converges (see Exercise 18).

Note that the expectation of \( X \) is a measure of location of \( X \).

**Example 2.1.** We perform an experiment with two possible outcomes: failure or success. If we have a success we set \( X = 1 \). If we have a failure we set \( X = 0 \). Let \( P(X = 1) = p \). What is the expectation of this Bernoulli random variable?

\[
E(X) = \sum_{k} k P(X = k) = 0 \times (1 - p) + 1 \times p = p.
\]

Thus we can state:

**Expectation of a Bernoulli Random Variable**

Let \( X \) be a Bernoulli random variable with probability of success \( p \). That is, \( X \) may take only values 0 and 1 and \( P(X = 1) = p \). Then,

\[
E(X) = p.
\]
2.3 Expectation

For instance, if we toss a fair coin and set $X = 1$ if we have heads and $X = 0$ if we get tails then $E(X) = 1/2$. What is the meaning of the value $1/2$ since $X$ can only take values 0 and 1?

The Law of Large Numbers that we will now (loosely) describe gives a physical meaning to the notion of expectation.

**Law of Large Numbers**
We make $n$ independent and identical random experiments. Each experiment has a random outcome with the same distribution as the random variable $X$. The Law of Large Numbers states that as $n$ goes to infinity the average over the $n$ outcomes approaches $E(X)$.

We now come back to Example 2.1. The Law of Large Numbers states that if we toss a coin many times then the ratio of heads over the total number of tosses will approach $1/2$. This gives a physical meaning to the expected value and also explains why this is a crucial notion.

**Example 2.2.** Roll a fair die. Let $X$ be the face shown. We have $P(X = k) = 1/6$ for every $k = 1, 2, \ldots, 6$. Thus,

$$E(X) = \sum_{k=1}^{6} k P(X = k) = \sum_{k=1}^{6} \frac{k}{6} = \frac{7}{2}.$$ 

**Example 2.3.** The preceding example gave the expected value of a discrete uniform random variable in a particular case. We now treat the general case. Assume that $X$ is a discrete uniform random variable on the set $\{1, 2, \ldots, n\}$. Thus, $P(X = k) = 1/n$ for $k = 1, 2, \ldots, n$. So

$$E(X) = \sum_{k=1}^{n} k P(X = k) = \frac{1}{n} \sum_{k=1}^{n} k.$$ 

Thus, we need to compute the sum of the first $n$ integers. Let $S_n$ be this sum and we write $S_n$ in two different ways.

$$S_n = 1 + 2 + \cdots + (n - 1) + n,$$

$$S_n = n + (n - 1) + \cdots + 2 + 1.$$ 

We now add both equations to get

$$2S_n = (n + 1) + (n + 1) + \cdots + (n + 1).$$
There are \( n \) terms equal to \( n + 1 \) on the r.h.s. Thus, 

\[
2S_n = n(n + 1)
\]

and we get 

\[
S_n = \frac{n(n + 1)}{2}.
\]

Going back to the computation of the expected value we have: 

\[
E(X) = \frac{n + 1}{2}.
\]

Note that if we let \( n = 6 \) we get the particular case of Example 1.

**Example 2.4.** We now deal with geometric random variables. Let \( X \) be the number of independent and identical trials to get the first success. We denote by \( p \) the probability that a given trial be a success and \( q = 1 - p \). The distribution of \( X \) is given by 

\[
P(X = k) = q^{k-1} p \quad \text{for all } k = 1, 2, \ldots.
\]

Thus, 

\[
E(X) = \sum_{k=1}^{\infty} kq^{k-1} p = p \sum_{k=1}^{\infty} kq^{k-1}.
\]

Recall that 

\[
\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x} \quad \text{for } x \in (-1, 1).
\]

Also recall that power series are infinitely differentiable on their interval of convergence (except possibly at the boundary points). Thus, by taking derivatives on both sides of the preceding equality we get: 

\[
\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1 - x)^2} \quad \text{for } x \in (-1, 1).
\]

We plug \( x = q \) and get for the expected value 

\[
E(X) = p \sum_{k=1}^{\infty} kq^{k-1} = p \frac{1}{(1 - q)^2} = \frac{1}{p}.
\]
2.3 Expectation

**Expectation of a Geometric Random Variable**

Let $X$ be the number of independent and identical trials up to and including the first success. We denote by $p$ the probability that a given trial be a success and $q = 1 - p$. Then,

$$E(X) = \frac{1}{p}.$$ 

**Example 2.5.** Roll a die until you get a 6. What is the expected number of rolls?

Let $T$ be the number rolls to get a 6. This is a geometric random variable with $p = 1/6$. Thus, $E(T) = 6$.

**2.3.1 Continuous Random Variables**

We start by defining the expected value for a continuous random variable.

**Expectation of a Continuous Random Variable**

Assume that $X$ is a continuous random variable with density $f$. The expected value (or mean) of $X$ is then

$$\int xf(x)\,dx,$$

where the integral is taken on the largest interval on which $f$ is strictly positive.

**Example 2.6.** Assume that $X$ in uniformly distributed on $[a, b]$. What is its expected value?

Using that the density of $X$ is $f(x) = \frac{1}{b-a}$ for $x$ in $[a, b]$, we get

$$E(X) = \int_a^b xf(x)\,dx = \frac{1}{b-a} \left( \frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$ 

We get

**Expectation of a Continuous Uniform Random Variable**

Assume that $X$ in uniformly distributed on $[a, b]$. Then

$$E(X) = \frac{a + b}{2}.$$
2 Random Variables

Example 2.7. Assume $T$ is exponentially distributed with rate $a$. What is its expected value?

We integrate by parts to get

$$E(T) = \int_0^\infty tf(t)\,dt = \int_0^\infty ta e^{-at}\,dt = -te^{-at}]_0^\infty + \int_0^\infty e^{-at}\,dt = \frac{1}{a}. \quad \text{(AQ1)}$$

Expectation of an Exponential Random Variable

Assume $T$ is exponentially distributed with rate $a$. Then,

$$E(T) = \frac{1}{a}.$$

2.3.2 Other Measures of Location

To summarize the location of a distribution it is often a good idea to use more than one number. Besides the mean, there are two other important measures of location. The first one is the median.

Median of a Random Variable

A median $m$ of a random variable $X$ is a number $m$ such that $P(X \leq m)$ and $P(X \geq m)$ are both at least 1/2.

As we will show in the exercises a median gives less weight to the extreme values of the distribution than the mean.

Example 2.8. Roll a die. Let $X$ be the face shown. Note that $P(X \geq 3) = 2/3$ and $P(X \leq 3) = 1/2$. So 3 is a median. Observe that 4 is also a median and actually any number in [3, 4] is a median. Recall that the mean in this case is 3.5. This example shows that a discrete random variable may have several medians.

Unlike what may happen for discrete random variables there is a unique median for continuous random variables. If the continuous variable $X$ has density $m$ then the median of $X$ is such that

$$\int_m^\infty f(x)\,dx = \int_{-\infty}^m f(x)\,dx = \frac{1}{2}.$$
2.3 Expectation

Example 2.9. Let $T$ be an exponential random variable with rate 1. What is its median?

We solve the equation as mentioned below

$$P(T > m) = P(T \geq m) = \int_m^\infty e^{-t} dt = e^{-m} = \frac{1}{2}. $$

Thus $m = \ln 2$. Note that $P(T < \ln 2) = 1 - P(T > \ln 2) = 1/2$. So $\ln 2$ is the unique median of this distribution.

Another measure of location, only defined for discrete random variables, is the mode.

Mode of a Discrete Random Variable

A mode $M$ of a discrete random variable $X$ is a number $M$ such that $P(X = M)$ is maximum.

Example 2.10. For the uniform distribution on $\{1, 2, \ldots, 6\}$ there are 6 modes: $M = 1, 2, 3, 4, 5, 6$.

2.3.3 The Addition Rule

The following rule holds for any type (continuous or discrete) of random variables.

Addition Rule

Let $X$ and $Y$ be two random variables defined on the same sample space $\Omega$. Then,

$$E(X + Y) = E(X) + E(Y).$$

More generally, if $X_1, X_2, \ldots, X_n$ are all defined on $\Omega$ we have

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

As the next example shows this is a very important rule. Its proof involves the joint distribution of several random variables. We will prove this formula when we will see joint distributions in 8.3.
Example 2.11. I roll two dice. Let $S$ the sum of the two dice. What is the expected value of $S$?

Let $X$ be the value of the first die and $Y$ be the value of the second die. Then $S = X + Y$. According to the addition rule we have

$$E(S) = E(X) + E(Y).$$

But Example 2.1 tells us that $E(X) = E(Y) = 7/2$. Thus,

$$E(S) = 7.$$

We could have computed $E(S)$ by first computing the distribution of $S$ and then averaging but this would have taken a lot longer.

### 2.3.4 Computing the Expectation By Breaking Up the Random Variable

In many cases the distribution of a given random variable is too involved to be computed. In some of those cases it is possible to break up a random variable into a sum of Bernoulli random variables. By using the addition rule we then get the mean of the random variable with the involved distribution without computing its distribution. We next give such an example.

Example 2.12. Assume that three people enter independently an elevator that goes to five floors. What is the expected number of stops $S$ that the elevator is going to make?

Instead of computing the distribution of $S$ we break $S$ into a sum of 5 Bernoulli random variables as follows. Let $X_1 = 1$ if at least one person goes to floor 1, otherwise we set $X_1 = 0$. Likewise let $X_2 = 1$ if at least one person goes to floor 2, otherwise we set $X_2 = 0$. We do the same for the five possible choices. We have

$$S = X_1 + X_2 + \cdots + X_5.$$

Note that $X_1 = 0$ if none of the three people pick floor 1. Thus,

$$P(X_1 = 0) = \left(\frac{4}{5}\right)^3.$$
2.3 Expectation

The probability of success for $X_1$ is $p = P(X_1 = 1) = 1 - (\frac{4}{5})^3$. All the $X_i$ have the same Bernoulli distribution. By the addition rule we have

$$E(S) = 5p = 5 \left( 1 - \left( \frac{4}{5} \right)^3 \right) = \frac{61}{25} = 2.44.$$ 

We now compute $E(S)$ by using the distribution of $S$. The random variable $S$ may only take values 1, 2 and 3. In order to have $S = 1$, the second and third person need to pick the same floor as the first person. Thus,

$$P(S = 1) = \left( \frac{1}{5} \right)^2.$$ 

To have $S = 2$, there are two possibilities: either the second person picks the same floor as the first one and the third a different floor (the probability of that is $(1/5)(4/5)$) or the second person picks a different floor from the first one and the third one picks one of the two floors that have already been picked (the probability of that is $(4/5)(2/5)$). Thus,

$$P(S = 2) = \left( \frac{1}{5} \right) \left( \frac{4}{5} \right) + \left( \frac{4}{5} \right) \left( \frac{2}{5} \right).$$ 

Finally, $S = 3$ happens only if the three persons pick distinct floors:

$$P(S = 3) = \left( \frac{4}{5} \right) \left( \frac{3}{5} \right).$$ 

Thus,

$$E(S) = 1 \times \frac{1}{25} + 2 \times \frac{12}{25} + 3 \times \frac{12}{25} = \frac{61}{25}.$$ 

So even in this very simple case ($S$ has only three values after all) it is better to compute the expected value of $S$ by breaking $S$ in a sum of 0–1 random variables rather than compute the distribution of $S$.

Example 2.13. Let $B$ be the number of distinct birthdays in a class of 50 students. What is the $E(B)$?

The distribution of $B$ is clearly fairly involved. We are going to break $B$ into a sum of Bernoulli random variables. Set $X_1 = 1$ if at least one student was born on
January 1, otherwise set $X_1 = 0$. Set $X_2 = 1$ if at least one student was born on
January 2, otherwise set $X_2 = 0$. We define $X_i$ like above for every one for 365
days of the calendar. We claim that

$$B = X_1 + X_2 + \cdots + X_{365}.$$  

This is so because the r.h.s. counts all the days on which at least one student has his
birthday. Moreover, the $X_i$ are Bernoulli random variables. In order for $X_1 = 0$ we
must have that none of the 50 students was born on January 1. Thus,

$$P(X_1 = 0) = \left( \frac{364}{365} \right)^{50}.$$  

The probability of success for $X_1$ is $p = 1 - \left( \frac{364}{365} \right)^{50}$. We do the same for every
$X_i$ and they all have the same $p$ (which is also the expected value of a Bernoulli
random variable). By the addition rule we have

$$E(B) = E(X_1) + E(X_2) + \cdots + E(X_{365}) = 365p = 365 \left( 1 - \left( \frac{364}{365} \right)^{50} \right).$$

Numerically, we get

$$E(B) = 46.79.$$  

From Example 2.2 in Sect. 1.4 we know that the probability of having at least
two students born on the same day is 0.96. However from the value of $E(B)$
we see that more than two students born on the same day or more than one set
of students born on the same day are not that likely, otherwise $E(B)$ would be
lower.

Example 2.14. The collector’s problem. Assume that a certain brand of cereals has
a cartoon character in each box. There are $r$ different cartoon characters. What is
the expected number of cereal boxes that need to be purchased in order to get all the
cartoon characters?

Let $T_1$ be the number of boxes needed to get the first character. Obviously, $T_1 = 1$. Let $T_2$ be the number of boxes needed to get the second (different) character.

Since we have already one character every time we buy a box there is a probability
$\frac{1}{r}$ of getting the same character we already have and a probability $\frac{r-1}{r}$ to get a
different one. Hence, $T_2$ is a geometric random variable with success probability
$\frac{r-1}{r}$. More generally, let $T_k$ be the number of boxes needed to get the $k$th different
character given that we have already $k - 1$ different characters. Since we have
already $k - 1$ characters every time we purchase a box the probability to get a
2.3 Expectation

The $k$th different character is $\frac{r-(k-1)}{r}$. That is, $T_k$ is a geometric random variable with probability of success $\frac{r-k+1}{r}$ for $k = 2, \ldots, r$. The number of boxes needed to have a complete collection is therefore

$$T_1 + T_2 + \cdots + T_r.$$ 

Recall that the expected value of a geometric random variable with success probability $p$ is $1/p$. Hence, the expected number of boxes needed to have the complete collection is:

$$E(T_1 + T_2 + \cdots + T_r) = 1 + \frac{r}{r-1} + \frac{r}{r-2} + \cdots + \frac{r}{2} + \frac{r}{1}.$$ 

It is convenient to rewrite the formula as

$$E(T_1 + T_2 + \cdots + T_r) = r \left( 1 + \frac{1}{2} + \cdots + \frac{1}{r} \right).$$

As $r$ goes to infinity one can show that

$$1 + \frac{1}{2} + \cdots + \frac{1}{r} \sim \ln r$$

in the sense that the ratio goes to 1. Hence, the expected number of boxes needed to complete the collection is approximately $r \ln r$.

2.3.5 Fair Gambling

Example 2.15. We roll a die. You pay me $b$ if the die shows 5 or 6. I pay you $1 otherwise. Clearly, the probabilities of winning are not the same for both players. Can we pick $b$ so that this is a fair game?

Assume we play this game many times. By the Law of Large Numbers my average winnings will be close to my expected winnings. We will say that the game is fair if the expected winnings (of each player) are 0. So that in the long run my average winnings will approach 0.

In this particular case let $W$ be my winnings in 1 bet. We have that $W = b$ with probability 1/3 and $W = -1$ with probability 2/3. Thus

$$E(W) = b \times \frac{1}{3} + (-1) \times \frac{2}{3}.$$ 

We want $E(W) = 0$. Solving for $b$ we get $b = 2$. Since I am twice less likely to win than you are, you should pay me twice as much when I win.
2.3.6 Expectation of a Function of a Random Variable

As we will see in the next section it is often necessary to compute $E(X^2)$ for a random variable $X$. This is NOT $E(X)^2$. We could compute the distribution of $X^2$ and use the distribution to compute the expected value. However, there is a quicker way to do things and it is contained in the following formula.

**Expectation of a Function of $X$**

Let $X$ be a random variable and $g$ be a real valued function. For instance, $g(x) = x^2$. Then if $X$ is discrete we have

$$E(g(X)) = \sum_k g(k)P(X = k).$$

If $X$ is continuous with density $f$ then

$$E(g(X)) = \int g(x)f(x)dx.$$

**Example 2.16.** Let $X$ be a discrete random variable such that $P(X = -1) = 1/3$, $P(X = 0) = 1/2$ and $P(X = 2) = 1/6$. What is $E(X^2)$?

We use the formula above with $g(x) = x^2$ to get

$$E(X^2) = (-1)^2 \times \frac{1}{3} + 0^2 \times \frac{1}{2} + (2)^2 \times \frac{1}{6} = 1.$$

**Example 2.17.** Let $X$ be uniformly distributed on $[0,1]$. What is $E(X^3)$?

This time we use the formula with $g(x) = x^3$. We get

$$E(X^3) = \int_0^1 x^3 f(x)dx = \frac{1}{4}.$$

Another case which is of particular interest is when $g(x) = ax + b$. Assume that $X$ is a discrete random variable. Then we use the formula above to get

$$E(aX + b) = \sum_k (ak + b)P(X = k) = a \sum_k kP(X = k) + b \sum_k P(X = k) = aE(X) + b.$$

The same formula may be derived for continuous random variables. We have the following for continuous and discrete random variables.
2.3 Expectation

Expectation of a Linear Function of X

\[ E(aX + b) = aE(X) + b. \]

The following observation gives the expectation without any computation provided. We have a symmetric distribution.

Symmetric Case

Let \( f \) be the density of a continuous random variable \( X \). Assume that there is \( a \geq 0 \) such that

\[ f(a + x) = f(a - x) \]

for every \( x \). Then, \( E(X) = a \) (if \( E(X) \) exists!).

We now show this property. Assume first that \( a = 0 \). That is,

\[ f(x) = f(-x) \]

for every \( x \). We have

\[ E(X) = \int_{-b}^{+b} xf(x)dx, \]

where \( b \) is a positive number or \( +\infty \). Let \( g(x) = xf(x) \), note that \( g \) is an odd function. That is,

\[ g(-x) = -g(x) \]

for every \( x \). Hence, \( E(X) \) is the integral of an odd function on a symmetric interval. It is easy to see that this integral must be 0 and therefore \( E(X) = 0 \). We are done in the case \( a = 0 \).

If \( a > 0 \) let \( Y = X - a \). One can check (this will be done in Sect. 8.1) that the density of \( Y \) is \( f_Y(x) = f(a + x) \). This implies that

\[ f_Y(x) = f(a + x) = f(a - x) = f_Y(-x). \]

That is, \( Y \) has a symmetry at 0 and therefore by the case \( a = 0 \) we know that \( E(Y) = 0 \). But

\[ E(Y) = E(X) - a = 0. \]

Hence, \( E(X) = a \) and we are done.
Exercises 2.3

2.1. What is the expected value of a random variable uniformly distributed on \{-1, 0, 3\}.

2.2. Toss two fair coins. What is the expected number of heads?

2.3. The probability of finding someone in favor of a certain initiative is 0.01. We interview people at random until we find a person in favor of the initiative. What is the expected number of interviews?

2.4. Roll two dice. What is the expected value of the maximum of the two dice?

2.5. Let $X$ be exponentially distributed with mean 1/2. What is the density of $X$?

2.6. Let $U$ be a random variable which is uniformly distributed on $[-1, 2]$.
   (a) Compute the mean of $U$.
   (b) What is the median of $U$?

2.7. Let $X$ have the following density.

\[
\begin{array}{c|c|c|c|c}
 & 0 & 1 & 2 \\
\hline
0 & , & , & , \\
\hline
1 & , & , & , \\
\hline
2 & , & , & , \\
\hline
\end{array}
\]

   (a) Find the expected value of $X$.
   (b) How good is $E(X)$ as a measure of location of $X$?

2.8. Let $f(x) = 3x^2$ for $x$ in $[0, 1]$. Let $X$ be a random variable with density $f$.
   (a) What is $E(X)$?
   (b) What is the median of $X$?

2.9. Let $X$ be a random variable such that $P(X = 0) = 1/5$ and $P(X = 4) = 4/5$. Find the mean, medians and modes.

2.10. Let $T$ be exponentially distributed with rate $a$. Find the median of $T$ in function of $a$.

2.11. Roll four dice. What is the expected value of the sum?

2.12. There are three components in a circuit. Each one of them fails with probability $p$. The failure of one component may influence the other components in a way that is not well understood. What is the expected number of working components?
2.13. Let $B$ be the number of distinct birthdays in a class of 200 students. What is the $E(B)$?

2.14. There are eight people in a bus and five bus stops ahead. What is the expected number of stops the bus will have to make for these eight people?

2.15. I roll four dice. If there is at least one 6 you pay me $1. If there are no 6's I pay you $1.

(a) Is this a fair game?
(b) How would you make it into a fair game?

2.16. Let $X$ be uniform on $\{1, 2, \ldots, 6\}$. What is $E(X^2)$?

2.17. Let $X$ be exponentially distributed with rate 1. What is $E(X^2)$?

2.18. In this problem we give an example of a discrete random variable for which the expectation is not defined.

(a) Use the fact that
\[ \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} \]

...to find $c$ so that $P(X = k) = c/k^2$ is a probability distribution.

(b) Show that the expectation of the random variable defined above does not exist.

2.19. This problem gives an example of a continuous random variable that has no expectation.

(a) Show that $f(x) = \frac{2}{\pi(1+x^2)}$ for $x > 0$ is a density function.
(b) Show that a random variable with the density above has no expectation.

2.20. I roll a die repeatedly.

(a) What is the expected number of rolls to get three different faces?
(b) What is the expected number of rolls to get all six faces?

2.4 Variance

We have seen in Sect. 2.3 that the expectation is a measure of location for a distribution. Next we are going to define a measure of dispersion: the variance. A small variance will mean that the distribution is concentrated around the mean and that the mean is a good measure of location. A large variance will mean that the distribution is dispersed and that no value is really typical for this distribution.
2 Random Variables

Variance of a Random Variable
Let $X$ be random variable with mean $E(X) = \mu$. The variance of $X$ is denoted by $\text{Var}(X)$ and is defined by

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

The following formula for the variance is useful for computational purposes

$$\text{Var}(X) = E(X^2) - \mu^2.$$

Finally, the standard deviation of $X$ is denoted by $SD(X)$ and is defined by

$$SD(X) = \sqrt{\text{Var}(X)}.$$

We now list the consequences of these definitions.

**Consequences**

C1. The variance of a random variable is ALWAYS positive or 0. This is so because the variance is the expected value of the positive random variable $(X - \mu)^2$.

C2. The variance of a random variable $X$ is 0 if and only if $X$ is a constant. For a discrete random variable this can be seen from the formula

$$E[(X - \mu)^2] = \sum_k (k - \mu)^2 P(X = k).$$

If this sum is 0 it means that every term must be 0 since these are all positive terms. But the sum of the $P(X = k)$ is 1 so at least some of these terms are non zero. It is easy to see that for exactly one $k$ $P(X = k)$ is not 0 and that corresponds to $k = \mu$. Thus, $X$ is a constant equal to $\mu$.

For a continuous random variable (that is a random variable whose density is strictly positive on some interval) one can show that the variance is always strictly positive.

C3. An easy consequence of the definition of variance is that

**Properties of Variance**

For any random variable $X$ and constants $a$ and $b$ we have that

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

$$SD(aX + b) = |a|SD(X).$$
2.4 Variance

Observe that the translation by $b$ has no effect on the variance of $aX + b$. Intuitively, this is clear since the variance measures the dispersion, not the location, of a random variable.

**Example 2.1.** We start with the Bernoulli distribution. Assume that $X$ takes values 0 and 1. We denote $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$. What is the variance of $X$?

We have that

$$E(X) = p.$$ 

We now compute

$$E(X^2) = 0^2 	imes q + 1^2 	imes p = p.$$ 

Thus,

$$\text{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = pq.$$ 

**Variance of a Bernoulli Random Variable**

Assume that $X$ takes values 0 and 1. We denote $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$. Then,

$$\text{Var}(X) = pq.$$ 

**Example 2.2.** What is the variance of the discrete random variable uniformly distributed on $\{1, 2, 3, 4, 5, 6\}$?

We know that $E(X) = 7/2$.

We now compute

$$E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6} = \frac{91}{6}.$$ 

Thus,

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$ 

So the standard deviation is approximately 1.7. It is large for a distribution on $\{1, \ldots, 6\}$. But this is not surprising since the extreme values have the same weight as the middle values for this distribution.
Example 2.3. We now turn to the variance of a geometric random variable. We have independent identical trials that have a probability $p$ of success. Let $T$ be the number of trials to get the first success. The random variable $T$ has a geometric distribution and we know that

$$E(T) = \frac{1}{p}.$$ 

As before we need to compute $E(T^2)$. In this case it is easier to compute $E(T(T-1))$ first. We need a new fact about geometric series. Recall that for every $x$ in $(-1, 1)$ we have

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$ 

Power series are infinitely differentiable on their interval of convergence. We take derivatives twice in the formula above to get:

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}.$$ (2.1)

Now we compute

$$E(T(T-1)) = \sum_{k=1}^{\infty} k(k-1)P(T = k)$$

$$= \sum_{k=2}^{\infty} k(k-1)q^{k-1}p = pq \sum_{k=2}^{\infty} k(k-1)q^{k-2}.$$ 

We let $x = q$ in (2.1) to get

$$E(T(T-1)) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}.$$ 

We have that

$$E(T^2) = E(T(T-1)) + E(T) = \frac{2q}{p^2} + \frac{1}{p}.$$ 

Finally,

$$\text{Var}(T) = E(T^2) - E(T)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q + p - 1}{p^2}.$$ 
2.4 Variance

Note that $p + q = 1$, so $2q + p - 1 = q$. Hence,

$$\text{Var}(T) = \frac{q}{p^2}.$$

### Variance of a Geometric Random Variable

Assume that we have independent identical trials that have a probability $p$ of success. Let $T$ be the number of trials to get the first success. Then,

$$\text{Var}(T) = \frac{q}{p^2}.$$

We now compute variances for a few continuous random variables.

**Example 2.4.** Assume that $X$ is uniformly distributed on $[a, b]$. Then

$$E(X) = \frac{a + b}{2}.$$

We compute $E(X^2)$.

$$E(X^2) = \int_a^b x^2 f(x) \, dx = \frac{1}{b - a} \int_a^b x^2 \, dx = \frac{1}{3(b - a)}(b^3 - a^3) = \frac{b^2 + ab + a^2}{3}.$$

Thus,

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{b^2 - 2ab + a^2}{12} = \frac{(b - a)^2}{12}.$$

### Variance of a Continuous Uniform Random Variable

Assume that $X$ is uniformly distributed on $[a, b]$. Then

$$\text{Var}(X) = \frac{(b - a)^2}{12}.$$

We now deal with an exponential random variable.

**Example 2.5.** Assume that $T$ is exponentially distributed with rate $a$. Then, $E(T) = 1/a$. We have

$$E(T^2) = \int_0^\infty t^2 f(t) \, dt = \int_0^\infty t^2 ae^{-at} \, dt.$$

We do an integration by parts to get

\[ E(T^2) = \int_0^\infty t^2 e^{-at} dt = \frac{2}{a} \int_0^\infty t e^{-at} dt = \frac{2}{a^2}, \]

where we have used \( E(T) = 1/a \) to get the last equality. So

\[
\text{Var}(T) = E(T^2) - (E(T))^2 = \frac{2}{a^2} - \frac{1}{a^2} = \frac{1}{a^2}.
\]

That is, the mean and the standard deviation are equal for an exponential distribution. This shows that exponential distributions are rather dispersed.

**Variance of an Exponential Random Variable**

Assume that \( T \) is exponentially distributed with rate \( a \). Then,

\[
\text{Var}(T) = \frac{1}{a^2}.
\]

**Example 2.6.** Consider \( Y \) with the following density.

The density of \( Y \) is \( f(y) = y \) for \( y \) in \([0, 1]\) and \( f(y) = 2 - y \) for \( y \) in \([1, 2]\). The mean of \( Y \) is 1 because of the symmetry of the density. We confirm this by computation.

\[
E(Y) = \int_0^2 y f(y) dy = \int_0^1 y^2 dy + \int_1^2 y(2-y) dy.
\]

Thus,

\[
E(Y) = \left. y^3 / 3 \right|_0^1 + \left. y^2 \right|_1^2 - \left. y^3 / 3 \right|_1^2 = 1.
\]
2.4 Variance

We now deal with $E(Y^2)$.

$$E(Y^2) = \int_0^2 y^2 f(y)dy = \int_0^1 y^3dy + \int_1^2 y^2(2-y)dy.$$  

So

$$E(Y^2) = \frac{y^4}{4}\bigg|_0^1 + \frac{2y^3}{3}\bigg|_1^2 - \frac{y^4}{4}\bigg|_1^2 = \frac{7}{6}.  

 Var(Y) = E(Y^2) - E(Y)^2 = \frac{7}{6} - 1 = \frac{1}{6}.  

2.4.1 Independent Random Variables

We will need to compute the variance of sums of random variables. This turns out to be a simple task only when the random variables in the sum are independent. We start by defining independence for random variables.

**Independent Random Variables**

Two discrete random variables $X$ and $Y$ are said to be independent if

$$P(\{X = x\} \cap \{Y = y\}) = P(X = x)P(Y = y) \text{ for ALL } x, y.$$  

Two continuous random variables $X$ and $Y$ are said to be independent if for ALL real numbers $a < b, c < d$ we have

$$P(\{a < X < b\} \cap \{c < Y < d\}) = P(a < X < b)P(c < Y < d).$$  

We now examine two examples.

**Example 2.7.** Roll two dice. Let $X$ be the face shown by the first die and $S$ be the sum of the two dice. Are $X$ and $S$ independent?

Intuitively it is clear that the answer should be no. It is enough to find one $x$ and one $y$ such that

$$P(\{X = x\} \cap \{S = y\}) \neq P(X = x)P(S = y)$$

in order to show that $X$ and $S$ are not independent. For instance, take $x = 1$ and $y = 12$. Clearly if one die shows 1 the sum cannot be 12. So $P(\{X = 1\} \cap \{S = 12\}) = 0$. However, $P(X = 1)$ and $P(S = 12)$ are strictly positive so $P(\{X = 1\} \cap \{S = 12\}) \neq P(X = 1)P(S = 12)$. $X$ and $S$ are not independent.
Example 2.8. Toss two fair coins. Set $X = 1$ if the first coin shows heads, set $X = 0$ otherwise. Set $Y = 1$ if the second coin shows heads, set $Y = 0$ otherwise. Are $X$ and $Y$ independent?

Our sample space is $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. We need to examine the 4 possible outcomes for $(X, Y)$. Note that the event $\{X = 0\} \cap \{Y = 0\}$ is the event $\{(T, T)\}$ and that has probability 1/4. Note that $P(X = 0) = 2/4 = P(Y = 0)$. So the product rule holds for $x = 0$ and $y = 0$. We now examine $x = 0$ and $y = 1$. The event $\{X = 0\} \cap \{Y = 1\}$ is the event $\{(T, H)\}$. This has probability 1/4. Since $P(Y = 1) = 2/4$ the product rule holds in this case as well. The two remaining cases are symmetric to the cases we just examined. We may conclude that $X$ and $Y$ are independent.

2.4.2 Variance of a Sum of Random Variables

If $X$ and $Y$ are independent it is easy to compute the variance of $X + Y$.

Variance of a Sum of Independent Random Variables

Assume that $X$ and $Y$ are two INDEPENDENT random variables defined on the same sample space $\Omega$. Then,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

More generally, if $X_1, X_2, \ldots, X_n$ are independent random variables then

$$\text{Var}(X_1, X_2 + \cdots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n).$$

Example 2.9. Roll 2 dice. Let $S$ be the sum of the two dice. What is the variance of $S$?

Let $X$ and $Y$ be the faces shown by each die. From Example 2 we know that $\text{Var}(X) = \text{Var}(Y) = 35/12$. Since $X$ and $Y$ are independent we get that

$$\text{Var}(S) = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 2 \times \frac{35}{12} = \frac{35}{6}.$$

Example 2.10. Assume that $X$ and $Y$ are independent random variables with

$$\text{Var}(X) = 2 \quad \text{Var}(Y) = 3.$$

What is the variance of $2X - 3Y$?
Exercises 2.4

From the definition of independence it is to see that if $X$ and $Y$ are independent so are $2X$ and $-3Y$. Thus,

$$\text{Var}(2X - 3Y) = \text{Var}(2X) + \text{Var}(-3Y) = 4\text{Var}(X) + 9\text{Var}(Y) = 35.$$ 

Exercises 2.4

2.1. What is the variance of a random variable uniformly distributed on $\{-1, 0, 3\}$?

2.2. Let $X$ be a random variable such that $P(X = 0) = 1/5$ and $P(X = 4) = 4/5$. Find the variance of $X$.

2.3. The probability of finding someone in favor of a certain initiative is 0.01. We interview people at random until we find a person in favor of the initiative. What is the standard deviation of the number of interviews?

2.4. Roll two dice.

(a) What is the variance of the maximum of the two dice?

(b) Compare the result of (a) to the variance of a single roll obtained in Example 2.

2.5. Let $X$ have density $f(x) = x^2e^{-x}/2$. What is the variance of $X$?

2.6. Let $U$ be a random variable which is uniformly distributed on $[1, 2]$. What is the variance of $U$?

2.7. Consider the random variables $X$ and $Y$ with densities $f(x) = \frac{3}{2}x^2$ for $x$ in $[-1, 1]$ and $g(x) = \frac{3}{4}(1 - x^2)$ for $x$ in $[-1, 1]$, respectively.

(a) Sketch the graphs of $f$ and $g$. Based on the graphs which random variable should have the largest variance?

(b) Compute the variances of $X$ and $Y$.

2.8. Let $f(x) = 3x^2$ for $x$ in $[0,1]$. Let $X$ be a random variable with density $f$. What is the variance of $X$?

2.9. Let $X$ have variance 2. What is the variance of $-3X + 1$?

2.10. Let $X$ be a measure in cm and let $Y$ be the measure of the same object in inches. How are $SD(X)$ and $SD(Y)$ related?

2.11. Roll two dice successively. Let $X$ be the face of the first die and $Y$ be the face of the second die.

(a) Find $\text{Var}(X - Y)$.

(b) Find $\text{Var}(|X - Y|)$.

2.12. A circuit has three components that work independently one of each other with probability $p_i$ for $i = 1, 2, 3$. Let $S$ be the number of components that work. Find the variance of $S$. 

2.13. What is the variance of a single roll obtained in Example 2?

2.14. What is the variance of the maximum of two dice?

2.15. What is the variance of a single roll obtained in Example 2?

2.16. What is the variance of the maximum of two dice?

2.17. What is the variance of a single roll obtained in Example 2?
2.5 Normal Random Variables

We start by giving the following definition.

Normal Random Variables
The continuous random variable $X$ is said to be a normal random variable with mean $\mu$ and standard deviation $\sigma$ if it has the density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}.$$ 

There are several things to be checked here: that $f$ is a density, that $E(X) = \mu$ and that $\text{Var}(X) = \sigma^2$. Since these computations involve calculus only they will be left as exercises.

The case $\mu = 0$ and $\sigma = 1$ is of particular interest. The density becomes

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$ 

Below is the graph of $f$. 

![Graph of Normal Distribution](image)
2.5 Normal Random Variables

We also graph below the densities of two normal densities with $\mu = 2$. One has a standard deviation equal to 1 and the other one a standard deviation equal to 2. They both have the characteristic bell shaped form. However, one can see below how much more spread out the curve with $\sigma = 2$ is compared to the one with $\sigma = 1$.

**Standard Normal Random Variable**

The continuous random variable $Z$ is said to be a standard normal random variable if it has the density

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

That is, $Z$ is a normal random variable with mean 0 and standard deviation 1.

The notation $Z$ will be reserved to standard normal random variables. In order to compute probabilities involving $Z$ we will need to integrate its density. Unfortunately, there is no explicit formula for antiderivatives of $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. We will
need to rely on a numerical table provided in the appendix. What is provided is a table for the function

\[ \Phi(x) = P(0 < Z < x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \]

**Example 2.1.** What is the probability that a standard normal random variable \( Z \) is larger than 1?

We have that

\[ P(Z > 1) = 1/2 - \Phi(1) = 1/2 - 0.34 = 0.16. \]

**Example 2.2.** What is the probability that a standard normal random variable \( Z \) is larger than \(-1\)?

By symmetry of the distribution of \( Z \) we have

\[ P(Z > -1) = P(Z < 1) = 0.84. \]

**Example 2.3.** What is the value below which a standard normal random variable is with probability 90%?

We want \( c \) such that

\[ P(Z < c) = \frac{1}{2} + \Phi(c) = 0.9. \]

We see from the table that \( c \) is between 1.28 and 1.29. Since \( c \) is closer to 1.28, we take \( c = 1.28 \).

**Example 2.4.** What is the value below which a standard normal random variable is with probability 20%?

This time we want \( c \) such that

\[ P(Z < c) = 0.2. \]

Note that \( c \) is negative. By symmetry we have that

\[ P(Z < c) = P(Z > -c) = \frac{1}{2} - \Phi(-c) = 0.2. \]

Thus,

\[ \Phi(-c) = 0.3. \]

We read in the table that \(-c \) is approximately 0.84. Thus, we have \( c = -0.84 \).

**Example 2.5.** What is the probability that a standard normal random variable \( Z \) is between \(-2 \) and \( 2 \)?

\[ P(-2 < Z < 2) = 2P(0 < Z < 2) = 2\Phi(2) \sim 0.95. \]
2.5 Normal Random Variables

So there is only a 5% chance that a standard normal distribution is larger than 2 or smaller than $-2$.

One of the nice properties of the normal distributions is that they can easily be transformed into standard normal distributions as the property below shows.

**Standardization**

If $X$ has normal distribution with mean $\mu$ and standard deviation $\sigma$ then the random variable

$$\frac{X - \mu}{\sigma}$$

is a standard normal random variable.

What is remarkable here is not that $\frac{X - \mu}{\sigma}$ has mean 0 and standard deviation 1. This is true for any random variable that has a mean and a standard deviation as it is shown below. What is remarkable is that after shifting and scaling a normal random variable we still get a normal random variable.

We now compute the expected value and standard deviation of $\frac{X - \mu}{\sigma}$.

$$E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} (E(X) - \mu) = 0,$$

where the last equality comes from the fact that $E(X) = \mu$. For the variance we have

$$\text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \text{Var}(X - \mu) = \frac{1}{\sigma^2} \text{Var}(X) = 1.$$

We will now give a few examples on how to use the property above.

**Example 2.6.** Assume that height of 6 years old are normally distributed with mean 100 cm and standard deviation 2 cm. What is the probability that a 6 years old taken at random is at least 105 cm tall?

Let $X$ be height of the child picked at random. We want $P(X > 105)$. We standardize $X$ to get

$$P(X > 105) = P\left(\frac{X - 100}{2} > \frac{105 - 100}{2}\right) = P(Z > 2.5) \sim 0.01.$$

So there is only a 1% probability that a child taken at random be at least 105 cm tall.
Example 2.7. What is the height above which 90% of the 6 years old are? We want $h$ such that $P(X > h)$. We standardize $X$ again to get

$$P(X > h) = P\left(\frac{X - 100}{2} > \frac{h - 100}{2}\right) = P\left(Z > \frac{h - 100}{2}\right) = 0.9.$$  

Note that $\frac{h - 100}{2}$ must be negative. By symmetry of the distribution of $Z$ we have that

$$P\left(Z > \frac{h - 100}{2}\right) = P\left(Z < \frac{-h + 100}{2}\right) = 0.9.$$  

So according to the Normal table we have

$$\frac{-h + 100}{2} = 1.28.$$  

We solve for $h$ and get $h$ that is approximately 97.44 cm.

Example 2.8. Let $X$ be normally distributed with mean $\mu$ and standard deviation $\sigma$. What is the probability that $X$ is $2\sigma$ or more away from its mean? We want

$$P(\{X > \mu + 2\sigma\} \cup \{X < \mu - 2\sigma\}) = P(X > \mu + 2\sigma) + P(X < \mu - 2\sigma),$$  

where the last equality comes from the fact that the two events are disjoint. We standardize $X$ to get

$$P(\{X > \mu + 2\sigma\} \cup \{X < \mu - 2\sigma\}) = P\left(\frac{X - \mu}{\sigma} > 2\right) + P\left(\frac{X - \mu}{\sigma} < -2\right)$$

$$= P(Z > 2) + P(Z < -2) = 0.05.$$  

2.5.1 Extreme Observations

As we have just seen the normal distribution is concentrated around its mean and it is unlikely that an observation taken at random is more than $2\sigma$ away from its mean (see Example 2.8). However, if we make several independent observations, what is the probability that the largest or the smallest of the observations is far away from the mean? Next we look at a particular example.

Example 2.9. Assume that height of 6 years old are normally distributed with mean 100 cm and standard deviation 2 cm. In a group of 25 children what is the probability that the tallest of the group is at least 105 cm tall?
Let $X_1, \ldots, X_{25}$ be the heights of the 25 children in the group. We are interested in the probability that the maximum of these random variables be at least 105. It is easier to deal with the complement of the preceding event. Note that the maximum of the 25 observations is less than 105 cm if and only if each one of the observations is less than 105 cm. Thus,

$$P(\max(X_1, \ldots, X_{25}) < 105) = P(\{X_1 < 105\} \cap \{X_2 < 105\} \cap \ldots \cap \{X_{25} < 105\})$$

$$= P(X_1 < 105) P(X_2 < 105) \ldots P(X_{25} < 105),$$

where the last equality comes from the independence of the $X_i$. According to Example 6, we have that $P(X_1 < 105)$ is $P(Z > 2.5) = 0.9876$ and this probability is the same for each $X_i$ since they all have the same distribution. Thus,

$$P(\max(X_1, \ldots, X_{25}) < 105) = (0.9876)^{25} \sim 0.73.$$  

That is, the probability that the tallest child in a group of 25 is at least 105 is 0.27. As the group increases this probability increases as well. For a group of 50 this probability becomes about 0.5. For a group of 100 this probability becomes about 0.7.

The important conclusion of this example is the following. Extreme observations (especially if there are many observations) are likely to be far from a typical observation.

**Exercises 2.5**

2.1. Let $Z$ be a standard normal random variable. Compute the following.

(a) $P(Z > 1.52)$.
(b) $P(Z > -1.15)$.
(c) $P(-1 < Z < 2)$.
(d) $P(-2 < Z < -1)$.

2.2. Let $Z$ be a standard normal random variable. What is the value above which $Z$ is with 99% of probability?

2.3. Assume that $X$ is normally distributed with mean 3 and standard deviation 2.

(a) $P(X > 3) = ?$
(b) $P(X > -1) = ?$
(c) $P(-1 < X < 3) = ?$
(d) $P(|X - 2| < 1) = ?$

2.4. Assume that the diameter of a ball bearing is normally distributed with mean 1 cm and standard deviation 0.05 cm. A ball bearing is considered defective if its diameter is larger than 1.1 cm or smaller than 0.9 cm.
2.5. Assume that $X$ is normally distributed with mean 5 and standard deviation $\sigma$. Find $\sigma$ so that $P(X > 4) = 0.95$.

2.6. Assume that the annual snow fall at some place is normally distributed with mean 20 in. and standard deviation 8 in.

(a) What is the probability that the snow fall be less than 5 in. on a given year?

(b) What is the probability that the smallest annual snow fall in the next 20 years will be less than 5 in.?

2.7. Let $Z$ be a standard normal random variable with density

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$ 

In this exercise we will check that $f$ is actually a density.

(a) Change the variables from Cartesian to polar to show that

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2} dx dy = \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\rho^2/2} \rho d\rho d\theta.$$ 

(b) Show that the r.h.s. of (a) is $2\pi$.

(c) Show that the l.h.s. of (a) is

$$\left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx\right)^2.$$ 

(d) Conclude that $f$ is a density.

2.8. Let $Z$ be a standard normal random variable.

(a) Compute $E(Z)$.

(b) Compute $\text{Var}(Z)$.

2.9. Let

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2}.$$ 

Show that $f$ has inflection points at $\mu + \sigma$ and $\mu - \sigma$. 
Review Exercises for Chap. 2

2.1. Three people toss one fair coin each. The winner is the one whose coin shows a face different from the two others. If the three coins show the same face then there is a new round of tosses, until someone wins.

(a) What is the probability of exactly one round of tosses?
(b) What is the probability that at least three rounds of tosses are necessary?

2.2. A and B take turns rolling a die. A starts. The winner is the first one that rolls a 6. What is the probability that A wins?

2.3. Two people play the following game. They toss two fair coins. If the two coins land on heads then A wins. If one coin lands on heads and the other on tails then B wins. If the two coins land on tails then the coins are tossed again until someone wins. What is the probability that B wins?

2.4. The probability of finding someone in favor of a certain initiative is 0.01. We interview people at random until we find a person in favor of the initiative. What is the probability that we need to conduct 50 or more interviews?

2.5. Draw five cards from a 52 cards deck.

(a) Explain why the probability the second card is red is the same as the probability of the second card is black.
(b) What is the expected number of red cards among the five cards that have been drawn.
(c) What is the expected number of hearts in five cards dealt from a deck of 52 cards?

2.6. Assume that car batteries lifetimes follow an exponential distribution with mean 3 years.

(a) What is the probability that a battery lasts 10 years or more?
(b) In a group of ten batteries what is the probability that at least one will last 10 years or more?
(c) How many batteries do we need in order to have at least one last 10 years or more with probability 0.9?

2.7. Let $X$ a random variable with density $f(x) = ce^{-x}$.

(a) Find $c$.
(b) What is the $P(X > 1)$?

2.8. Let $X$ have density $g(x) = c(x - 1)^2$ for $x$ in [0, 2].

(a) Find $c$.
(b) Find $E(X)$.
(c) Find $Var(X)$.
2.9. Let \( Y \) be a random variable with density \( f(y) = c(-(y - 1)^2 + 2) \) for \( y \) in \([0, 2]\).

(a) Sketch the graphs of \( g \) in Exercise 8 and of \( f \).
(b) Which random variable \( X \) or \( Y \) do you expect to have the highest variance?
(c) Confirm your prediction by doing a computation.

2.10. Roll two dice. I win $1 if the sum is 7 or more. I lose $b$ if the sum is 6 or less. Find \( b \) so that this is a fair game.

2.11. Toss five fair coins.

(a) What is the expected number of heads?
(b) What is the variance of the number of heads?

2.12. Suppose atoms of a given kind have an exponential distributed lifetime with mean 30 years. What is the expected number of atoms still present after 30 years if we start with \( 10^{23} \) atoms?

2.13. Ball bearings are manufactured with diameters that are normally distributed with mean 1 cm and standard deviation 0.05 cm. Assume that 1,000 ball bearings are manufactured. What is the expected number of ball bearings whose diameter is at least 1.1 cm?

2.14. Assume that the random variable \( T \) is such that \( E(T) = 1 \) and \( ET(T - 1) = 2 \). What is the standard deviation of \( T \)?

2.15. It is believed that in the 1700s in Europe life expectancy at birth was only around 40 years. That is, a newborn baby could expect on average to live 40 years. It is also known that child mortality was extremely high. Maybe, as many as 50% of all babies did not make it to their fifth birthday.

(a) Compare the median life span to the expected life span.
(b) Were people old at 35?

2.16. I have 100 balls in an urn numbered from 1 to 100. I draw at random one ball at a time and then I put it back in the urn.

(a) What is the expected number of draws to get ten different numbers?
(b) What is the expected number of draws to get all the 100 different numbers?
Probability with Statistical Applications
Schinazi, R.B.
2012, XI, 347 p. 28 illus., Hardcover
ISBN: 978-0-8176-8249-1
A product of Birkhäuser Basel