Probability I

Games of chance, such as those involving dice, have been played for over 5,000 years. Ever since 1550 A.D., mathematicians have been involved with calculating the probabilities or chances of winning at gambling. For instance, in approximately 1620, Galileo wrote a paper on dice probabilities. The year 1654 is often thought of as the beginning of probability theory since it was at that time that Blaise Pascal and Pierre de Fermat began correspondence on the subject. Pierre Simon, the Marquis de Laplace, was an important contributor to probability theory; in 1812 he proved the central limit theorem which provides an explanation as to why so many data sets follow a distribution that is bell-shaped, i.e., normally distributed. In Laplace’s book *Analytical Theory of Probability* he says:

“We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it ... It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge. ... The most important questions in life are, for the most part, really only problems of probability.”

While Laplace may have overstated the case, probability theory is certainly a very important modeling tool in quantum mechanics, in the financial markets, and in biology.

Probability is an extension of the idea of a proportion, or the ratio of a part to the whole. Suppose that there are 6 green balls and 4 red balls in a urn. You mix them well and then reach in without looking and pull one out. Since 60% of the balls are green, we expect “on average” that about 60% of the time we will get a green ball. Intuitively, here is what the phrase “on average” means. After noting the color of the ball, we put it back, mix well and select again. Let $N_g$ denote the number of green balls that we have obtained after performing this experiment $N$ times. Then, we expect the ratio $N_g/N$ to get close to 0.6 as $N$ increases and similarly, the probability of getting a red ball is 0.4. Selecting a ball is called a “random event” because the outcome is uncertain.

An experiment is a well-defined procedure. The set of possible outcomes of the experiment is called the sample space.

So, in our example above the experiment is putting 6 green balls and 4 red balls in an urn, mixing well, and selecting one ball at random (without looking). The sample space has two elements in it, $G$ (for green) and $R$ (for red).

**Example 1.** Consider the experiment of rolling a single die. There are six possible outcomes: one of the numbers 1, 2, 3, 4, 5, or 6. In this case, the sample space has six outcomes. Assuming that we have a fair die (that is, each side is equally likely to turn up), we say that the probability of each of these outcomes is $1/6$. 
Events

An event is a subset of the sample space of possible outcomes. If $A$ is an event, we denote by $\mathbb{P}(A)$ the probability that the event will happen.

In the case of rolling a die, an event is a subset of the set of six integers $\{1, 2, 3, 4, 5, 6\}$. Suppose that $A$ is the event that the die shows an even number. That is, $A = \{2, 4, 6\}$. Then

$$\mathbb{P}(A) = \mathbb{P}(\{2, 4, 6\}) = \frac{3}{6},$$

since the event $\{2, 4, 6\}$ contains 3 out of the 6 equally likely outcomes. Similarly, suppose that $B$ is the event that the outcome shows a number less than 3. Then, $B = \{1, 2\}$ and $\mathbb{P}(B) = \frac{1}{3}$.

There are two fundamental principles to remember:

1. The probability of an outcome or an event is always a number between zero and one $\left(0 \leq p \leq 1\right)$, because it is the proportion of repeated experiments (on average) in which the outcome or event occurs.

2. The sum of the probabilities of the possible outcomes of an experiment must add up to one; that is, one of the outcomes must occur.

Example 2. Consider the experiment of flipping a coin three times. If we denote a head by $H$ and a tail by $T$, we can list the eight possible outcomes:

$$(H, H, H), (H, H, T), (H, T, H), (T, H, H), (T, T, H), (T, H, T), (H, T, T), (T, T, T)$$

each of which occurs with probability $\frac{1}{8}$. To calculate the probability of any particular event (subset of the sample space of outcomes), we add the probabilities of the outcomes in the event set. Thus,

(a) The probability that all three flips are heads = $\frac{1}{8}$.

(b) The probability that exactly two flips are heads = $\frac{3}{8}$.

(c) The probability that the first flip is tails = $\frac{1}{2}$.

(d) The probability that at least one flip is heads = $\frac{7}{8}$.

(e) The probability that heads and tails alternate = $\frac{1}{4}$. 
Adding Probabilities

Suppose that we are considering two events $A$ and $B$ and we would like to know the probability of either $A$ or $B$ happening? Is it $P(A) + P(B)$? Let’s look at another example.

**Example 3.** Suppose that we roll a die twice. We denote the outcome of the first roll by $X$ and the outcome of the second roll by $Y$. Suppose we want to find

$$P(X \geq 3 \text{ or } Y \geq 2).$$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,1</td>
<td>1,2</td>
<td>1,3</td>
<td>1,4</td>
<td>1,5</td>
<td>1,6</td>
</tr>
<tr>
<td>2</td>
<td>2,1</td>
<td>2,2</td>
<td>2,3</td>
<td>2,4</td>
<td>2,5</td>
<td>2,6</td>
</tr>
<tr>
<td>3</td>
<td>3,1</td>
<td>3,2</td>
<td>3,3</td>
<td>3,4</td>
<td>3,5</td>
<td>3,6</td>
</tr>
<tr>
<td>4</td>
<td>4,1</td>
<td>4,2</td>
<td>4,3</td>
<td>4,4</td>
<td>4,5</td>
<td>4,6</td>
</tr>
<tr>
<td>5</td>
<td>5,1</td>
<td>5,2</td>
<td>5,3</td>
<td>5,4</td>
<td>5,5</td>
<td>5,6</td>
</tr>
<tr>
<td>6</td>
<td>6,1</td>
<td>6,2</td>
<td>6,3</td>
<td>6,4</td>
<td>6,5</td>
<td>6,6</td>
</tr>
</tbody>
</table>

Table 1: Outcomes of rolling a die twice

This is the probability of either getting a 3, 4, 5, or 6 on the first roll, or getting a 2, 3, 4, 5, or 6 on the second roll. When we say $(A \text{ or } B)$ we mean either $A$ or $B$ or both. The probability that $X$ is greater than or equal to three is $4/6$ and the probability that $Y$ is greater than or equal to two is $5/6$. If we add these probabilities we get $9/6$, which should set off alarm bells since we cannot have a probability greater than one.

Let’s figure out what is going on here.

First, the only time when $(X \geq 3 \text{ or } Y \geq 2)$ is not true is when $X = 1$ and $Y = 1$ or when $X = 2$ and $Y = 1$. Thus there are only two events out of 36 possible events where our condition is not satisfied; therefore, the correct probability is $(36 - 2)/36 = 34/36$. The condition $X \geq 3$ (which has probability $4/6$) corresponds to all the events in the bottom 4 rows of Table 1. The condition $Y \geq 2$ (probability $5/6$) corresponds to all events in the 5 right-hand columns of Table 1. When we add the probability of $X \geq 3$ to the probability of $Y \geq 2$ we are double counting all of the events which are both in the bottom 4 rows and in the 5 right-hand columns! Notice that there are $5 \cdot 4 = 20$ different outcomes in the intersection of the bottom 4 rows and the 5 right-hand columns. Thus, we need to subtract the probability of the outcomes in the intersection, $20/36$. Notice that the intersection of the bottom 4 rows and the 5 right-hand rows corresponds to the event $(X \geq 3 \text{ and } Y \geq 2)$. Thus, we have

$$P(X \geq 3 \text{ or } Y \geq 2) = P(X \geq 3) + P(Y \geq 2) - P(X \geq 3 \text{ and } Y \geq 2)$$

$$= \frac{4}{6} + \frac{5}{6} - \frac{20}{36}$$

$$= \frac{34}{36}.$$
In general, we have the following rule:

**Addition Rule:** \( P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B). \)

Thus, the probability of the event \((A \text{ or } B)\) is equal to the probability of \(A\) plus the probability of \(B\) minus the probability of \((A \text{ and } B)\), see the picture below.

Note that in terms of sets the event \((A \text{ or } B)\) corresponds to the set \(A \cup B\) and the event \((A \text{ and } B)\) corresponds to the set \(A \cap B\). Thus, we can rewrite the above rule as

**Addition Rule:** \( P(A \cup B) = P(A) + P(B) - P(A \cap B). \)

**Example 4:** A card is selected at random from a deck of 52 cards. We wish to find the probability that the card selected will be a Queen or a diamond. Letting \(Q\) represent the event that a Queen is selected and \(D\) represent the event that a diamond is selected, and using the addition rule, we have

\[
P(Q \text{ or } D) = P(Q) + P(D) - P(Q \text{ and } D)
\]

\[
= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}
\]

\[
= \frac{4}{13}.
\]

We can confirm this computation by noting that a deck of cards has 13 diamonds and 3 queens which are not diamonds; thus \(\frac{16}{52} = \frac{4}{13}\) of the cards we are after are either queens or diamonds.

**Independence**

Suppose that we are considering the outcomes of a particular experiment. If \(A\) and \(B\) are events (subsets of the sample space of outcomes), we say that \(A\) and \(B\) are **independent** if

\[
P(A \text{ and } B) = P(A)P(B). \tag{1}
\]

As we shall see in the following example, \(A\) and \(B\) are independent if knowing that \(A\) is true (that is, that the outcome is in \(A\)) does not affect the probability that \(B\) is true and vice-versa. This is why we use the word “independent.”

**Example 5.** Consider the experiment in Example 2 where we flipped a coin 3 times. Let \(A\) be the event that a head comes up on the first flip. There are 4 outcomes in the sample space
for which this is true – each with probability $\frac{1}{8}$. Therefore, $P(A) = \frac{1}{2}$. Let $B$ be the event that a head comes up on the third flip. Again, we calculate easily that $P(B) = \frac{1}{2}$. What is the probability of $(A$ and $B)$? This is the subset $A \cap B$ of outcomes such that there is a head on the first flip and a head on the third flip. There are two such outcomes, $(H, T, H)$ and $(H, H, H)$, each with probability $\frac{1}{8}$. Thus, $P(A$ and $B) = \frac{1}{4}$. Therefore,

$$P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A$ and $B).$$

So, by definition, the events $A$ and $B$ are independent. This makes sense because the outcome of the first flip should not affect the probability of the outcome of the third flip. To see an example of events that are not independent, let $A$ again denote the event that the first flip is a head. Let $C$ denote the event that at least two of the three flips are heads. Looking at the list of outcomes, we see that in four of them there are two or more heads. Thus, $P(C) = \frac{1}{2}$. The event $(A$ and $C)$ is the set of outcomes in which the first flip is a head and at least two out of the three are heads. Consulting the list of possible outcomes we see that there are three outcomes for which this is true. Thus $P(A$ and $C) = \frac{3}{8}$. Since

$$P(A)P(C) = \frac{1}{2} \cdot \frac{1}{2} \neq \frac{3}{8} = P(A$ and $C),$$

we see that $A$ and $C$ are not independent. This makes sense because getting a head on the first flip surely affects the probability that we will get at least two heads in the three flips.

Example 6. Suppose that we roll a die three times. Each time we roll there are 6 possible results for that roll, so the total number of outcomes is $6^3$. Therefore, the probability of any particular outcome, say $(5, 1, 4)$, is $\frac{1}{6^3}$. Let $A$ be the event that we get a 4 on the second roll and not a 4 on both the first and third rolls. Let $B_1$ be the event that we do not get a 4 on the first roll; then $P(B_1) = \frac{5}{6}$. Let $B_2$ be the event that we do get a 4 on the second roll; then $P(B_2) = \frac{1}{6}$. Let $B_3$ be the event that we do not get a 4 on the third roll. Then $P(B_3) = \frac{5}{6}$. Now, $A = (B_1$ and $B_2$ and $B_3)$. Therefore, if $B_1$, $B_2$, and $B_3$ are all independent (as we believe they are), then the probability of $A$ should just be the product of the probabilities of $B_1$, $B_2$, and $B_3$. That is,

$$P(A) = P(B_1)P(B_2)P(B_3) = \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6}.$$ 

Now, let’s use these ideas to calculate the probability that if we roll a die three times we will get exactly one 4. This event could happen in three ways. We could get a 4 the first roll and then two non-4s; call this the event $C_1$. We could get a non-4, a 4, and then a non-4; this is just the event $A$ above. Finally, we could get two non-4s, and then a 4; call this the event $C_3$. Each of these three events has probability

$$\left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^2$$
so,

\[ \mathbb{P}(C_1) + \mathbb{P}(A) + \mathbb{P}(C_3) = 3 \left( \frac{1}{6} \right) \cdot \left( \frac{5}{6} \right)^2 \]

We can add the probabilities here because if there is exactly one 4 it could be on the first roll or the second but not both; that is, \( \mathbb{P}(C_1 \cap A) = 0 \). Similarly, it could be on the second or the third roll but not both; that is, \( \mathbb{P}(A \cap C_3) = 0 \).
Homework Problems

1. Suppose that a bag contains 7 black balls, 6 yellow balls, 4 green balls, and 3 red balls.
You shake the bag well, and remove one ball without looking into the bag.

   (a) What is the probability that the ball you remove is red? black? yellow? green? white?
   (b) What is the probability that the ball you pick is either black or green?
   (c) What is the probability that you have picked a ball whose color is not red?

2. A die is painted so that three sides are red, two sides are blue and one side is green.
Thus, rolling the die has three possible outcomes $R$, $B$, and $G$.

   (a) What is the probability the die will come up blue?
   (b) What is the probability the die will not come up red?
   (c) What is the probability that the face showing is either red or blue?

3. Suppose that we flip a fair coin four times.

   (a) How many outcomes are in the sample space?
   (b) What is the probability that all four flips are tails?
   (c) What is the probability that there will be one or fewer heads?
   (d) What is the probability that there will be equal numbers of heads and tails?

4. The painted die from problem 2 is rolled twice. Denote the nine possible outcomes by $RR$, $RB$, etc..

   (a) Find the probability of each element of the sample space.
   (b) What is the probability that at least one roll will be red?
   (c) What is the probability that neither roll is blue?
   (d) What is the probability that the two rolls will have different colors?

5. If we roll a fair die, what is the probability that after six rolls we:

   (a) do not get a 6?
   (b) get a 6 on the first roll, but not after?
   (c) get exactly one 6?

6. The painted die from problem 2 is rolled twice. Use the addition rule to find the following probabilities.

   (a) The probability that either both rolls are red or both rolls are blue.
   (b) The probability that either both rolls are red or exactly one roll is blue.
   (c) The probability that either at least one roll is red or exactly one roll is blue.
   (d) The probability that either at least one roll is red or at least one roll is blue.
7. Suppose that a fair coin is flipped three times.

(a) Let $A$ be the event that the first flip is heads. Let $B$ be the event that the second and third flips are the same. Find $P(A \text{ and } B)$, and prove that $A$ and $B$ are independent.

(b) Let $A$ be the event that the first flip is heads. Let $B$ be the event that at least two flips are heads. Find $P(A \text{ and } B)$ and prove that $A$ and $B$ are not independent. Why does this make sense?

8. Suppose that a fair die is rolled twice.

(a) Let $A$ be the event that the first roll is $\geq 2$. Let $B$ be the event that the second roll is $\leq 4$. Find $P(A \text{ and } B)$ and prove that $A$ and $B$ are independent.

(b) Let $A$ be the event that the first roll is $\geq 2$. Let $B$ be the event that the sum of the rolls is $\leq 4$. Find $P(A \text{ and } B)$, and prove that $A$ and $B$ are not independent. Why does this make sense?

Selected Answers.

1. (i) $\frac{3}{20}, \frac{7}{20}, \frac{6}{20}, \frac{4}{20}, 0$; (ii) $\frac{11}{20}$; (iii) $\frac{17}{20}$.

3. (i) 16; (ii) $\left(\frac{1}{2}\right)^4$; (iii) $\frac{5}{16}$; (iv) $\frac{3}{8}$.

5. (i) $\left(\frac{5}{6}\right)^6$; (ii) $\left(\frac{5}{6}\right)^5 \cdot \frac{1}{6}$; (iii) $6 \left(\frac{5}{6}\right)^5 \frac{1}{6}$.

7. (i) $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \text{ and } B) = \frac{1}{4}$.

(ii) $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \text{ and } B) = \frac{3}{8}$.
**Probability II**

In this lesson we introduce the concept of random variable, a central idea of probability theory. Consider an experiment which has a set, $S$, of possible outcomes. $S$ is the sample space. A random variable is a function defined on $S$ which takes values in the real numbers. That is, if $X$ is the name of the random variable and $s$ is a possible outcome of the experiment, then $X(s)$ is a number.

**Example 1.** Consider the experiment of flipping a coin three times. The sample space of possible outcomes is

\[
\{(H, H, H), (H, H, T), (H, T, H), (T, H, H), (T, H, T), (H, T, T), (T, T, H), (T, T, T)\}
\]

Let $X$ be the random variable which assigns to each outcome the number of heads in the outcome. Then,

\[
X((H, H, H)) = 3, \quad X((H, H, T)) = 2, \quad X((H, T, H)) = 2, \quad X((T, H, H)) = 2, \\
X((T, H, T)) = 1, \quad X((T, T, H)) = 1, \quad X((H, T, T)) = 1, \quad X((T, T, T)) = 0.
\]

Thus, $X$ can take the values 0, 1, 2, and 3. How likely is it that $X$ will have each of these values? Each outcome has probability $\frac{1}{8}$. There is only one outcome with 3 heads, so the probability that $X = 3$ is $\frac{1}{8}$. We write

\[
P(X = 3) = \frac{1}{8}.
\]

Since there are three outcomes where $X = 2$, we have

\[
P(X = 2) = \frac{3}{8}.
\]

Similarly, $P(X = 1) = \frac{3}{8}$ and $P(X = 0) = \frac{1}{8}$.

Here is another random variable on the same sample space. Let $Y$ be the number of heads minus the number of tails. If we consult the list of possible outcomes above, we see that $Y$ can take the values 3, 1, −1, and −3, and $P(Y = 3) = \frac{1}{8}$, $P(Y = 1) = \frac{3}{8}$, $P(Y = -1) = \frac{3}{8}$, $P(Y = -3) = \frac{1}{8}$.

**Example 2.** Suppose that we have a special die that is painted red on three sides, blue on two sides, and green on one side. We are going to play the following game. You roll the die and if red comes up, you pay me one dollar. If blue comes up you receive two dollars. If green comes up, no money changes hands. Let $Z$ be the random variable which is the payoff to you after you roll. Then, $Z$ takes on three values, 2, 0, and −1 with the following probabilities:

\[
P(Z = 2) = \frac{1}{3}, \quad P(Z = 0) = \frac{1}{6}, \quad P(Z = -1) = \frac{1}{2}.
\]

Since we will be very interested in the probabilities that a given random variable takes various values, it is convenient to define a function, $p(x)$, which has this information in it. Suppose that $X$ is a random variable. If $x$ is a possible value of $X$, we define

\[
p(x) = P(X = x).
\]
If $x$ is not a possible value of $X$, we define $p(x) = 0$. The function $p(x)$ is called the **mass density** of the random variable $X$.

**Example 1 (again).** If $X$ is the number of heads in three flips, then, $p(0) = \frac{1}{8}$, $p(1) = \frac{3}{8}$, $p(2) = \frac{3}{8}$, $p(3) = \frac{1}{8}$. The value of the function $p$ at any other $x$ is 0.

**Example 2 (again).** Let $p(x)$ be the mass density for the payoff function $Z$ in Example 2. Then, $p(-1) = \frac{1}{2}$, $p(0) = \frac{1}{6}$, $p(2) = \frac{1}{3}$. The value of the function $p$ at any other $x$ is 0.

It is sometimes helpful to graph the nonzero values of $p$ with straight lines from the values down to the $x$-axis. Figures 1(a) and 1(b) graph the mass density functions for the random variables $X$ of Example 1 and $Z$ of Example 2. The graph shows visually which values have positive probability and the length of the line shows how much probability is associated with that value of $x$.

In Example 2 we see that the sum of the non-zero values of $p$ is $p(-1) + p(0) + p(2) = 1$. In fact, we must always have the condition

$$\sum_{\{x|p(x)>0\}} p(x) = 1 \quad (2)$$

because the sum of the probabilities that $X$ takes on all possible values is 1. The sum is over the set of numbers $x$ such that $p(x) > 0$.

**Example 3.** Consider the experiment of rolling two dice. Let the random variable $X$ be the sum of the numbers on the two faces showing. The possible outcomes of the experiment are shown in Table 1 of the last lesson. Each outcome has probability $\frac{1}{36}$. The random variable $X$ takes the values $2, 3, 4, \ldots, 12$. There is only one outcome, $\{1, 1\}$, such that the sum of the faces is 2. Thus,

$$p(2) = \mathbb{P}(X = 2) = \frac{1}{36},$$

There are two outcomes, $\{1, 2\}$ and $\{2, 1\}$, such that $X$ has the value 3, so

$$p(3) = \mathbb{P}(X = 3) = \frac{2}{36}.$$
We can continue in this way and determine $p(k)$ for all the numbers $k = 2, 3, 4, \ldots, 12$. The function $p$ is largest when $k = 7$. There are six outcomes, \{1, 6\}, \{2, 5\}, \{3, 4\}, \{4, 3\}, \{5, 2\}, \{6, 1\}, for which $X = 7$, so

$$p(7) = \mathbb{P}(X = 7) = \frac{6}{36}.$$ 

The graph of the mass density function for $X$ is shown in Figure 2. Check that the formula (1) is true by calculating that $p(2) + p(3) + \ldots + p(12) = 1$. 

In all of our examples so far, the random variables could take only finitely many values. Here is an example of a random variable with infinitely many values.

**Example 4.** Consider the following experiment. We flip a coin until heads comes up. The sample space for this experiment is:

$$\{H, TH, TTH, TTTTH, TTTTTH, \ldots\}$$

Notice that there are infinitely many different outcomes in the sample space. Let $X$ be the number of flips required. The probability of getting a head on the first flip is $\frac{1}{2}$, so

$$p(1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$ 

If heads occurs for the first time on the second flip, that means that we got a tail on the first flip. The probability of flipping a coin twice and getting tails the first time and heads the second time is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Thus,

$$p(2) = \mathbb{P}(X = 2) = \frac{1}{4}.$$ 

If heads occurs for the first time on the third flip, that means that we got a tail on the first two flip. The probability of flipping a coin three times and getting tails the first two times and heads the third time time is $\left(\frac{1}{2}\right)^3$, so

$$p(3) = \mathbb{P}(X = 3) = \frac{1}{8}.$$ 

Continuing in this way, we see that

$$p(n) = \mathbb{P}(X = n) = \left(\frac{1}{2}\right)^n.$$ 

(3)
Thus, $X$ is a random variable whose value can be any positive integer and the probability that it takes the value $n$ is $\left(\frac{1}{2}\right)^n$. In this case, condition (1) implies:

$$\frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 + \ldots = 1.$$ 

The left-hand side is an infinite sum of positive numbers. Could this sum add up to something finite? In 112L, we shall see that, indeed, infinite sums can sometimes make sense and add up to finite numbers. In particular, we shall see that the sum on the left, which is a special case of something called a “geometric series”, does add up to 1.

**Example 5.** Let us suppose that you own one share of stock for two years and that each year the value of the stock either rises by one dollar (probability $= \frac{1}{3}$), stays the same (probability $= \frac{1}{3}$), or declines by one dollar (probability $= \frac{1}{3}$). Suppose that the change in value in the first year is independent of the change in value in the second year. Let $X$ be the profit after two years. What is the probability density of $X$?

The sample space is the set of pairs ($c_1$, $c_2$) where $c_1$ is the change in the first year and $c_2$ is the change in the second year. That is, $c_1$ is either +$1$ or 0 or −$1$ and similarly for $c_2$. Thus, there are nine possible outcomes, (−$1$, −$1$), (−$1$, 0), . . . The probability of a decline of one dollar in the first year is $\frac{1}{3}$ and the probability of a decline of one dollar in the second year is $\frac{1}{3}$. Since we are assuming that the changes in the two years are independent, we calculate that the probability of the outcome (−$1$, −$1$) (i.e. two declines) is $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$. Similarly, we see that the probability of each of the nine outcomes ($c_1$, $c_2$) is $\frac{1}{9}$.

The profit $X$ is $c_1 + c_2$, that is, the total change in value after two years. Since $c_1$ and $c_2$ are each +$1$ or 0 or −$1$, the possible values of $X$ are −$2$, −$1$, 0, +$1$, +$2$. There is only one outcome, namely (−$1$, −$1$), such that $X = −2$ so,

$$p(−2) = P\{X = −2\} = \frac{1}{9}.$$ 

However, notice that there are two outcomes in the sample space, (−$1$, 0) and (0, −$1$) such that $X = −1$. Thus,

$$p(−1) = P\{X = −1\} = \frac{2}{9}.$$ 

And, there are three outcomes in the sample space, (−$1$, +$1$), (0, 0), and(+$1$, −$1$) such that $X = 0$. Thus,

$$p(0) = P\{X = 0\} = \frac{3}{9}.$$ 

Similarly, we see that

$$p(1) = P\{X = 1\} = \frac{2}{9}.$$ 

and

$$p(2) = P\{X = 2\} = \frac{1}{9}.$$ 

**Example 6.** As in Example 5, we suppose that you own one share of stock for two years and that each year the value of the stock either rises by one dollar, stays the same, or declines by one dollar and the changes in the two years are independent. However, we assume that in the first year, the change +$1$ has probability $\frac{1}{2}$, the change 0 has probability $\frac{1}{4}$ and
the change \(-$1\) has probability \(\frac{1}{4}\). In the second year, we assume that the change \(+$1\) has probability \(\frac{1}{3}\), the change \($0\) has probability \(\frac{1}{2}\) and the change \(-$1\) has probability \(\frac{1}{6}\). Find the probability density of the profit, \(X\), after two years. The sample space is the same as in Example 6 and, as before, \(X\) has the possible values \(-$2\), \(-$1\), \($0\), \(+$1\), \(+$2\). However, now the different outcomes in the sample space have different probabilities. For example, the probability of the outcome \((-$1,-$1)\) is \(\frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24}\). Once one knows the probabilities of all the outcomes, one can find the probability that \(X = $n\) by adding the probabilities of all outcomes \((c_1, c_2)\) such that \(c_1 + c_2 = $n\). The student is asked to do this in problem 7 below.
Homework Problems

1. Consider the experiment of flipping a fair coin twice. Let $X$ be the number of heads minus the number of tails. What are the possible values of $X$? Find and graph the probability mass density of $X$.

2. Consider the experiment of flipping a biased coin (which comes up heads with probability $\frac{3}{4}$) twice. Let $X$ be the number of heads minus the number of tails. What are the possible values of $X$? Find and graph the probability mass density of $X$.

3. Consider the experiment of flipping a fair coin three times. Let $X$ be the square of the number of heads. What are the possible values of $X$? Find and graph the probability mass density of $X$.

4. Consider the experiment of flipping a fair coin three times. Let $X$ be the square of the number of heads minus two times the number of tails. What are the possible values of $X$? Find and graph the probability mass density of $X$.

5. An encyclopedia salesman visits three customers each day and with each he has a probability of $\frac{1}{4}$ of making a sale. For each sale he earns a commission of $100 and if he makes three sales in a day he earns a $50 bonus from his company. Let $X$ be his daily earnings. What are the possible values of $X$? What is the mass density of $X$?

6. Consider the following game. We choose a ball at random from an urn containing 7 red, 3 green, and 2 amber balls. We win $2 if we choose a green ball, we lose $1 if we choose a red ball, and we don’t win or lose any money if we choose an amber ball. Let $X$ represent our winnings. Find the probability density of $X$.

7. Complete the calculations in Example 6 by finding the probability of each of the nine outcomes. Then find the probability mass density of $X$, the profit.

8. You own one share of stock for three years and each year the value of the stock either rises by one dollar (probability $= \frac{2}{3}$) or declines by one dollar (probability $= \frac{1}{3}$); it never stays the same. Suppose that the yearly changes are independent of each other. What is the sample space? What is the probability of each outcome in the sample space? Let $X$ be the profit after three years. What values can $X$ have? What is the probability mass density of $X$?

9. You own one share of stock for two years and each year the value of the stock changes by +$2, +$1, $0, -$1 each with probability $\frac{1}{4}$. Suppose that the changes in the two years are independent. What is the sample space? What is the probability of each outcome in the sample space? Let $X$ be the profit after two years. What values can $X$ have? What is the probability mass density of $X$?

10. You own one share of stock for two years. The first year the value of the stock changes by +$1, $0, -$1, with probabilities $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$, respectively. The second year the value of the stock changes by +$1, +$2, +$3, with probabilities $\frac{1}{3}, \frac{1}{2}, \frac{1}{6}$, respectively. Suppose that the changes in the two years are independent. What is the sample space? What is the probability of each outcome in the sample space? Let $X$ be the profit after two years. What values can $X$ have? What is the probability mass density of $X$?
Answers to Selected Homework Problems

1. $\P\{X = 2\} = \frac{1}{4}$, $\P\{X = 0\} = \frac{1}{2}$, $\P\{X = -2\} = \frac{1}{4}$.

3. $\P\{X = 0\} = \frac{1}{8}$, $\P\{X = 1\} = \frac{3}{8}$, $\P\{X = 4\} = \frac{3}{8}$, $\P\{X = 9\} = \frac{1}{8}$.

5. $\P\{X = 0\} = \frac{27}{64}$, $\P\{X = 100\} = \frac{27}{64}$, $\P\{X = 200\} = \frac{9}{64}$, $\P\{X = 350\} = \frac{1}{64}$.

7. $\P\{X = -2\} = \frac{1}{24}$, $\P\{X = -1\} = \frac{4}{24}$, $\P\{X = 0\} = \frac{7}{24}$, $\P\{X = +1\} = \frac{8}{24}$, $\P\{X = +2\} = \frac{4}{24}$.

9. $\P\{X = 4\} = \frac{1}{16}$, $\P\{X = 3\} = \frac{2}{16}$, $\P\{X = 2\} = \frac{3}{16}$, $\P\{X = 1\} = \frac{4}{16}$, $\P\{X = 0\} = \frac{3}{16}$, $\P\{X = -1\} = \frac{2}{16}$, $\P\{X = -2\} = \frac{1}{16}$. 
Probability III

In this lecture we introduce the concept of expectation (or average value) of a random variable and show through several examples how it can be used in gaming situations, economics, and the financial markets.

Consider an experiment in which we roll two dice many times and tally the sums of the numbers on the two dice. What would we “expect” the average of those rolls to be? To get the answer, we carry out a thought experiment. Let $N$ be the number of times we roll the two dice. (We assume that $N$ is very, very large.) Remembering the probability mass density computed in Example 3 of the last section, we expect that the sum 2 would appear about $\frac{1}{36}$ of the time. That is, we expect to roll 2’s approximately $\frac{1}{36}N$ times. Since the sum 3 would appear about $\frac{2}{36}$ of the time, we expect to roll 3’s about $\frac{2}{36}N$ times. Continuing this line of thought, we expect 4’s about $\frac{3}{36}N$ times, 5’s about $\frac{4}{36}N$ times, 6’s about $\frac{5}{36}N$ times, etc. Now, we compute the average of all expected rolls using the usual definition of average (i.e. take the sum of all the rolls and divide by $N$):

$$(\text{Expected}) \text{ Average} = \frac{\frac{1}{36}N(2) + \frac{2}{36}N(3) + \frac{3}{36}N(4) + \ldots + \frac{1}{36}N(12)}{N}$$

$$\quad = \frac{1}{36}(2) + \frac{2}{36}(3) + \frac{3}{36}(4) + \ldots + \frac{1}{36}(12)$$

A remarkable thing happened: the $N$ canceled out of the equation! Thus, the average value of the rolls we expect to get is

$$\frac{1}{36}(2) + \frac{2}{36}(3) + \frac{3}{36}(4) + \ldots + \frac{1}{36}(12) = 7$$

We call 7 the "expected value" of rolling two dice; i.e., if we roll two dice a very large number of times, we would expect the average of the rolls to be 7. (Of course, you may roll two dice 10 billion times, compute the average, and get a number slightly different from 7. Just keep rolling...) The result leads us to an important observation: The average value that we computed is simply the sum of each of the possible values of the random variable (in this case the sum of the faces), multiplied by the probability of that value. In light of this, we make the following general definition:

**Definition.** Let $X$ be a random variable with mass density $p(x)$. Recall that $p(x) = P\{X = x\}$, the probability that the random variable $X$ has the value $x$. The **expectation** (or expected value) of $X$ is denoted by $\mathbb{E}(X)$ and defined as

$$\mathbb{E}(X) = \sum_{\{x:p(x)>0\}} x \cdot p(x). \quad (4)$$

Thus, $\mathbb{E}(X)$ is the weighted average of the possible values of $X$ (those are the $x$’s in the formula) with the weights being the probabilities (those are the $p(x)$’s in the formula). This is exactly the computation that we did in the introductory example. The expectation of $X$ is also called the **mean** or **average value** of $X$.

**Example 1.** Consider the experiment of flipping a coin three times. Let $X$ be the random variable which gives the number of heads obtained in three flips. The possible values of $X$
are 0, 1, 2, and 3. In Example 1 of Probability II, we determined that \( p(0) = \frac{1}{8}, \ p(1) = \frac{3}{8}, \ p(2) = \frac{3}{8}, \text{ and } p(3) = \frac{1}{8} \). Therefore,

\[
\mathbb{E}(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5
\]

Notice that the expectation, \( \mathbb{E}(X) \), need not be one of the values that \( X \) can have. Indeed, you couldn’t flip a coin three times and get 1.5 heads! The fact that \( \mathbb{E}(X) = 1.5 \) says that if we do the experiment of flipping three coins many times, then we would expect about 1.5 heads per experiment. That is, on the average, about half the flips should be heads.

**Example 2.** Let \( Z \) be the payoff for playing the dice game in Example 2 of Probability II. Then, \( Z \) has the possible values 2, 0, \(-1\), and

\[
\mathbb{E}(Z) = (-1) \cdot \frac{1}{2} + 0 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} = \frac{1}{6}.
\]

Thus, on the average, you should expect to win \( \frac{1}{6} \) of a dollar every time you play this game. So, if you play the game 60 times you could expect to be about $10 dollars ahead.

Notice that all you need to compute \( \mathbb{E}(X) \) is the probability mass density of \( X \). If you have (or are given) the probability mass density you don’t need to know the sample space.

**Example 3.** A certain slot machine that costs one dollar to play returns two dollars if you get three cherries and returns 15 dollars if you get three sevens. It gives you nothing for anything else. Suppose that the probability of getting three cherries is \( \frac{1}{4} \) and the probability of getting three sevens is \( \frac{1}{100} \). Since the sum of the probabilities is always one, the probability of “anything else” must be \( \frac{74}{100} \). How much will you win or loss on the average if you play this game? Let \( Z \) be your profit. \( Z \) takes the value \(-1\) with probability \( \frac{74}{100}\), takes the value 1 with probability \( \frac{1}{4} \) and takes the value 14 with probability \( \frac{1}{100} \). Therefore,

\[
\mathbb{E}(Z) = (-1) \cdot \frac{74}{100} + (1) \cdot \frac{25}{100} + (14) \cdot \frac{1}{100} = -\frac{35}{100}
\]

Thus, you will lose on the average 35 cents per time that you play this game.

In the first three examples the random variable had only finitely many possible values. If \( X \) has infinitely many possible values, then formula (4) is an infinite series. In this case the mean, \( \mathbb{E}(X) \), makes sense only if the infinite series has a finite sum. We will study infinite series in detail in 112L.

**Example 4.** In Example 4 of Probability II we considered the experiment of flipping a coin until a head comes up. If we let \( X \) be the number of flips required, then \( X \) can take the values 1, 2, 3, 4, \ldots, so \( X \) has infinitely many possible values. We showed there that

\[
p(n) = \mathbb{P}(X = n) = \left( \frac{1}{2} \right)^n.
\]

Thus, according to the definition (1),

\[
\mathbb{E}(X) = 1 \cdot \frac{1}{2} + 2 \cdot \left( \frac{1}{2} \right)^2 + 3 \cdot \left( \frac{1}{2} \right)^3 + 4 \cdot \left( \frac{1}{2} \right)^4 + \cdots
\]

\[
= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \cdots
\]
Using techniques that we shall see in 112L one can show that this infinite series has a sum of 2. Thus, on the average it should take two flips to get a head, which makes sense.

**Example 5.** An encyclopedia salesman visits three customers each day and with each he has a probability of \( \frac{1}{4} \) of making a sale. For each sale he earns a commission of $100 and if he makes three sales in day he earns a $50 bonus from his company. What should he expect to earn each day? If the random variable \( X \) represents his daily earnings, then \( X \) takes the possible values $0, $100, $200, and $350 with probabilities \( P\{X = 0\} = \frac{27}{64}, \ P\{X = 100\} = \frac{27}{64}, \ P\{X = 200\} = \frac{9}{64}, \ P\{X = 350\} = \frac{1}{64} \) (problem 5 of the last section). Therefore, his expected earning each day is:

\[
E(X) = 0 \cdot \frac{27}{64} + 100 \cdot \frac{27}{64} + 200 \cdot \frac{9}{64} + 350 \cdot \frac{1}{64} = 75.7812
\]

Thus, he can expect to earn on the average $76 each day.

Suppose, now, that the salesman is disgruntled and threatens to quit unless he can earn on the average $90 per day. How should the company set his bonus (for selling 3 in one day) so that his expected earnings will be $90? If the bonus is denoted by \( B \), then his expected earnings will be:

\[
E(X) = 0 \cdot \frac{27}{64} + 100 \cdot \frac{27}{64} + 200 \cdot \frac{9}{64} + (300 + B) \cdot \frac{1}{64}
\]

Setting his expected earnings equal to $90 and solving for \( B \) yields \( B = $960 \). Note that this is a huge bonus but the company is only going to be paying it \( \frac{1}{64} \)th of the time, so on the average the company will be paying him an extra $14 per day approximately. That makes sense since $90 - $76 = $14.

**Example 6.** Let us consider again the change of price of a single share of stock considered in Example 6 of Probability II. We assumed that in each of two years the price would increase by one dollar, decrease by one dollar, or stay the same, each with probability \( \frac{1}{3} \). We computed the probability mass density of \( X \), the profit after two years, so it is easy to calculate the expected profit:

\[
E(X) = \frac{1}{9}(-2) + \frac{2}{9}(-1) + \frac{3}{9}(0) + \frac{2}{9}(1) + \frac{1}{9}(2) = 0.
\]

This makes sense since the probability of going up by one unit each year is the same as the probability of going down by one unit.
Example 7. Here we reconsider Example 7 of Probability II in which the stock price rises by one dollar, decline by one dollar, or stays the same in each year, but the probabilities of these changes are not equal. In problem 7 of Probability II, the student was asked to compute the probability mass density of the profit, $X$, after two years:

\[
P\{X = -2\} = \frac{1}{24} \quad P\{X = -1\} = \frac{4}{24} \quad P\{X = 0\} = \frac{7}{24}
\]

\[
P\{X = 1\} = \frac{8}{24} \quad P\{X = 2\} = \frac{4}{24}
\]

Therefore, the expected profit from holding this one share of this stock for the next two years is:

\[
E(X) = \frac{1}{24}(-2) + \frac{4}{24}(-1) + \frac{7}{24}(0) + \frac{8}{24}(1) + \frac{4}{24}(2) = \frac{10}{24}.
\]

Summary. To compute the expected value $E(X)$ of a random variable $X$, one first computes the probability mass density of the random variable and then uses formula (1). Note that $E(X)$ need not be a possible value of $X$. $E(X)$ is also called the “mean” or “average value” of $X$.

Remark I (expected value is not guaranteed). It is important to keep in mind that $E(X)$ is the average value one would expect after a large number of trials of the experiment (a large number of rolls in Example 2 or a large number of days selling encyclopedias in Example 5). In real life, the actual experiment may be performed only once. In Example 6 the stock price will change over the next two years only in one way. Thus, the meaning of $X$ is that if, hypothetically, one was in this same situation many, many times (like the movie “Groundhog Day”), then we would expect an average return of $E(X)$.

Remark II (how large is “large”). There is a deep mathematical question that we have swept under the rug in this discussion. We keep saying that after a large number of trials of an experiment, the average of values of $X$ which we have obtained should be near $E(X)$. How near? And, how many trials is a “large number”? These are deep questions that are considered in more advanced courses.

Remark III (guessing the future). Notice that our discussions of stock prices and options involved assumptions about the probabilities of different future events. Where do such assumptions (a less polite word would be “guesses”) come from anyway? Well, sometimes they come from analyzing data from the past and sometimes they come from theories about the underlying mechanisms that produce the changes. Mathematicians have made important contributions to the development of such theories about the financial markets.
Homework Problems

1. For each random variable in problems 1 – 4 of Probability II, find \( E(X) \).

2. Consider the following game. We choose a ball at random from an urn containing 7 red, 3 green, and 2 amber balls. We win $2 if we choose a green ball, we lose $1 if we choose a red ball, and we don’t win or lose any money if we choose an amber ball. Let \( X \) represent our winnings. Find the probability mass density of \( X \) and the expected winnings, \( E(X) \), each time we play.

3. A casino offers the following game. You pay $8 to play. Then you roll a fair die twice and the casino pays you the sum of the faces. What is your expected payoff each time you play? Hint: The mass density for the sum of the faces was calculated in Example 3 of Probability II.

4. A casino offers the following game. They have the die from Example 2 of Probability II that has three red faces, two blue faces and one green face. You pay $1 to play. Then you roll the die twice. If both rolls are the same, the casino pays you $2. What is the casino’s expected payoff everytime someone plays?

5. Consider the following game played at a casino. A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears \( i \) times, for \( i = 1, 2, 3 \), then the player wins \( i \) dollars; on the other hand, if the number bet by the player does not appear on any of the dice, then the player loses 1 dollar. Is the game fair? If the game is not fair, who has the advantage, the player or the casino? Hint: Calculate \( E(X) \) where \( X \) equals the number of dollars the player wins or loses after each game.

6. A casino offers the following game. They have the die from Example 2 of Probability II that has three red faces, two blue faces and one green face. You pay $10 to play. Then you roll the die twice. If both rolls are red, the casino pays you $10. If both rolls are blue, the casino pays you $25 and if both rolls are green, the casino pays you $100. What is the expected profit each time you play?

7. Suppose that we have a biased coin which comes up tails \( \frac{5}{8} \) of the time. Consider the experiment of flipping the coin until a head comes up and let \( X \) be the number of flips required. What values can \( X \) take? What is the probability mass density of \( X \). Find an infinite series for the expected number of flips and use your calculator to estimate the sum of the series.

8. You start a small insurance business on the side by providing car insurance for three of your neighbors. Each has a probability of \( \frac{1}{10} \) of having an accident this year (assume no probability of each having more than one). For each accident you will have to pay out $3000. How much should you charge for this coverage so that your expected profit for the year is $1000?

9. Find the expected profit for the purchase of one share of stock in the situation described in problem 8 of Probability II.

10. Find the expected profit for the purchase of one share of stock in the situation described in problem 9 of Probability II.
11. Find the expected profit for the purchase of one share of stock in the situation described in problem 10 of Probability II.

Answers to Selected Problems

1. 0, 1, 3, 0.

3. $-$1.

5. $-$\frac{17}{216}$.

7. $\frac{3}{8}\{1 + 2\left(\frac{5}{8}\right)^2 + 3\left(\frac{5}{8}\right)^3 + \ldots\} = \frac{8}{5}$.

9. $1$.

11. $\frac{7}{3}$. 