14 Probability: Discrete Random Variables

Read the supplementary notes

14.1 Introduction

We will view probability as dealing with repeatable experiments such as flipping a coin, rolling a die or measuring a distance. Anytime there is some uncertainty as to the outcome of an experiment probability has a role to play. Gambling, polling, measuring are typical places where probability is used. In general polling and measuring involve analyzing data which is usually called statistics. In this sense statistics is just the art of applied probability.

14.2 Random outcomes and probability functions

Many ‘experiments’ have a finite number of possible outcomes each with a probability of occurring. It is often possible to list the set of all possible outcomes and give their probabilities.

The fair coin. One of the main examples used in probability is a fair coin. When we toss a fair coin we have the following:

Set of possible outcomes = \{heads, tails\}
Probability of heads = 1/2 = probability of tails.

This seems clear, though it is surprisingly difficult to make physical sense of the notion of probability. One standard interpretation is that if we flip a fair coin many times we expect that close to 1/2 of the flips will land heads.

Naturally there is a notation for probability, e.g.

\[ P(\text{heads}) = \frac{1}{2} \]

Example 14.1. Roll 1 six-sided die. What are the possible outcomes. If the die is fair what is the probability of each of these outcomes.

**answer:** Set of possible outcomes = \{1, 2, 3, 4, 5, 6\}, with \( P(j) = 1/6 \) for \( j = 1, \ldots, 6 \). That is, if we roll a die many times we expect close to 1/6 of the outcomes to be a 1 (or a 2, 3, 4, 5, 6).

Example 14.2. Roll 2 fair dice. Describe the possible outcomes and their probabilities.

**answer:** There are several choices for describing the outcomes.

1. Ordered pairs. Ordered pairs = \{(1,1), (1,2), \ldots, (6,6)\}. This means that we distinguish the dice, e.g. color one red and the other blue, and the outcome (2,5) means the first die lands 2 and the second lands 5. All of these 36 possible outcomes are equally likely so \( P(i, j) = 1/36 \) for any of the pairs.
2. **Total.** We can also describe the outcome of a roll as the total on the two dice. In this case the possible outcomes are:

\[
\text{Outcomes } = \{2, 3, 4, \ldots, 12\}
\]

The outcomes aren’t equally likely, for example we see there is only one way to get a total of two, i.e. roll a (1,1). Therefore \( P(2) = P(1,1) = 1/36 \).

There are two ways to get a total of 3, i.e. (1,2) and (2,1). Since each of these has probability \( 1/36 \) we have \( P(3) = P(1,2) + P(2,1) = 2/36 \). Proceeding in this way we get the following table:

<table>
<thead>
<tr>
<th>Total of two dice: ( x )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
</table>

**Example 14.3.** Suppose that \( p \) is the fraction of voters who support candidate A. Ask a random voter if they support candidate A. Assume everyone answers yes or no. Give the possible outcomes and make a table of outcomes and probabilities:

**answer:** The possible outcomes are yes and no. The table of probabilities is

<table>
<thead>
<tr>
<th>Outcome</th>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( p )</td>
<td>( 1-p )</td>
</tr>
</tbody>
</table>

**Note:** In all the above examples we carefully state how to run the repeatable experiment.

**General terminology.**

Suppose there are \( n \) possible outcomes, call them \( x_1, x_2, \ldots x_n \). Then we can list the outcomes and probabilities in a table

<table>
<thead>
<tr>
<th>Outcome</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( \cdots )</th>
<th>( x_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( P(x_1) )</td>
<td>( P(x_2) )</td>
<td>( \cdots )</td>
<td>( P(x_n) )</td>
</tr>
</tbody>
</table>

There are a few standard names for the things we have been discussing:

- **Sample space** = set of possible outcomes = \( \{x_1, x_2, \ldots, x_n\} \).
- **Probability function**, \( P(x_j) = \) probability of outcome \( x_j \).
- **Trial** = one run of the 'experiment'.
- **Independent**: Two trials are independent if their outcomes have no effect on each other. E.g., repeated flips of a coin are independent.

### 14.3 Probability Laws

There are a rules for probability that we used above without making them explicit. We formulate them first in words and then in symbols:

In words:

(i) Probabilities are between 0 and 1, i.e., the fraction of the time that any one outcome occurs must be between 0 and 1.

(ii) The total probability of all outcomes is 1, i.e. the probability that one of the possible outcomes occurs is 100%.
(iii) The probability that one of several outcomes occurs is just the sum of their probabilities. More precisely, using symbols:
Suppose there are \( n \) possible outcomes and the sample space (set of all possible outcomes) is
\[
sample \ space \ S = \{ x_1, x_2, \ldots, x_n \}
\]
Then the probability function \( P(x_i) \) must satisfy the following three properties
(i) For each outcome \( x_i \) we have \( 0 \leq P(x_i) \leq 1 \) (probability is between 0 and 1)
(ii) \( \sum_{j=1}^{n} P(x_j) = 1 \) (total probability is 1).
(iii) \( P(x_1 \text{ or } x_2 \text{ or } x_3) = P(x_1) + P(x_2) + P(x_3) \) (probabilities add)

Example 14.4. (a) Roll two dice. Let \( A = \text{’the total is < 4’} \). What is \( P(A) \)?
answer: We know \( A = \{2,3\} \), that is, the ‘total is less than 4’ means that the total is either 2 or 3. Using the table for the probabilities of the sum of two dice given in an earlier example we get
\[
P(A) = P(2 \text{ or } 3) = P(2) + P(3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}.
\]
(b) Let \( B = \text{’the total is odd’}, \) so \( B = \{3,5,7,9,11\} \). Find \( P(B) \).
answer: \( P(B) = P(3)+P(5)+P(7)+P(9)+P(11) = \frac{2}{36}+\frac{4}{36}+\frac{6}{36}+\frac{4}{36}+\frac{2}{36} = \frac{1}{2}\).

14.4 Independence and the multiplication law

We say two random events are independent if the outcome of one does not have any effect on the outcome of the other. For example, it is reasonable to assume that the result of the first roll of two dice has no effect on the result of a second roll.

Multiplication law: Suppose we run two independent trials of an experiment and we get the outcome \( x_i \) for the first trial and \( x_j \) for the second, then the probability of the combined result is
\[
P(x_i \text{ on the first trial and } x_j \text{ on the second trial}) = P(x_i) \cdot P(x_j).
\]

Example 14.5. It is reasonable to assume that different tosses of a coin are independent. Suppose we toss a fair coin twice. Describe the outcomes and give their probabilities.
answer: The following notation for the outcomes is common:
\( HH \): heads on both tosses
\( HT \): heads on the first and tails on the second toss
\( TH \): tails on the first toss and heads on the second
\( TT \): tails on both tosses.

Using the multiplication law we get
\[
P(HH) = P(H)P(H) = \frac{1}{4}; \quad P(HT) = P(H)P(T) = \frac{1}{4}; \quad P(TH) = \frac{1}{4}; \quad P(TT) = \frac{1}{4}.
\]
That is, all 4 outcomes have equal probability.
Example 14.6. Toss a fair die 3 times, what is the probability of getting and odd number each time?

**answer:** Let \( A = \{1, 3, 5\} \). On any one toss \( P(A) = 1/2 \). Since repeated tosses are independent \( P(A \text{ then } A \text{ then } A) = P(A) \cdot P(A) \cdot P(A) = 1/8 \).

### 14.5 Discrete random variables

When the outcomes of an experiment are numbers we have a random variable. More precisely: a **finite random variable** \( X \) consists of

(i) A finite list \( x_1, x_2 \ldots, x_n \) of values \( X \) can take.

(ii) A probability function \( P(x_j) \)

**Examples:**

(i) Roll a die, let \( X = \) number of spots.

(ii) Roll a die, let \( Y = (\text{number of spots})^2 \).

(iii) Toss a coin, let \( X = 1 \) if the result is heads and \( X = 0 \) if the result is tails.

There is no reason a random variable has to take only a finite number of values. If it has an infinite number of values that can be listed we call it an **infinite discrete random variable**. That is, \( X \) is an infinite discrete random variable if:

(i) \( X \) takes values \( x_1, x_2, \ldots \)

(ii) There is a probability function \( P(x_i) \) that satisfies the probability laws given above.

Later we will look at so called continuous random variables. These take values in an entire interval, e.g. \([0, 1]\) or \([0, \infty)\).

### 14.6 Expectation

The **expectation**, (also called **average**, **mean** or **expected value**) of the finite random variable \( X \) is defined by

\[
E(X) = x_1 P(x_1) + \ldots + x_n P(x_n) = \sum_{i=1}^{n} x_i P(x_i).
\]

**Example 14.7.** Roll a die. Define the random variable \( X = \) number of spots. Compute the expected value of \( X \).

**answer:** \( E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5 \).

**Interpretation of expected value**

1. The expected value is the expected average over a lot of trials. To understand this, suppose I roll a die, and I pay you $1 per spot then over a lot of rolls, say 6000, I would expect to roll (about) 1000 ones, 1000 twos, etc. In that case I would pay you a total of

\[
1000 \cdot 1 + 1000 \cdot 2 + 1000 \cdot 3 + 1000 \cdot 4 + 1000 \cdot 5 + 1000 \cdot 6 = $3,500
\]

So the average payment per roll would be

\[
\frac{1000 \cdot 1 + 1000 \cdot 2 + 1000 \cdot 3 + 1000 \cdot 4 + 1000 \cdot 5 + 1000 \cdot 6}{6000} = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{1}{6} = 3.5
\]
Notice the formula for the average above is exactly the formula we used to define expected value of a random variable.

Note, you would never be paid $3.5 on any one turn, it is the expected average over many turns.

Also note, expectation is a weighted average like center of mass.

**Concept question:** If I pay you $1 per spot how much would you be willing to pay to roll the die?

**Example 14.8.** Roll a die, you win $5 if it’s a 1 and lose $2 if it’s not. Model this with a random variable and its expected value.

**answer:** In detail here are the steps we need.

Define the random variable: Let \( X \) the amount you win or lose on a given roll. That is, \( X \) takes the values 5, -2.

Give the probability function:
\[
P(X = 5) = P(\text{roll a 1}) = \frac{1}{6}, \quad P(X = -2) = P(\text{roll 2 to 6}) = \frac{5}{6}.
\]

(Or in a table:)

<table>
<thead>
<tr>
<th>Payment</th>
<th>5</th>
<th>-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(x) )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{5}{6} )</td>
</tr>
</tbody>
</table>

Compute the expected value:
\[
E(X) = P(5) \cdot 5 + P(-2) \cdot (-2) = \frac{5}{6} - \frac{10}{6} = -\frac{5}{6}.
\]

**Concept question:** Is the above bet a good one?

**Example 14.9.** (From supplementary notes) A trial consists of tossing a fair coin until it comes up heads. Let \( X = \) number of tosses. Verify that the total probability of all outcomes is 1.

**answer:** First we show all the outcomes and their probabilities in a table.

<table>
<thead>
<tr>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>toss pattern</td>
<td>H</td>
<td>TH</td>
<td>TTH</td>
<td>TTH</td>
<td>...</td>
</tr>
<tr>
<td>( P(X = n) )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{16} )</td>
<td>...</td>
</tr>
</tbody>
</table>

So, the total probability = \( \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1 \). (This is a geometric series.)

**Example 14.10.** If I paid you $n for a trial of length \( n \), what would you pay to take a turn?

**answer:** We compute the expectation.

\[
E(X) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \ldots = 2.
\]

(The supplementary notes give a nice method for finding this sum.) Assuming you can play the game many times, for any price less than $2 you would expect to win a positive amount on average.
14.7 Poisson random variable with parameter $m$

There are lots of random variables that model various situations. One of them is the Poisson random variable. To describe it we need to give the possible values, their probabilities, and we should say what it models.

Let $X$ be a Poisson random variable with parameter $m$.

(i) Values: $X$ takes values 0, 1, 2, 3, 

(ii) Probability: $P(X = k) = e^{-m} \frac{m^k}{k!}$. (The factor $e^{-m}$ is chosen to give total probability 1.)

(iii) Model: A Poisson random variable models the number of times a low probability event occurs in a given time interval.

Examples: A Poisson random variable can be used to model the following counts.

(i) Number of defects in a manufacturing process in a day.
(ii) Number of errors in data transmission.
(iii) Number of cars in an hour at a rural tollbooth.
(iv) Number of chocolate chips in a cookie.

Theorem. $E(X) = m$.

Proof.

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} k e^{-m} \frac{m^k}{k!}$$

$$= e^{-m} \left( m \frac{m}{1!} + \frac{m^2}{2!} + \ldots \right)$$

$$= m e^{-m} \left( 1 + \frac{m}{1!} + \frac{m^2}{2!} + \ldots \right)$$

$$= m e^{-m} e^m$$

$$= m \quad \text{QED.}$$

(To get from the second to last line to the last line, you need to remember the Taylor series for $e^m$.)

Example 14.11. A manufacturer of widgets knows that 1/200 will be defective.

(a) What is the probability that a box of 144 contains no defective widgets?

answer: Since a single widget has a 1/200 chance of being defective, the expected number of defective widgets in a box is 144/200. Since being defective is rare we can model the number of defective widgets in a box as a Poisson random variable $X$ with mean $m = 144/200$.

The problem asks us to find $P(X = 0)$. We know this is $P(X = 0) = e^{-m} \frac{m^0}{0!} = e^{-144/200} = 0.487$.

(b) What is the probability that more than 2 are defective?

answer: This problem asks us to find $P(X > 2)$. We do this using the formula $P(X > 2) = 1 - P(X \leq 2)$.
\[ 2) = 1 - P(X = 0, 1, 2). \]

\[ 1 - P(X = 0, 1, 2) = 1 - e^{-m} \left( 1 + m + \frac{m^2}{2!} \right) = 0.037. \]

### 14.8 Histograms

If we ran many trials and made a histogram of the percentage of each outcome, the result should look like the graph of \( P(x_j) \) vs. \( x_j \).

Or we make a histogram of the count for each outcome.

The histograms at right give examples.

![Histogram of percentages (Poisson distr. with \( \mu = 4.5 \))](image)