Random Variables

A quantitative variable $x$ is a random variable if the value that $x$ takes on in a given experiment or observation is a chance or random outcome.

A discrete random variable can take on only a finite number of values or a countable number of values.

A continuous random variable can take on any of the countless number of values in a line interval.

The distinction between discrete and continuous random variables is important because of the different mathematical techniques associated with the two kinds of random variables.
Random Variables

Most of the continuous random variables we will see will occur as the result of a measurement on a continuous scale.

For example, the air pressure in an automobile tire represents a continuous random variable. The air pressure could, in theory, take on any value from 0 lb/in$^2$ (psi) to the bursting pressure of the tire.

Values such as 20.126 psi, 20.12678 psi, and so forth are possible.

Probability Distribution of a Discrete Random Variable

A random variable has a probability distribution whether it is discrete or continuous.

A probability distribution is an assignment of probabilities to each distinct value of a discrete random variable or to each interval of values of a continuous random variable.

Features of the probability distribution of a discrete random variable

1. The probability distribution has a probability assigned to each distinct value of the random variable.
2. The sum of all the assigned probabilities must be 1.

Example 1 – Discrete probability distribution

Dr. Mendoza developed a test to measure boredom tolerance. He administered it to a group of 20,000 adults between the ages of 25 and 35.

The possible scores were 0, 1, 2, 3, 4, 5, and 6, with 6 indicating the highest tolerance for boredom. The test results for this group are shown in Table 5-1.

<table>
<thead>
<tr>
<th>Score</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1400</td>
</tr>
<tr>
<td>1</td>
<td>2800</td>
</tr>
<tr>
<td>2</td>
<td>3600</td>
</tr>
<tr>
<td>3</td>
<td>6000</td>
</tr>
<tr>
<td>4</td>
<td>4400</td>
</tr>
<tr>
<td>5</td>
<td>1600</td>
</tr>
<tr>
<td>6</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 5-1

Statistical Literacy: Which of the following are continuous variables, and which are discrete?

(a) Speed of an airplane
(b) Age of a college professor chosen at random
(c) Number of books in the college bookstore
(d) Weight of a football player chosen at random
(e) Number of lightning strikes in Rocky Mountain National Park on a given day

Statistical Literacy: Consider each distribution. Determine if it is a valid probability distribution or not, and explain your answer.

(a) $x$ 0 1 2
    $P(x)$ 0.25 0.60 0.15
(b) $x$ 0 1 2
    $P(x)$ 0.25 0.60 0.10
Example 1 – *Discrete probability distribution*  

**a.** If a subject is chosen at random from this group, the probability that he or she will have a score of 3 is $6000/20,000$, or $0.30$. In a similar way, we can use relative frequencies to compute the probabilities for the other scores (Table 5-2).

<table>
<thead>
<tr>
<th>Score x</th>
<th>Probability P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.08</td>
</tr>
<tr>
<td>6</td>
<td>0.01</td>
</tr>
</tbody>
</table>

| $\sum P(x) = 1$ |

Probability Distribution of Scores on Boredom Tolerance Test  

Table 5-2

Example 1 – *Discrete probability distribution*  

**c.** The Topnotch Clothing Company needs to hire someone with a score on the boredom tolerance test of 5 or 6 to operate the fabric press machine.

Since the scores 5 and 6 are mutually exclusive, the probability that someone in the group who took the boredom tolerance test made either a 5 or a 6 is the sum

$$P(5 \text{ or } 6) = P(5) + P(6)$$

$$= 0.08 + 0.02 = 0.10$$

Example 1 – *Discrete probability distribution*  

**These probability assignments make up the probability distribution.**

Notice that the scores are mutually exclusive: No one subject has two scores.

The sum of the probabilities of all the scores is 1.

---

Probability Distribution of a Discrete Random Variable

A probability distribution can be thought of as a relative-frequency distribution based on a very large $n$.

As such, it has a mean and standard deviation. If we are referring to the probability distribution of a *population*, then we use the Greek letters $\mu$ for the mean and $\sigma$ for the standard deviation.

When we see the Greek letters used, we know the information given is from the *entire population* rather than just a sample.
Probability Distribution of a Discrete Random Variable

If we have a sample probability distribution, we use $\bar{x}$ (x bar) and $s$, respectively, for the mean and standard deviation.

The mean and standard deviation of a discrete population probability distribution are found by using these formulas:

$$
\mu = \sum xP(x), \mu \text{ is called the expected value of } x
$$

$$
\sigma = \sqrt{\sum (x - \mu)^2P(x)}, \sigma \text{ is called the standard deviation of } x
$$

where $x$ is the value of a random variable,

$P(x)$ is the probability of that variable, and

the sum $\sum$ is taken for all the values of the random variable.

Note: $\mu$ is the population mean and $\sigma$ is the underlying population standard deviation because the sum $\sum$ is taken over all values of the random variable (i.e., the entire sample space).

Example 2 – Expected value, standard deviation

Are we influenced to buy a product by an ad we saw on TV?

National Infomercial Marketing Association determined the number of times buyers of a product had watched a TV infomercial before purchasing the product. The results are shown here:

<table>
<thead>
<tr>
<th>Number of Times Buyers Saw Infomercial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Buyers</td>
<td>27%</td>
<td>51%</td>
<td>18%</td>
<td>9%</td>
<td>19%</td>
</tr>
</tbody>
</table>

*This category includes those who have not purchased the product. Which is different from the expected value $\mu$. The mean of a probability distribution is often called the expected value of the distribution.

This terminology reflects the idea that the mean represents a “central point” or “cluster point” for the entire distribution. Of course, the mean or expected value is an average value, and as such, it need not be a point of the sample space.

The standard deviation is often represented as a measure of risk. A larger standard deviation implies a greater likelihood that the random variable $x$ is different from the expected value $\mu$.

We can treat the information shown as an estimate of the probability distribution because the events are mutually exclusive and the sum of the percentages is 100%.

Compute the mean and standard deviation of the distribution.
Example 2 – Solution

We put the data in the first two columns of a computation table and then fill in the other entries (see Table 5-5).

<table>
<thead>
<tr>
<th>( x ) (number of viewings)</th>
<th>( f(x) )</th>
<th>( x - \mu )</th>
<th>( (x - \mu)^2 )</th>
<th>( x - \mu )^2 ( f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.27</td>
<td>-1.54</td>
<td>2.372</td>
<td>0.0640</td>
</tr>
<tr>
<td>2</td>
<td>0.31</td>
<td>0.46</td>
<td>0.212</td>
<td>0.038</td>
</tr>
<tr>
<td>3</td>
<td>0.18</td>
<td>0.46</td>
<td>0.212</td>
<td>0.038</td>
</tr>
<tr>
<td>4</td>
<td>0.09</td>
<td>1.46</td>
<td>2.152</td>
<td>0.192</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
<td>1.46</td>
<td>2.152</td>
<td>0.192</td>
</tr>
</tbody>
</table>

\[ \mu = \Sigma f(x) = 2.54 \text{ (sum of column 3)} \]

To find the standard deviation, we take the square root of the sum of column 6:

\[ \sigma = \sqrt{\Sigma (x - \mu)^2 f(x)} = \sqrt{1.869} = 1.37 \]

Example 2 – Solution (cont’d)

The average number of times a buyer views the infomercial before purchase is

\[ \mu = \Sigma x f(x) = 2.54 \text{ (sum of column 3)} \]

8. Basic Computation: Expected Value For a fundraiser, 1,000 raffle tickets are sold and the winner is chosen at random. There is only one prize, $500 in cash. You buy one ticket.
(a) What is the probability you will win the prize of $500?
(b) Your expected earnings can be found by multiplying the value of the prize by the probability you will win the prize. What are your expected earnings?
(c) Interpretation If a ticket costs $2, what is the difference between your “cost” and “expected earnings”? How much are you effectively contributing to the fundraiser?
This problem is typical of an entire class of problems that are characterized by the feature that there are exactly two possible outcomes (for each trial) of interest. These problems are called binomial experiments, or Bernoulli experiments, after the Swiss mathematician Jacob Bernoulli, who studied them extensively in the late 1600s.

Features of a binomial experiment
1. There is a fixed number of trials. We denote this number by the letter n.
2. The n trials are independent and repeated under identical conditions.
3. Each trial has only two outcomes: success, denoted by S, and failure, denoted by F.
4. For each individual trial, the probability of success is the same. We denote the probability of success by p and that of failure by q. Since each trial results in either success or failure, \( p + q = 1 \) and \( q = 1 - p \).
5. The central problem of a binomial experiment is to find the probability of \( r \) successes out of \( n \) trials.

Computing Probabilities for a Binomial Experiment Using the Binomial Distribution Formula

The central problem of a binomial experiment is finding the probability of \( r \) successes out of \( n \) trials. Now we’ll see how to find these probabilities.

Suppose you are taking a timed final exam. You have three multiple-choice questions left to do. Each question has four suggested answers, and only one of the answers is correct.

You have only 5 seconds left to do these three questions, so you decide to mark answers on the answer sheet without even reading the questions.

Computing Probabilities for a Binomial Experiment Using the Binomial Distribution Formula

Assuming that your answers are randomly selected, what is the probability that you get zero, one, two, or all three questions correct?

This is a binomial experiment. Each question can be thought of as a trial, so there are \( n = 3 \) trials. The possible outcomes on each trial are success \( S \), indicating a correct response, or failure \( F \), meaning a wrong answer.

The trials are independent—the outcome of any one trial does not affect the outcome of the others.

What is the probability of success on anyone question? Since you are guessing and there are four answers from which to select, the probability of a correct answer is 0.25.

The probability \( q \) of a wrong answer is then 0.75. In short, we have a binomial experiment with \( n = 3 \), \( p = 0.25 \), and \( q = 0.75 \).

Now, what are the possible outcomes in terms of success or failure for these three trials? Let’s use the notation \( SSF \) to mean success on the first question, success on the second, and failure on the third.
There are eight possible combinations of $S$s and $F$s. They are

$$SSS \quadSSF \quad SFS \quad SFF \quad FSF \quad FFS \quad FFF$$

To compute the probability of each outcome, we can use the multiplication law because the trials are independent. For instance, the probability of success on the first two questions and failure on the last is

$$P(SSF) = P(S) \cdot P(S) \cdot P(F) = p \cdot p \cdot q = p^2q$$

$$= (0.25)^2(0.75) \approx 0.047$$

Now we can compute the probability of $r$ successes out of three trials for $r = 0, 1, 2$ or 3. Let's compute $P(1)$. The notation $P(1)$ stands for the probability of one success. For three trials, there are three different outcomes that show exactly one success. They are the outcomes $SFF$, $FSF$, and $FFS$.

$$P(F) = P(SFF) + P(FSF) + P(FFS)$$

$$= 3pq^2$$

$$= 3(0.25)(0.75)^2$$

$$= 0.422$$
Computing Probabilities for a Binomial Experiment Using the Binomial Distribution Formula

In the same way, we can find \( P(0) \), \( P(2) \), and \( P(3) \). These values are shown in Table 5.9.

<table>
<thead>
<tr>
<th>( r ) (number of successes)</th>
<th>( P(r) ) (probability of ( r ) successes in 3 trials)</th>
<th>( P(r) ) for ( p = 0.25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( P(0) = P(000) = q^3 )</td>
<td>0.422</td>
</tr>
<tr>
<td>1</td>
<td>( P(1) = P(001) + P(010) + P(100) = 3pq^2 )</td>
<td>0.422</td>
</tr>
<tr>
<td>2</td>
<td>( P(2) = P(011) + P(101) + P(110) = 3p^2q )</td>
<td>0.141</td>
</tr>
<tr>
<td>3</td>
<td>( P(3) = P(111) )</td>
<td>0.016</td>
</tr>
</tbody>
</table>

| \( P(r) \) for \( n = 3 \) Trials, \( p = 0.25 \) |

Table 5.9

Example 5 – Compute \( P(r) \) using the binomial distribution formula

Privacy is a concern for many users of the Internet. One survey showed that 59% of Internet users are somewhat concerned about the confidentiality of their e-mail. Based on this information, what is the probability that for a random sample of 10 Internet users, 6 are concerned about the privacy of their e-mail?

Solution:

a. This is a binomial experiment with 10 trials. If we assign success to an Internet user being concerned about the privacy of e-mail, the probability of success is 59%. We are interested in the probability of 6 successes.

By the formula,

\[
P(6) = C_{10,6}(0.59)^6(0.41)^{10-6}
\]

We have

\[
 n = 10 \quad p = 0.59 \quad q = 0.41 \quad r = 6
\]

Use Table 2 of Appendix II or a calculator.

Use a calculator.

\[
\approx 210(0.0542)(0.0283)
\]

\[
\approx 0.25
\]

There is a 25% chance that exactly 6 of the 10 Internet users are concerned about the privacy of e-mail.
Using a Binomial Distribution Table

In many cases we will be interested in the probability of a range of successes. In such cases, we need to use the addition rule for mutually exclusive events. For instance, for \( n = 6 \) and \( p = 0.50 \),

\[
P(4 \text{ or fewer successes}) = P(r \leq 4) = P(r = 4 \text{ or } 3 \text{ or } 2 \text{ or } 1 \text{ or } 0) = P(4) + P(3) + P(2) + P(1) + P(0)
\]

Using a Binomial Distribution Table

Table 5-10 is an excerpt from Table 3 of Appendix II showing the section for \( n = 6 \). Notice that all possible \( r \) values between 0 and 6 are given as row headers.

<table>
<thead>
<tr>
<th>( r )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(r) )</td>
<td>0.094</td>
<td>0.234</td>
<td>0.312</td>
<td>0.234</td>
<td>0.094</td>
<td>0.016</td>
</tr>
</tbody>
</table>

The value \( p = 0.50 \) is one of the column headers. For \( n = 6 \) and \( p = 0.50 \), you can find the value of \( P(4) \) by looking at the entry in the row headed by 4 and the column headed by 0.50. Notice that \( P(4) = 0.234 \).

Using a Binomial Distribution Table

Likewise, you can find other values of \( P(r) \) from the table. In fact, for \( n = 6 \) and \( p = 0.50 \),

\[
P(r \leq 4) = P(4) + P(3) + P(2) + P(1) + P(0) = 0.234 + 0.312 + 0.234 + 0.094 + 0.016 = 0.890
\]

Alternatively, to compute \( P(r \leq 4) \) for \( n = 6 \), you can use the fact that the total of all \( P(r) \) values for \( r \) between 0 and 6 is 1 and the complement rule.

Using a Binomial Distribution Table

Since the complement of the event \( r \leq 4 \) is the event \( r \geq 5 \), we have

\[
P(r \leq 4) = 1 - P(5) - P(6) = 1 - 0.094 - 0.016 = 0.890
\]
Example 6 – Using the binomial distribution table to find $P(r)$

A biologist is studying a new hybrid tomato. It is known that the seeds of this hybrid tomato have probability 0.70 of germinating. The biologist plants six seeds.

**a.** What is the probability that exactly four seeds will germinate?

**Solution:**
This is a binomial experiment with $n = 6$ trials. Each seed planted represents an independent trial.

We’ll say germination is success, so the probability for success on each trial is 0.70.

Example 6 – Solution

$n = 6 \quad p = 0.70 \quad q = 0.30 \quad r = 4$

We wish to find $P(4)$, the probability of exactly four successes. In Table 3, Appendix II, find the section with $n = 6$ (excerpt is given in Table 5-10).

Then find the entry in the column headed by $p = 0.70$ and the row headed by $r = 4$.

This entry is 0.324.

$P(4) = 0.324$

**b.** What is the probability that at least four seeds will germinate?

**Solution:**
In this case, we are interested in the probability of four or more seeds germinating. This means we are to compute $P(r \geq 4)$. Since the events are mutually exclusive, we can use the addition rule

\[
P(r \geq 4) = P(r = 4 \text{ or } r = 5 \text{ or } r = 6) = P(4) + P(5) + P(6)
\]

We already know the value of $P(4)$. We need to find $P(5)$ and $P(6)$.

Example 6 – Using the binomial distribution table to find $P(r)$

cont'd
Example 6 – Solution

Use the same part of the table but find the entries in the row headed by the \( r \) value 5 and then the \( r \) value 6. Be sure to use the column headed by the value of \( p \), 0.70.

\[
P(5) = 0.303 \quad \text{and} \quad P(6) = 0.118
\]

Now we have all the parts necessary to compute \( P(r \geq 4) \).

\[
P(r \geq 4) = P(4) + P(5) + P(6)
\]

\[
= 0.324 + 0.303 + 0.118
\]

\[
= 0.745
\]

Example 7 – Compute \( \mu \) and \( \sigma \)

Let’s compute the mean and standard deviation for the distribution of Example 7 that describes that probabilities of lone diners leaving tips at the Green Spot Restaurant.

Solution:
In Example 7,
\( n = 6 \quad \quad p = 0.7 \quad \quad q = 0.3 \)

For the binomial distribution,
\( \mu = np = 6(0.7) = 4.2 \)

Example 8 – Compute \( \mu \) and \( \sigma \)

Mean and Standard Deviation of a Binomial Distribution

**Procedure:**

<p>| ( \mu ) is the expected number of successes for the random variable ( r ) |</p>
<table>
<thead>
<tr>
<th>( \sigma ) is the standard deviation for the random variable ( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu = np )</td>
</tr>
<tr>
<td>( \sigma = \sqrt{npq} )</td>
</tr>
<tr>
<td>where</td>
</tr>
<tr>
<td>( r ) is a random variable representing the number of successes in a binomial distribution,</td>
</tr>
<tr>
<td>( n ) is the number of trials,</td>
</tr>
<tr>
<td>( p ) is the probability of success on a single trial, and</td>
</tr>
<tr>
<td>( q = 1 - p ) is the probability of failure on a single trial.</td>
</tr>
</tbody>
</table>

Example 8 – Solution

The balance point of the distribution is at \( \mu = 4.2 \).

The standard deviation is given by

\[
\sigma = \sqrt{npq}
\]

\[
= \sqrt{6(0.7)(0.3)}
\]

\[
= \sqrt{1.26}
\]

\[
\approx 1.12
\]
Insurance: Auto  State Farm Insurance studies show that in Colorado, 55% of the auto insurance claims submitted for property damage are submitted by males under 25 years of age. Suppose 10 property damage claims involving automobiles are selected at random.
(a) Let \( r \) be the number of claims made by males under age 25. Make a histogram for the \( r \)-distribution probabilities.
(b) What is the probability that six or more claims are made by males under age 25?
(c) What is the expected number of claims made by males under age 25? What is the standard deviation of the \( r \)-probability distribution?

(a) The probabilities can be taken directly from Table 3 in Appendix II. \( n = 10, p = 0.55. \)

(b) \( P(r \geq 6) = P(6) + P(7) + P(8) + P(9) + P(10) = 0.504 \)
(c) \( \mu = np = 10(0.55) = 5.5 \)
The expected number of claims made by males under age 25 is 5.5.
\( \sigma = \sqrt{npq} = \sqrt{10(0.55)(0.45)} \approx 1.57 \)