Random variable (RV)—a variable that assumes numerical values associated with the random outcomes of an experiment, where only one numerical value is assigned to each sample point

- Discrete RVs—random variables that can assume a countable number of values
- Continuous RVs—random variables that can assume values corresponding to any of the points contained in one or more intervals

### Chapter 5—Discrete Probability Distributions

<table>
<thead>
<tr>
<th>$x$</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x)$</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

### Random Variables

Identify the following RVs as discrete or continuous
1. The diameter of a tree
2. The number of chapters in your statistics textbook
3. Number of commercials during your favorite TV show
4. The length of the first commercial shown during your favorite TV show
5. The number of registered voters who vote in a national election

### Expectations for RVs

The expected value ($EV$) of a RV is the mean value of the variable $X$ in the sample space, or population of possible outcomes.

$EV$ can be interpreted as the mean value that would be obtained from an infinite number of observations of the random variable.

### Probability Distributions for Discrete RVs

A complete description of a discrete RV requires specification of
- The possible values that the RV can assume
- The probability associated with each value

The probability distribution of a discrete RV, $X$, can be represented by a graph, table, or formula that specifies the probabilities associated with each possible value of $x$.

#### Requirements
1. $0 \leq P(X = x) \leq 1$ for any value of $x$
2. $\sum_{x} P(X = x) = 1$

### Example

Consider the following probability distribution function (pdf):

<table>
<thead>
<tr>
<th>$x$</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x)$</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Cumulative Distribution Function (cdf)

<table>
<thead>
<tr>
<th>$x$</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X \leq x)$</td>
<td>0.2</td>
<td>0.5</td>
<td>0.7</td>
<td>0.8</td>
<td>1.0</td>
</tr>
</tbody>
</table>
The mean, or expected value, of a discrete RV is determined by its probability distribution

\[ \mu = \sum xP(x) \]

The variance of a discrete RV is

\[ \sigma^2 = \sum (x - \mu)^2 P(x) \]

Calculator formula:

\[ \sigma^2 = \sum [x_i^2 \times P(x_i)] - \mu^2 \]

The standard deviation of a discrete RV is equal to the square root of the variance:

\[ \sigma = \sqrt{\sigma^2} \]

This quantifies how spread out the possible values of a discrete RV might be, weighted by how likely each value is to occur.


Given the probability distribution of \( Y \), find the mean, variance and SD for \( Y \)

\[ E(Y) = \sum yP(y) \]

\[ \sigma = \sqrt{\sum (y - E(Y))^2 P(y)} \]

\[ \sigma = \sqrt{\sum [y_i^2 \times P(y_i)] - \mu^2} \]

Based on past results found in the Information Please Almanac, there is a 0.1818 probability that a baseball World Series contest will last four games, a 0.2121 probability that it will last five games, a 0.2323 probability that it will last six games, and a 0.3737 probability that it will last seven games. Is it unusual for a team to “sweep” by winning in four games?

Let \( Y = \) number of games played in a World Series

\[ P(Y = y) \]

Range rule of thumb: \( \mu \pm 2\sigma \)

- Unusually high number of successes: \( P(y \text{ or more}) \leq 0.05 \)
- Unusually low number of successes: \( P(y \text{ or fewer}) \leq 0.05 \)

Example

Based on past results found in the Information Please Almanac, there is a 0.1818 probability that a baseball World Series contest will last four games, a 0.2121 probability that it will last five games, a 0.2323 probability that it will last six games, and a 0.3737 probability that it will last seven games. Is it unusual for a team to “sweep” by winning in four games?
Characteristics of a binomial experiment

1. The experiment consists of \( n \) identical (fixed) trials
2. The trials are independent
3. The experiment results in a dichotomous response; i.e., there are only two possible outcomes on each trial. One outcome is denoted by \( S \) (success) and the other by \( F \) (failure)
4. The probability of \( S \), denoted as \( p \), remains the same from trial to trial. The probability of \( F \), denoted as \( q \), is equal to \( 1 - p \).

The binomial random variable, \( X \), is the number of \( S \)'s in \( n \) trials

Example

Record the sequence of heads and tails in 3 tosses of an unfair coin where the \( P(H) = .6 \) and the \( P(T) = .4 \). We are interested in the distribution of the number of tails.

What is \( n \), the number of trials?
Are the trials identical?
Are the trials independent?
What is \( S \)?
What is \( X \)?

Example (continued)

How many possible outcomes are there?
Number of outcomes = \( 2^n = 8 \)
Possible outcomes are:

<table>
<thead>
<tr>
<th>( X )</th>
<th>( P(X) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.216</td>
</tr>
<tr>
<td>1</td>
<td>0.432</td>
</tr>
<tr>
<td>2</td>
<td>0.288</td>
</tr>
<tr>
<td>3</td>
<td>0.064</td>
</tr>
</tbody>
</table>

The Binomial Probability Distribution

\[
P(X = x) = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, 2, \ldots, n)
\]

where
- \( p \) = probability of success on a single trial
- \( q = 1 - p \)
- \( n \) = number of trials
- \( x \) = number of successes in \( n \) trials

Mean, Variance and SD for a Binomial RV

Mean: \( \mu = np \)
Variance: \( \sigma^2 = npq \)
SD: \( \sigma = \sqrt{npq} \)

The Binomial Probability Distribution

For the tossing three coins example we can calculate various quantities:

\[
P(X = 2) = \binom{3}{2} .6^2 .4^1 = 3(.36)(.4) = .288
\]

\[
\mu_x = np = 3(.4) = 1.2
\]

\[
\sigma_x = \sqrt{npq} = \sqrt{3(.4)(.6)} = .72
\]
Generic Example: \( n = 6; p = 0.4 \)

Can calculate various quantities:

\[
P(X = 0) = \binom{6}{0} \cdot (0.4)^0 \cdot (0.6)^6 = 0.087
\]

\[
P(X = 1) = \binom{6}{1} \cdot (0.4)^1 \cdot (0.6)^5 = 0.338
\]

\[
P(X = 2) = \binom{6}{2} \cdot (0.4)^2 \cdot (0.6)^4 = 0.366
\]

Can also calculate cumulative probabilities:

\[
P(X \leq 3) = P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0) = 0.8208
\]

\[
P(X > 2) = 1 - P(X \leq 2) = 0.5443
\]

Example


A Federal Trade Commission (FTC) study of the pricing accuracy of electronic checkout scanners at stores found that one of every 30 items is priced incorrectly. Suppose the FTC randomly selects five items at a retail store and checks the accuracy of the scanner price of each. Let \( X \) represent the number of the five items that is priced incorrectly.

a) Show that \( X \) is a binomial RV.

b) Use the information in the FTC study to estimate \( p \) for the binomial experiment.

c) What is the probability that exactly one of the five items is priced incorrectly by the scanner?

d) What is the probability that at least one of the five items is priced incorrectly by the scanner?

e) What is the probability that \( X \) is in the interval \( \mu \pm 2\sigma \)?

Example

Consider the discrete probability distribution:

\[
\begin{array}{cccc}
X & 10 & 12 & 18 & 20 \\
P(X) & 2 & 3 & 1 & 4 \\
\end{array}
\]

Calculate \( \mu \), \( \sigma \), and \( \sigma \).

What is \( P(X < 15) \)?

Calculate \( \mu \pm 2\sigma \).

What is the probability that \( X \) is in the interval \( \mu \pm 2\sigma \)?