Chapter 8

Binomial and Geometric Distributions
Lesson 8-1, Part 1

Binomial Distribution
What is a Binomial Distribution?

- Specific type of discrete probability distribution
- The outcomes belong to **two** categories
  - pass or fail
  - acceptable or defective
  - success or failure
Example 1 – Cereal

Suppose a cereal manufacturer puts pictures of famous athletes on cards in boxes of cereal, in the hope of increasing sales. The manufacture announces that 20% of the boxes contain a picture of Tiger Woods, 30% a picture of Lance Armstrong, and the rest a picture of Serena Williams.

You buy 5 boxes of cereal. What’s the probability you get exactly 2 pictures of Tiger Woods?
Requirements for a Binomial Distribution

- There is a \textbf{fixed} number \((n)\) of trials.
- Trials are \textbf{independent}.
  - Outcome of any individual trial doesn’t affect the probabilities in the other trial.
- Outcomes are classified into \textbf{two categories}.
  - Success or failure.
- The probability of success \((p)\) is the \textbf{same} for each trial.
If $X$ is a binomial random variable, it is said to have a **binomial distribution**

- $X$ = number of success
  - Whole numbers from 0 to $n$
- Is denoted as $B(n, p)$
  - $n$ is the number of trials
  - $p$ is the probability of a success on any one observation

- The **probability distribution function** (or p.d.f) assigns a probability to each value of $X$.
- The **cumulative distribution function** (or c.d.f) calculates the sum of probabilities up to $X$. 

Methods for Finding Probabilities of a Binomial Distribution

- Using the Binomial Probability Formula
- Using the TI-83
TI – Binomial Probability

○ Computing **exact** probabilities
  - 2nd/Vars/Binompdf
    - binompdf(n, p, x)
  - pdf: probability distribution function

○ Computing **less than or equal** to probabilities
  - 2nd/Vars/binomcdf
    - binomcdf(n, p, x)
  - cdf: cumulative distribution function
There is a mathematical way to count the total number of ways to arrange $k$ out of $n$ objects. This is called “$n$ choose $k$” or binomial coefficient.

\[
\binom{n}{k} = n \binom{k}{n} \quad \text{and is called “$n$ choose $k$” is given by the formula}
\]

\[
\binom{n}{k} = n \binom{k}{n} = \frac{n!}{k!(n-k)!}
\]
Binomial Formula

\[ n = \text{number of trials} \]
\[ p = \text{probability of success and } q = 1 - p \text{ for failures} \]
\[ X = \text{number of success in } n \text{ trials} \]

\[ P(X = k) = \binom{n}{k} \left( p^k q^{n-k} \right) \]
Example 1 – Cereal

Suppose you buy 5 boxes of cereal. Where \( n = 5 \) and \( p = 0.2 \). What’s the probability you get exactly 2 pictures of Tiger Woods?

\[
\binom{5}{2} = 5 \cdot C_2 = \frac{5!}{2!(5-2)!} = 10
\]

There are 10 ways to get 2 Tiger pictures in 5 boxes.
Example 1 – Cereal

Suppose you buy 5 boxes of cereal. Where \( n = 5 \) and \( p = 0.2 \). What’s the probability you get exactly 2 pictures of Tiger Woods?

There are 10 ways to get 2 Tiger pictures in 5 boxes.

\[
P(X = 2) = 10(0.20)^2(0.80)^3 = 0.2048
\]
## Example – Cereal

The following table shows the probability distribution function (p.d.f) for the binomial random variable, $X$.

<table>
<thead>
<tr>
<th>$X =$ Tiger</th>
<th>$P(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.32768</td>
</tr>
<tr>
<td>1</td>
<td>0.4096</td>
</tr>
<tr>
<td>2</td>
<td>0.2048</td>
</tr>
<tr>
<td>3</td>
<td>0.0512</td>
</tr>
<tr>
<td>4</td>
<td>0.0064</td>
</tr>
<tr>
<td>5</td>
<td>0.00032</td>
</tr>
</tbody>
</table>

\[
P(X = 0) = P(FFFFF) = 0.80^5 = 0.32768 \]

\[
\text{Binompdf} (5, 0.20, 0) = 0.32768
\]

\[
\text{Binompdf} (5, 0.20, 1) = 0.4096
\]

\[
\text{Binompdf} (5, 0.20, 2) = 0.2048
\]

\[
\text{Binompdf} (5, 0.20, 3) = 0.0512
\]

\[
\text{Binompdf} (5, 0.20, 4) = 0.0064
\]

\[
\text{Binompdf} (5, 0.20, 5) = 0.00032
\]
Example – Cereal

The following table shows the cumulative distribution function (c.d.f) for the binomial random variable, X.

<table>
<thead>
<tr>
<th>X</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X_{pdf})$</td>
<td>0.32768</td>
<td>0.4096</td>
<td>0.2048</td>
<td>.0512</td>
<td>.0064</td>
<td>.00032</td>
</tr>
<tr>
<td>$P(X_{cdf})$</td>
<td>$P(X \leq 0)$</td>
<td>$P(X \leq 1)$</td>
<td>$P(X \leq 2)$</td>
<td>$P(X \leq 3)$</td>
<td>$P(X \leq 4)$</td>
<td>$P(X \leq 5)$</td>
</tr>
<tr>
<td></td>
<td>0.32768</td>
<td>0.73728</td>
<td>0.94208</td>
<td>0.99328</td>
<td>0.99968</td>
<td>1</td>
</tr>
</tbody>
</table>

$\text{Binomcdf}(5, 0.20, 0) = 0.32768$

$\text{Binomcdf}(5, 0.20, 2) = 0.73728$

$\text{Binomcdf}(5, 0.20, 3) = 0.99328$

$\text{Binomcdf}(5, 0.20, 4) = 0.99968$

$\text{Binomcdf}(5, 0.20, 5) = 1$
Example – Cereal

Construct a histogram of the pdf and cdf using $X[0, 6]_1$ and $Y[0, 1]_{0.01}$

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X_{pdf})$</td>
<td>0.32768</td>
<td>0.4096</td>
<td>0.2048</td>
<td>0.0512</td>
<td>0.0064</td>
<td>0.00032</td>
</tr>
<tr>
<td>$P(X_{cdf})$</td>
<td>$P(X \leq 0)$</td>
<td>$P(X \leq 1)$</td>
<td>$P(X \leq 2)$</td>
<td>$P(X \leq 3)$</td>
<td>$P(X \leq 4)$</td>
<td>$P(X \leq 5)$</td>
</tr>
<tr>
<td></td>
<td>0.32768</td>
<td>0.73728</td>
<td>0.94208</td>
<td>0.99328</td>
<td>0.99968</td>
<td>1</td>
</tr>
</tbody>
</table>
In each of the following cases, decide whether or not a binomial distribution is an appropriate model, and give your reasons.

A). Fifty students are taught about the binomial distributions by a television program. After completing their study, all students take the same examination. The number who pass is counted.

Yes, it would be reasonable to assume that the results for the 50 students are independent, and each has the same chance of passing.
B). A student studies binomial distributions using computer instruction. After the initial instruction is completed, the computer presents 10 problems. The student solves each problem and enters the answer: the computer gives additional instruction between problems if the student’s answer is wrong. The number of problems that the student solves correctly is counted.

No; since the student receives instruction after incorrect answers, her probability of success is likely to increase.
C). A chemist repeats a solubility test 10 times on the same substance. Each test is conducted at temperature 10° higher than the previous test. She counts the number of times that the substance dissolves completely.

No; temperature may affect the outcome of the test.
Suppose that James guesses on each question of a 50-item true-false quiz. Find the probability that James passes if

A). a score of 25 or more correct is needed to pass.

\[ X = \text{the number of correct answers. } X \text{ is binomial with } n = 50 \text{ and } p = 0.50 \]

\[
P(X \geq 25) = P(X = 25) + P(X = 26) + \ldots + P(X = 50)
\]

\[
= 1 - \text{binomialcdf}(50, 0.50, 24) = 0.556
\]
B). a score of 30 or more correct is needed to pass.

\[ X = \text{the number of correct answers. } X \text{ is binomial with } n = 50 \text{ and } p = 0.50 \]

\[ P(X \geq 30) = P(X = 30) + P(X = 31) + \ldots + P(X = 50) \]

\[ = 1 - \text{binocdf(50,0.50,29)} = 0.101 \]
C). a score of 32 or more correct is needed to pass.

\[ X = \text{the number of correct answers. } X \text{ is binomial with } n = 50 \text{ and } p = 0.50 \]

\[ P(X \geq 32) = P(X = 32) + P(X = 33) + \ldots + P(X = 50) \]

\[ = 1 - \text{binocdf}(50,0.50,31) = 0.032 \]
According to a 2000 study by the Bureau of Justice Statistics, approximately 2% of the nation’s 72 million children had a parent behind bars – nearly 1.5 million minors. Let X be the number of children who had an incarcerated parent. Suppose that 100 children are randomly selected.

A) Does X satisfy the requirements for a binomial setting? Explain. If $X = B(n, p)$, what are $n$ and $p$?

Yes, if the 100 children are randomly selected, it is extremely likely that the result for one child will not influence the result for any other child. “Success” in this context means having an incarcerated parent. Where $n = 100$ and $p = 0.02$
X = \text{B}(100, 0.02)

B). Describe \( P(X = 0) \) in words. Then find \( P(X = 0) \) and \( P(X = 1) \).

\[ P(X = 0) = \text{the probability of none of the 100 selected children having incarcerated parent.} \]

\[ P(X = 0) = \text{binompdf} (100, 0.02, 0) = 0.133 \]

\[ P(X = 1) = \text{binompdf} (100, 0.02, 1) = 0.271 \]
$X = B(100, 0.02)$

C). What is the probability that 2 or more of the 100 children have a parent behind bars.

$$P(X \geq 2) = P(X = 2) + P(X = 3) + \ldots + P(X = 100)$$

$$= 1 - \text{binomcdf}(100, 0.02, 1) = 0.596$$

About 60% of the time we’ll find 2 or more children with parents behind bars among the 100 children.
Example – Page 449, #8.10

Suppose you purchase a bundle of 10 bare-root broccoli plants. The sales clerk tells you that on average you can expect 5% of the plants to die before purchasing any broccoli. Assume that the bundle is a random sample of plants. Use the binomial formula to find the probability that you will lose at most one of the broccoli plants.

Let $X =$ the number of broccoli plants that you lose

$n = 10$ and $p = 0.05$

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= \binom{10}{0} (0.05)^0 (0.95)^{10} + \binom{10}{1} (0.05)^1 (0.95)^9$$

$$= 0.59874 + 0.31512 = 0.914$$
Lesson 8-1, Part 2

Mean and Standard Deviation
Mean and Standard Deviation

If $X$ is binomial random variable with parameters $n$ and $p$, then the **mean** and **standard deviation** of $X$ are:

$$\mu_X = np$$

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{npq}$$
A) What is the mean number of Hispanics on randomly chosen committees of 15 workers in Exercise 8.13 (page 449)?

\[ n = 15 \quad \quad \mu = np \]
\[ p = 0.3 \quad \quad = 15(0.3) \]
\[ \quad = 4.5 \]

B) What is the standard deviation \( \sigma \) of the count \( X \) of Hispanic members?

\[ \sigma = \sqrt{npq} = \sqrt{15(0.3)(0.7)} = 1.77482 \]
C) Suppose that 10% of the factory workers were Hispanic. Then $p = 0.1$. What is $\sigma$ in this case? What is $\sigma$ if $p = 0.01$? What does your work show about the behavior of the standard deviation of binomial distribution as the probability of a success gets closer to 0?

\[ n = 15 \]
\[ p = 0.10 \]
\[ \sigma = \sqrt{npq} = \sqrt{15(0.1)(0.9)} = 1.1619 \]

\[ n = 15 \]
\[ p = 0.01 \]
\[ \sigma = \sqrt{15(0.01)(0.99)} = 0.385357 \]

As $p$ gets closer to 0, $\sigma$ gets closer to 0.
Approximate a Binomial Distribution with a Normal Distribution if:

\[ np \geq 10 \]
\[ nq \geq 10 \]

then \( \mu = np \) and \( \sigma = \sqrt{npq} \)

and the random variable has a distribution.

(normal)
You operate a restaurant. You read that a sample survey by the National Restaurant Association shows that 40% of adults are committed to eating nutritious food when eating away from home. To help plan your menu, you decide to conduct a sample survey in your own area. You will use random digit dialing to contact an SRS of 200 households by telephone.

A). If the national results holds in area, it is reasonable to use the binomial distribution with \( n = 200 \) and \( p = 0.4 \) to describe the count \( X \) of respondents who seek nutritional food when eating out. Explain why.

Yes, this study satisfies the requirements of a binomial setting.
B). What is the mean number of nutrition-conscious people in your sample if \( p = 0.4 \) it true? What is the standard deviation?

\[
\mu = np = 200(0.4) = 80
\]

\[
\sigma = \sqrt{npq} = \sqrt{200(0.4)(0.6)} = \sqrt{48} = 6.9282
\]
\[ \mu = 80 \quad \sigma = \sqrt{48} = 6.9282 \]

C). What is the probability that \( X \) lies between 75 and 85? Make sure that the rule of thumb conditions are satisfied, and then use a normal approximation to answer the question.

\[ np \geq 10 \quad nq \geq 10 \]
\[ 80 \geq 10 \quad 200(.60) \geq 10 \]
\[ 120 \geq 10 \]

Rule of thumb is satisfied

\[ P(75 \leq X \leq 85) = P\left( \frac{75 - 80}{\sqrt{48}} \leq Z \leq \frac{85 - 80}{\sqrt{48}} \right) \]
Example – Page 455, #8.20

\[ \mu = 80 \quad \sigma = \sqrt{48} = 6.9282 \]

\[ P(75 \leq X \leq 85) = P\left( \frac{75 - 80}{\sqrt{48}} \leq Z \leq \frac{85 - 80}{\sqrt{48}} \right) \]

\[ = P(-0.72 \leq Z \leq 0.72) = 0.5285 \]

\[ \text{normalcdf} \ (-0.72, 0.72, 0, 1) = 0.528475 \]

\[ \text{normalcdf} \ (75, 85, 80, \sqrt{48}) = 0.5295 \]
Lesson 8-2

Geometric Distributions
Example 2 – Cereal

Suppose a cereal manufacturer puts pictures of famous athletes on cards in boxes of cereal, in the hope of increasing sales. The manufacture announces that 20% of the boxes contain a picture of Tiger Woods, 30% a picture of Lance Armstrong, and the rest a picture of Serena Williams.

You’ve got to have the Tiger Woods picture, so you start madly opening boxes of cereal, hoping to find one. Assuming that the pictures are randomly distributed, there’s a 20% chance you succeed on any box you open.
Example 2 – Cereal

What’s the probability you find his picture in the first box of cereal? It’s 20%, of course. We could write

\[ P(\# \text{ of boxes} = 1) = 0.20. \]

How about the probability that you don’t find Tiger until the second box? \[ P(\# \text{ of boxes} = 2) = (0.8)(0.2) = 0.16 \]

Of course, you could have a run of bad luck. Maybe you won’t find Tiger until the fifth box of cereal. What are the chances of that?

\[ P(\# \text{ of boxes} = 5) = (0.80)^4(0.20) = 0.08192 \]
Geometric Distributions

- Random variable $X =$ the number of trials required to obtain the *first success*
- $X$ is a geometric random variable
  - There are only two outcomes: success or failure.
  - The variable of interest is the number of trials required to obtain the first success.
  - The $n$ observations are independent.
  - The probability of success $p$ is the same for each observation.
- Since $n$ is not fixed there could be an infinite number of $X$ values.
- The probability histogram for a geometric is always skewed to the right.
The probability formula that $X$ is equal to $n$ is given by the following formula:

$$P(X = n) = (1 - p)^{n-1} p = q^{n-1} p$$

The probability that $X$ is greater than $n$ is given by the following formula:

$$P(X > n) = (1 - p)^n = q^n$$
The expected value of a geometric random variable is:

\[ \mu = \frac{1}{p} \]

The standard deviation of geometric random variable is:

\[ \sigma = \sqrt{\frac{q}{p^2}} \]
An experiment consists of rolling a die until a prime number (2, 3, or 5) is observed. Let $X =$ number of rolls required to get the first prime number.

A). Verify that $X$ has a geometric distribution.

The four conditions of geometric setting hold, with probability of success $\frac{1}{2}$.
B). Construct probability distribution table to include at least 5 entries for the probability of X. Record probabilities to four decimal places.

\[
\text{geometpdf}(0.50, L1)
\]
Example – Page 468, #38

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>0.50</td>
<td>0.25</td>
<td>0.125</td>
<td>0.0625</td>
<td>0.03125</td>
</tr>
<tr>
<td>c.d.f</td>
<td>0.50</td>
<td>0.75</td>
<td>0.875</td>
<td>0.9375</td>
<td>0.96875</td>
</tr>
</tbody>
</table>

geometcdf(0.50, L1)
C). Construct a graph of the pdf of $X$. 

```
Plot1 Plot2 PLOT3
ON Off ON
Type:  1 1 1
Xlist:L1
Freq:L2

WINDOW
Xmin=●
Xmax=6
Xscl=1
Ymin=0
Ymax=.75
Yscl=.05
Xres=1
```
D). Compute the cdf of $X$ and plot its histogram.

```
Plot1 Plot2 Plot3
On Off
Type: L:
Xlist: L1
Freq: L3

WINDOW
Xmin=1
Xmax=6
Xscl=1
Ymin=0
Ymax=1
Yscl=.05
Xres=1
```
The State Department is trying to identify an individual who speaks Farsi to fill a foreign embassy position. They have determined that 4% of the applicants pool are fluent in Farsi.

A) If applicants are contacted randomly, how many individuals can they expect to interview in order to find one who is fluent in Farsi?

\[ \mu = \frac{1}{p} = \frac{1}{0.04} = 25 \] applicants
B) What is the probability that they will have to interview more than 25 until they find one who speaks Farsi? More than 40?

\[ P(x > 25) = (1 - p)^n = (1 - 0.04)^{25} = 0.3604 \]

\[ P(X > 25) = 1 - P(X \leq 25) = 1 - \text{geometcdf} (0.04, 25) = 0.3604 \]

\[ P(X > 40) = 1 - P(X \leq 40) = 1 - \text{geometcdf} (0.04, 40) = 0.1954 \]