Chapter 6
Random Variables

6.1 Discrete and Continuous Random Variables
6.2 Transforming and Combining Random Variables
6.3 Binomial and Geometric Random Variables
Section 6.3
Binomial and Geometric Random Variables

Learning Objectives

After this section, you should be able to…

- DETERMINE whether the conditions for a binomial setting are met
- COMPUTE and INTERPRET probabilities involving binomial random variables
- CALCULATE the mean and standard deviation of a binomial random variable and INTERPRET these values in context
- CALCULATE probabilities involving geometric random variables
Binomial Settings

When the same chance process is repeated several times, we are often interested in whether a particular outcome does or doesn’t happen on each repetition. In some cases, the number of repeated trials is fixed in advance and we are interested in the number of times a particular event (called a “success”) occurs. If the trials in these cases are independent and each success has an equal chance of occurring, we have a binomial setting.

**Definition:**
A binomial setting arises when we perform several independent trials of the same chance process and record the number of times that a particular outcome occurs. The four conditions for a binomial setting are

- **B**inary? The possible outcomes of each trial can be classified as “success” or “failure.”
- **I**ndependent? Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial.
- **N**umber? The number of trials \( n \) of the chance process must be fixed in advance.
- **S**uccess? On each trial, the probability \( p \) of success must be the same.
Binomial Random Variable

Consider tossing a coin \( n \) times. Each toss gives either heads or tails. Knowing the outcome of one toss does not change the probability of an outcome on any other toss. If we define heads as a success, then \( p \) is the probability of a head and is 0.5 on any toss.

The number of heads in \( n \) tosses is a **binomial random variable** \( X \). The probability distribution of \( X \) is called a **binomial distribution**.

**Definition:**

The count \( X \) of successes in a binomial setting is a **binomial random variable**. The probability distribution of \( X \) is a **binomial distribution** with parameters \( n \) and \( p \), where \( n \) is the number of trials of the chance process and \( p \) is the probability of a success on any one trial. The possible values of \( X \) are the whole numbers from 0 to \( n \).

**Note:** When checking the Binomial condition, be sure to check the BINS and make sure you’re being asked to count the number of successes in a certain number of trials!
Alternate Example – Dice, Cars, and Hoops

Problem: Determine whether the random variables below have a binomial distribution. Justify your answer.

- (a) Roll a fair die 10 times and let $X=\text{the number of sixes}$.
- (b) Shoot a basketball 20 times from various distances on the court. Let $Y=\text{number of shots made}$.
- (c) Observe the next 100 cars that go by and let $C=\text{color}$.

Solution:
(a) Binary? Yes; success = six, failure = not a six, Independent? Yes; knowing the outcomes of past rolls doesn’t provide additional information about the outcomes of future rolls. Number? Yes; there are 10 trials. Success? Yes; the probability of success is always $\frac{1}{6}$. This is a binomial setting. $X$ is binomial with $n = 10$ and $p = \frac{1}{6}$.
(b) Binary? Yes; success = make the shot failure = miss the shot. Independent? Yes; evidence suggests that it is reasonable to assume that making a shot doesn’t change the probability of making the next shot. Number? Yes; there are 20 trials. Success? No; the probability of success changes because the shots are taken from various distances.
(c) Binary? No. There are more than two possible colors. Also, $C$ is not even a random variable since the outcomes aren’t numerical.
Binomial and Geometric Random Variables

Binomial Probabilities

In a binomial setting, we can define a random variable (say, $X$) as the number of successes in $n$ independent trials. We are interested in finding the probability distribution of $X$.

Each child of a particular pair of parents has probability 0.25 of having type O blood. Genetics says that children receive genes from each of their parents independently. If these parents have 5 children, the count $X$ of children with type O blood is a binomial random variable with $n = 5$ trials and probability $p = 0.25$ of a success on each trial. In this setting, a child with type O blood is a “success” (S) and a child with another blood type is a “failure” (F).

What’s $P(X = 2)$?

$P(SSFFF) = (0.25)(0.25)(0.75)(0.75)(0.75) = (0.25)^2(0.75)^3 = 0.02637$

However, there are a number of different arrangements in which 2 out of the 5 children have type O blood:

- SSFFF
- SFSFF
- SFFSF
- SFFFS
- FSSFF
- FSFSF
- FSFFS
- FFSSF
- FFSFS
- FFFSS

Verify that in each arrangement, $P(X = 2) = (0.25)^2(0.75)^3 = 0.02637$

Therefore, $P(X = 2) = 10(0.25)^2(0.75)^3 = 0.2637$
Alternate Example – Rolling Doubles

In many games involving dice, rolling doubles is desirable. Rolling doubles mean the outcomes of two dice are the same, such as 1&1 or 5&5. The probability of rolling doubles when rolling two dice is $6/36 = 1/6$. If $X =$ the number of doubles in 4 rolls of two dice, then $X$ is binomial with $n = 4$ and $p = 1/6$.

What is $P(X = 0)$? That is, what is the probability that all 4 rolls are not doubles? Since the probability of not getting doubles is $1 – 1/6 = 5/6$, $P(X = 0) = P(FFFF) = (5/6)(5/6)(5/6)(5/6) = (5/6)^4 = 0.482$. Note: $F$ represents a failure and $S$ represents a Success.

What is $P(X = 1)$? There are four different ways to roll doubles once in four tries. For example, the doubles could occur on the first try (SFFF), the second try (FSFF), the third try (FFSF) or the fourth try (FFFS).

$P(SFFF) = (1/6)(5/6)(5/6)(5/6) = (1/6)(5/6)^3$

$P(FSFF) = (5/6)(1/6)(5/6)(5/6) = (1/6)(5/6)^3$

$P(FFSF) = (5/6)(5/6)(1/6)(5/6) = (1/6)(5/6)^3$

$P(FFFS) = (5/6)(5/6)(5/6)(1/6) = (1/6)(5/6)^3$

Thus, the probability of rolling doubles once in 4 attempts is $P(X = 1) = 4(1/6)(5/6)^3 = 0.386$. 
Binomial Coefficient

Note, in the previous example, *any one arrangement* of 2 S’s and 3 F’s had the same probability. This is true because no matter what arrangement, we’d multiply together 0.25 twice and 0.75 three times.

We can generalize this for any setting in which we are interested in $k$ successes in $n$ trials. That is,

$$P(X = k) = P(\text{exactly } k \text{ successes in } n \text{ trials}) = \text{number of arrangements} \cdot p^k (1 - p)^{n-k}$$

**Definition:**

The number of ways of arranging $k$ successes among $n$ observations is given by the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

for $k = 0, 1, 2, \ldots, n$ where

$$n! = n(n - 1)(n - 2)\cdots(3)(2)(1)$$

and $0! = 1$. 
Binomial Probability

The binomial coefficient counts the number of different ways in which $k$ successes can be arranged among $n$ trials. The binomial probability $P(X = k)$ is this count multiplied by the probability of any one specific arrangement of the $k$ successes.

If $X$ has the binomial distribution with $n$ trials and probability $p$ of success on each trial, the possible values of $X$ are $0, 1, 2, \ldots, n$. If $k$ is any one of these values,

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Number of arrangements of $k$ successes
- Probability of $k$ successes
- Probability of $n-k$ failures
Example: Inheriting Blood Type

Each child of a particular pair of parents has probability 0.25 of having blood type O. Suppose the parents have 5 children.

(a) Find the probability that exactly 3 of the children have type O blood.

Let \( X \) = the number of children with type O blood. We know \( X \) has a binomial distribution with \( n = 5 \) and \( p = 0.25 \).

\[
P(X = 3) = \binom{5}{3} (0.25)^3 (0.75)^2 = 10(0.25)^3 (0.75)^2 = 0.08789
\]

(b) Should the parents be surprised if more than 3 of their children have type O blood?

To answer this, we need to find \( P(X > 3) \).

\[
P(X > 3) = P(X = 4) + P(X = 5)
\]

\[
= \binom{5}{4} (0.25)^4 (0.75)^1 + \binom{5}{5} (0.25)^5 (0.75)^0
\]

\[
= 5(0.25)^4 (0.75)^1 + 1(0.25)^5 (0.75)^0
\]

\[
= 0.01465 + 0.00098 = 0.01563
\]

Since there is only a 1.5% chance that more than 3 children out of 5 would have Type O blood, the parents should be surprised!
Alternate Example: Rolling Doubles

When rolling two dice, the probability of rolling doubles is 1/6. Suppose that a game player rolls the dice 4 times, hoping to roll doubles.

(a) Find the probability that the player gets doubles twice in four attempts.

Let $X = \text{number of doubles}$. $X$ has a binomial distribution with $n = 4$ and $p = 1/6$.

$$P(X = 2) = \binom{4}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = 6 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = 0.116$$

(b) Should the player be surprised if he gets doubles more than twice in four attempts? Justify your answer.

To answer this, we need to find $P(X > 2)$.

$$P(X > 2) = P(X = 3) + P(X = 4)$$

$$= \binom{4}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^1 + \binom{4}{4} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^0 = 0.015 + 0.001 = 0.016$$

Since there is only a 1.6% chance of getting more than 2 doubles in four rolls, the player should be surprised if this happens.
Mean and Standard Deviation of a Binomial Distribution

We describe the probability distribution of a binomial random variable just like any other distribution – by looking at the shape, center, and spread. Consider the probability distribution of \( X \) = number of children with type O blood in a family with 5 children.

<table>
<thead>
<tr>
<th>( x_i )</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_i )</td>
<td>0.2373</td>
<td>0.3955</td>
<td>0.2637</td>
<td>0.0879</td>
<td>0.0147</td>
<td>0.00098</td>
</tr>
</tbody>
</table>

**Shape:** The probability distribution of \( X \) is skewed to the right. It is more likely to have 0, 1, or 2 children with type O blood than a larger value.

**Center:** The median number of children with type O blood is 1. Based on our formula for the mean:

\[
\mu_X = \sum x_i p_i = (0)(0.2373) + 1(0.3955) + ... + 5(0.00098) = 1.25
\]

**Spread:** The variance of \( X \) is

\[
\sigma^2_X = \sum (x_i - \mu_X)^2 p_i = (0 - 1.25)^2(0.2373) + (1 - 1.25)^2(0.3955) + ... + (5 - 1.25)^2(0.00098) = 0.9375
\]

The standard deviation of \( X \) is \( \sigma_X = \sqrt{0.9375} = 0.968 \)
Alternate Example – Rolling Doubles

If $X =$ number of doubles in 4 attempts, then $X$ follows a binomial distribution with $n = 4$ and $p = 1/6$. Here is the probability distribution and a histogram of the probability distribution.

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_i$</td>
<td>0.482</td>
<td>0.386</td>
<td>0.116</td>
<td>0.015</td>
<td>0.001</td>
</tr>
</tbody>
</table>

$$
\mu_X = \sum x_i p_i = (0)(0.482) + 1(0.386) + ... + 4(0.001) = 4 \left( \frac{1}{6} \right) = 0.667
$$

$$
\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i = (0 - 0.667)^2(0.482) + (1 - 0.667)^2(0.386) + ... + (4 - 0.667)^2(0.001) = 4 \left( \frac{1}{6} \right) \left( \frac{5}{6} \right) = 0.556
$$

So the standard deviation of $X$ is

$$
\sigma_X = \sqrt{0.556} = 0.746
$$
Mean and Standard Deviation of a Binomial Distribution

Notice, the mean $\mu_X = 1.25$ can be found another way. Since each child has a 0.25 chance of inheriting type O blood, we’d expect one-fourth of the 5 children to have this blood type. That is, $\mu_X = 5(0.25) = 1.25$. This method can be used to find the mean of any binomial random variable with parameters $n$ and $p$.

If a count $X$ has the binomial distribution with number of trials $n$ and probability of success $p$, the mean and standard deviation of $X$ are

$$\mu_X = np$$
$$\sigma_X = \sqrt{np(1-p)}$$

Note: These formulas work ONLY for binomial distributions. They can’t be used for other distributions!
Example: Bottled Water versus Tap Water

Mr. Bullard’s 21 AP Statistics students did the Activity on page 340. If we assume the students in his class cannot tell tap water from bottled water, then each has a $1/3$ chance of correctly identifying the different type of water by guessing. Let $X = \text{the number of students who correctly identify the cup containing the different type of water}$.

Find the mean and standard deviation of $X$.

Since $X$ is a binomial random variable with parameters $n = 21$ and $p = 1/3$, we can use the formulas for the mean and standard deviation of a binomial random variable.

\[
\mu_X = np = 21 \times \frac{1}{3} = 7
\]

\[
\sigma_X = \sqrt{np(1-p)} = \sqrt{21 \times \frac{1}{3} \times \frac{2}{3}} = 2.16
\]

We’d expect about one-third of his 21 students, about 7, to guess correctly.

If the activity were repeated many times with groups of 21 students who were just guessing, the number of correct identifications would differ from 7 by an average of 2.16.
Alternate Example: Tastes as Good as the Real Thing

The makers of a diet cola claim that its taste is indistinguishable from the full calorie version of the same cola. To investigate, an AP Statistics student named Emily prepared small samples of each type of soda in identical cups. Then, she had volunteers taste each cola in a random order and try to identify which was the diet cola and which was the regular cola. Overall, 23 of the 30 subjects made the correct identification.

**Explain why X is a binomial random variable.**

The chance process is each volunteer guessing which sample is the diet cola.
Binary? Yes, guesses are either correct or incorrect.
Independent? Yes, the results of one volunteer’s guess should have no effect on the results of other volunteers.
Number? Yes, there are 30 trials.
Success? Yes, the probability of guessing correctly is always 50%.
Since \( X \) is counting the number of successful guesses, \( X \) is a binomial random variable.
Alternate Example: Tastes as Good as the Real Thing

The makers of a diet cola claim that its taste is indistinguishable from the full calorie version of the same cola. To investigate, an AP Statistics student named Emily prepared small samples of each type of soda in identical cups. Then, she had volunteers taste each cola in a random order and try to identify which was the diet cola and which was the regular cola. Overall, 23 of the 30 subjects made the correct identification.

Find the mean and standard deviation of $X$. Interpret each value in context.

$$
\mu_X = np \\
= 30(0.5) = 15
$$

$$
\sigma_X = \sqrt{np(1-p)} \\
= \sqrt{30(0.5)(1-0.5)} = 2.74
$$

If this experiment were repeated many times and the volunteers were randomly guessing, the average number of correct guesses would be about 15 and the number of correct guesses would vary from 15 by about 2.74, on average.

Of the 30 volunteers, 23 made correct identifications. Does this give convincing evidence that the volunteers can taste the difference between the diet and regular colas?

$$
P(X \geq 23) = 1 - P(X \leq 22) = 1 - \text{binomialcdf}(30,0.5,22) = 1 - 0.9974 = 0.0026
$$

There is a very small chance that there would be 23 or more correct guesses if the volunteers couldn’t tell the difference in the colas. Therefore, we have convincing evidence that the volunteers can taste the difference.
Binomial Distributions in Statistical Sampling

The binomial distributions are important in statistics when we want to make inferences about the proportion $p$ of successes in a population.

Suppose 10% of CDs have defective copy-protection schemes that can harm computers. A music distributor inspects an SRS of 10 CDs from a shipment of 10,000. Let $X =$ number of defective CDs. **What is $P(X = 0)$?** Note, this is not quite a binomial setting. Why?

The actual probability is $P(\text{no defectives}) = \frac{9000}{10000} \cdot \frac{8999}{9999} \cdot \frac{8998}{9998} \cdot ... \cdot \frac{8991}{9991} = 0.3485$

Using the binomial distribution, $P(X = 0) = \binom{10}{0} (0.10)^0 (0.90)^{10} = 0.3487$

In practice, the binomial distribution gives a good approximation as long as we don’t sample more than 10% of the population.

Sampling Without Replacement Condition

When taking an SRS of size $n$ from a population of size $N$, we can use a binomial distribution to model the count of successes in the sample as long as $n \leq \frac{1}{10} N$. 
In the NASCAR Cards and Cereal Boxes example from section 5.1, we read about a cereal company that put one of 5 different cards into each box of cereal. Each card featured a different driver: Jeff Gordon, Dale Earnhardt, Jr., Tony Stewart, Danica Patrick, or Jimmie Johnson. Suppose that the company printed 20,000 of each card, so there were 100,000 total boxes of cereal with a card inside. If a person bought 6 boxes at random, what is the probability of getting no Danica Patrick cards?

Let X be the number of Danica Patrick cards obtained from 6 different boxes of cereal. Since we are sampling without replacement, the trials are not independent, so the distribution of X is not quite binomial—but it is close.

If we assume X is binomial with \( n = 6 \) and \( p = 0.2 \), then

\[
P(X = 0) = \binom{6}{0} (0.2)^0 (0.8)^6 = 0.262144
\]

The actual probability is:

\[
P(\text{no DP cards}) = \frac{80,000}{100,000} \cdot \frac{79,999}{99,999} \cdot \frac{79,998}{99,998} \cdot \frac{79,997}{99,997} \cdot \frac{79,996}{99,996} \cdot \frac{79,995}{99,995} = 0.262134
\]

These two probabilities are quite close!
Alternate Example – Dead Batteries

Almost everyone has one—a drawer that holds miscellaneous batteries of all sizes. Suppose that your drawer contains 8 AAA batteries but only 6 of them are good. You need to choose 4 for your graphing calculator. If you randomly select 4 batteries, what is the probability that all 4 of the batteries you choose will work?

Problem: Explain why the answer isn’t \( P(X = 4) = \binom{4}{4}(0.75)^4(0.25)^0 = 0.3164 \). The actual probability is 0.2143.

Solution: Since we are sampling without replacement, the selections of batteries aren’t independent. We can ignore this problem if the sample we are selecting is less than 10% of the population. However, in this case we are sampling 50% of the population (4/8), so it is not reasonable to ignore the lack of independence and use the binomial distribution. This explains why the binomial probability is so different from the actual probability.
Normal Approximation for Binomial Distributions

As $n$ gets larger, something interesting happens to the shape of a binomial distribution. The figures below show histograms of binomial distributions for different values of $n$ and $p$. What do you notice as $n$ gets larger?

![Histograms of binomial distributions for different $n$ and $p$.](image)

Suppose that $X$ has the binomial distribution with $n$ trials and success probability $p$. When $n$ is large, the distribution of $X$ is approximately Normal with mean and standard deviation

$$\mu_X = np \quad \sigma_X = \sqrt{np(1-p)}$$

As a rule of thumb, we will use the Normal approximation when $n$ is so large that $np \geq 10$ and $n(1-p) \geq 10$. That is, the expected number of successes and failures are both at least 10.
Example: Attitudes Toward Shopping

Sample surveys show that fewer people enjoy shopping than in the past. A survey asked a nationwide random sample of 2500 adults if they agreed or disagreed that “I like buying new clothes, but shopping is often frustrating and time-consuming.” Suppose that exactly 60% of all adult US residents would say “Agree” if asked the same question. Let \( X \) = the number in the sample who agree. **Estimate the probability that 1520 or more of the sample agree.**

1) Verify that \( X \) is approximately a binomial random variable.

- **B:** Success = agree, Failure = don’t agree
- **I:** Because the population of U.S. adults is greater than 25,000, it is reasonable to assume the sampling without replacement condition is met.
- **N:** \( n = 2500 \) trials of the chance process
- **S:** The probability of selecting an adult who agrees is \( p = 0.60 \)

2) Check the conditions for using a Normal approximation.

Since \( np = 2500(0.60) = 1500 \) and \( n(1 - p) = 2500(0.40) = 1000 \) are both at least 10, we may use the Normal approximation.

3) Calculate \( P(X \geq 1520) \) using a Normal approximation.

\[
\mu = np = 2500(0.60) = 1500 \\
\sigma = \sqrt{np(1 - p)} = \sqrt{2500(0.60)(0.40)} = 24.49 \\
z = \frac{1520 - 1500}{24.49} = 0.82 \\
P(X \geq 1520) = P(Z \geq 0.82) = 1 - 0.7939 = 0.2061
\]
Alternate Example: Teens and Debit Cards

In a survey of 506 teenagers ages 14-18, subjects were asked a variety of questions about personal finance. (http://www.nclnet.org/personal-finance/66-teens-and-money/120-ncl-survey-teens-and-financial-education) One question asked teens if they had a debit card.

1) Verify that $X$ is approximately a binomial random variable.

Binary? Yes, teens either have a debit card or they don’t. Independent? No, since we are sampling without replacement. However, since the sample size ($n = 506$) is much less than 10% of the population size (there are millions of teens ages 14-16), the responses will be very close to independent. Number? Yes, there is a fixed sample size, $n = 506$. Success? Yes, the unconditional probability of selecting a teen with a debit card is 10%.

2) Check the conditions for using a Normal approximation.

Since $np = 506(0.10) = 50.6$ and $n(1 - p) = 506(0.90) = 455.4$ are both at least 10, we should be safe using the normal approximation.

3) Calculate $P(X \geq 1520)$ using a Normal approximation.

$\mu_X = np = 506(0.10) = 50.6$

$\sigma_X = \sqrt{np(1 - p)} = \sqrt{506(0.1)(0.9)} = 6.75$

$P(X \leq 40) = \text{normalcdf}(-9999, 40, 50.6, 6.75) = 0.058$

(The probability using the binomial distribution is 0.064). Note: to get full credit on the AP exam when using the calculator command normalcdf, students must clearly identify the shape (Normal), center (mean = 50.6) and spread (standard deviation = 6.75) somewhere in their work.
## Geometric Settings

In a binomial setting, the number of trials $n$ is fixed and the binomial random variable $X$ counts the number of successes. In other situations, the goal is to repeat a chance behavior until a success occurs. These situations are called geometric settings.

**Definition:**
A geometric setting arises when we perform independent trials of the same chance process and record the number of trials until a particular outcome occurs. The four conditions for a geometric setting are

<table>
<thead>
<tr>
<th>B</th>
<th>Binary? The possible outcomes of each trial can be classified as “success” or “failure.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Independent? Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial.</td>
</tr>
<tr>
<td>T</td>
<td>Trials? The goal is to count the number of trials until the first success occurs.</td>
</tr>
<tr>
<td>S</td>
<td>Success? On each trial, the probability $p$ of success must be the same.</td>
</tr>
</tbody>
</table>
Geometric Random Variable

In a geometric setting, if we define the random variable $Y$ to be the number of trials needed to get the first success, then $Y$ is called a geometric random variable. The probability distribution of $Y$ is called a geometric distribution.

**Definition:**
The number of trials $Y$ that it takes to get a success in a geometric setting is a geometric random variable. The probability distribution of $Y$ is a geometric distribution with parameter $p$, the probability of a success on any trial. The possible values of $Y$ are 1, 2, 3, ... .

*Note:* Like binomial random variables, it is important to be able to distinguish situations in which the geometric distribution does and doesn’t apply!
Example: The Birthday Game

Read the activity on page 398. The random variable of interest in this game is $Y =$ the number of guesses it takes to correctly identify the birth day of one of your teacher’s friends. What is the probability the first student guesses correctly? The second? Third? What is the probability the $k^{th}$ student guesses correctly?

Verify that $Y$ is a geometric random variable.

B: Success = correct guess, Failure = incorrect guess
I: The result of one student’s guess has no effect on the result of any other guess.
T: We’re counting the number of guesses up to and including the first correct guess.
S: On each trial, the probability of a correct guess is $1/7$.

Calculate $P(Y = 1)$, $P(Y = 2)$, $P(Y = 3)$, and $P(Y = k)$

$P(Y = 1) = 1/7$

$P(Y = 2) = (6/7)(1/7) = 0.1224$

$P(Y = 3) = (6/7)(6/7)(1/7) = 0.1050$

Notice the pattern?

Geometric Probability

If $Y$ has the geometric distribution with probability $p$ of success on each trial, the possible values of $Y$ are 1, 2, 3, … . If $k$ is any one of these values,

$$P(Y = k) = (1 - p)^{k-1} p$$
Alternate Example: Monopoly

Problem: Let the random variable \( Y \) be defined as in the previous alternate example.

(a) Find the probability that it takes 3 turns to roll doubles.
\[
P(Y = 3) = \left( \frac{5}{6} \right)^2 \left( \frac{1}{6} \right) = 0.116
\]

(b) Find the probability that it takes more than 3 turns to roll doubles and interpret this value in context.

Since there are an infinite number of possible values of \( Y \) greater than 3, we will use the complement rule.
\[
P(Y > 3) = 1 - P(Y \leq 3) = 1 - P(Y = 3) - P(Y = 2) - P(Y = 1)
\]
\[
= 1 - \left( \frac{5}{6} \right)^2 \left( \frac{1}{6} \right) - \left( \frac{5}{6} \right)^1 \left( \frac{1}{6} \right) - \left( \frac{5}{6} \right)^0 \left( \frac{1}{6} \right) = 0.5787
\]

If a player tried to get out of jail many, many times by trying to roll doubles, about 58% of the time it would take more than 3 attempts. Note: if the first success occurs after the 3\textsuperscript{rd} attempt, this means that the first three attempts are all failures. Thus, an alternative solution would be \( P(Y > 3) = P(\text{no doubles and no doubles and no doubles}) = (5/6)^3 = 0.5787. \)
Mean of a Geometric Distribution

The table below shows part of the probability distribution of $Y$. We can’t show the entire distribution because the number of trials it takes to get the first success could be an incredibly large number.

<table>
<thead>
<tr>
<th>$y_i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_i$</td>
<td>0.143</td>
<td>0.122</td>
<td>0.105</td>
<td>0.090</td>
<td>0.077</td>
<td>0.066</td>
<td></td>
</tr>
</tbody>
</table>

**Shape:** The heavily right-skewed shape is characteristic of any geometric distribution. That’s because the most likely value is 1.

**Center:** The mean of $Y$ is $\mu_Y = 7$. We’d expect it to take 7 guesses to get our first success.

**Spread:** The standard deviation of $Y$ is $\sigma_Y = 6.48$. If the class played the Birth Day game many times, the number of homework problems the students receive would differ from 7 by an average of 6.48.

---

**Mean (Expected Value) of Geometric Random Variable**

If $Y$ is a geometric random variable with probability $p$ of success on each trial, then its **mean** (expected value) is $E(Y) = \mu_Y = 1/p$. 
A binomial setting consists of \( n \) independent trials of the same chance process, each resulting in a success or a failure, with probability of success \( p \) on each trial. The count \( X \) of successes is a **binomial random variable**. Its probability distribution is a **binomial distribution**.

The **binomial coefficient** counts the number of ways \( k \) successes can be arranged among \( n \) trials.

If \( X \) has the binomial distribution with parameters \( n \) and \( p \), the possible values of \( X \) are the whole numbers 0, 1, 2, . . . , \( n \). The binomial probability of observing \( k \) successes in \( n \) trials is

\[
P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]
Section 6.3
Binomial and Geometric Random Variables

Summary
In this section, we learned that...

- The **mean** and **standard deviation** of a binomial random variable $X$ are
  \[ \mu_X = np \]
  \[ \sigma_X = \sqrt{np(1 - p)} \]

- The **Normal approximation** to the binomial distribution says that if $X$ is a count having the binomial distribution with parameters $n$ and $p$, then when $n$ is large, $X$ is approximately Normally distributed. We will use this approximation when $np \geq 10$ and $n(1 - p) \geq 10$.  

Summary

In this section, we learned that...

✓ A **geometric setting** consists of repeated trials of the same chance process in which each trial results in a success or a failure; trials are independent; each trial has the same probability $p$ of success; and the goal is to count the number of trials until the first success occurs. If $Y$ = the number of trials required to obtain the first success, then $Y$ is a **geometric random variable**. Its probability distribution is called a **geometric distribution**.

✓ If $Y$ has the geometric distribution with probability of success $p$, the possible values of $Y$ are the positive integers 1, 2, 3, . . . . The **geometric probability** that $Y$ takes any value is

$$P(Y = k) = (1 - p)^{k-1} p$$

✓ The **mean** (expected value) of a geometric random variable $Y$ is $1/p$. 

Section 6.3
Binomial and Geometric Random Variables
We’ll learn how to describe sampling distributions that result when data are produced by random sampling.

We’ll learn about

✓ **Sampling Distributions**
✓ **Sample Proportions**
✓ **Sample Means**