Section 6.3: Binomial and 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.**

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

Binary? The possible outcomes of each trial can be classified as "success" or "failure."

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.

Number? The number of trials n of the chance process must be fixed in advance.

Success? On each trial, the probability p of success must be the same.

Binomial Random Variable

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

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!

Example 1: Here are three scenarios involving chance behavior. In each case, determine whether the given random variable has a binomial distribution. Justify your answer.

a) Genetics says that children receive genes from each of their parents independently. Each child of a particular pair of parents has probability 0.25 of having type O blood. Suppose these parents have 5 children. Let X = the number of children with type O blood.

***Binary?** "Success" = has type O blood. "Failure" = doesn't have type O blood.

***Independent?** The problem states that children receive genes from each of their parents independently.

***Number?** There are n = 5 trials of this chance process.

***Success?** The probability of a success is p = 0.25 on each trial.

b) Shuffle a deck of cards. Turn over the first 10 cards, one at a time. Let Y = the number of aces you observe.

***Independent?** No. If the first card you turn over is an ace, then the next card is less likely to be an ace because you're not replacing the top card in the deck. Similarly, if the first card isn't an ace, the second card is more likely to be an ace.

Because the trials are not independent, this is not a binomial setting.

c) Shuffle a deck of cards. Turn over the top card. Put the card back in the deck, and shuffle again. Repeat this process until you get an ace. Let W = the number of cards required.

***Number?** The number of trials is not set in advance. You could get an ace on the first card you turn over, or it may take many cards to get an ace.

Because there is no fixed number of trials, this is not a binomial setting.

Binomial Probabilities

In a binomial setting, we can define a random variable (say, X) as the number of successes in n independent trials. What's the probability distribution of X?

Example 2: Each child of a particular set 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 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 $P(X = 0)$? That is, what's the probability that *none* of the 5 children has type O blood? It's the chance that all 5 children *don't* have type O blood. The probability that any one of this couple's children doesn't have type O blood is 1 − 0.25 = 0.75 (complement rule).

By the multiplication rule for independent events (**Chapter 5**),

$P(X = 0) = P(\text{FFFFF}) = (0.75)(0.75)(0.75)(0.75)(0.75) = (0.75)^5 = 0.2373$

How about $P(X = 1)$? There are several different ways in which exactly 1 of the 5 children could have type O blood. For instance, the first child born might have type O blood, while the remaining 4 children don't have type O blood. The probability that this happens is

$P(\text{SFFFF}) = (0.25)(0.75)(0.75)(0.75)(0.75) = (0.25)(0.75)^4$

Alternatively, Child 2 could be the one that has type O blood. The corresponding probability is

$P(\text{FSFFF}) = (0.75)(0.25)(0.75)(0.75)(0.75) = (0.25)(0.75)^4$

There are three more possibilities to consider:

$P(\text{FFSFF}) = (0.75)(0.75)(0.25)(0.75)(0.75) = (0.25)(0.75)^4$
$P(\text{FFFSF}) = (0.75)(0.75)(0.75)(0.25)(0.75) = (0.25)(0.75)^4$
$P(\text{FFFFS}) = (0.75)(0.75)(0.75)(0.75)(0.25) = (0.25)(0.75)^4$

In all, there are five different ways in which exactly 1 child would have type O blood, each with the same probability of occurring. As a result,

$$\begin{aligned} P(X = 1) &= P(\text{exactly 1 child with type O blood}) \\ &= P(\text{SFFFF}) + P(\text{FSFFF}) + P(\text{FFSFF}) + P(\text{FFFSF}) + P(\text{FFFFS}) \\ &= 5(0.25)(0.75)^4 = 0.39551 \end{aligned}$$

There's about a 40% chance that exactly 1 of the couple's 5 children will have type O blood.

Binomial Coefficient

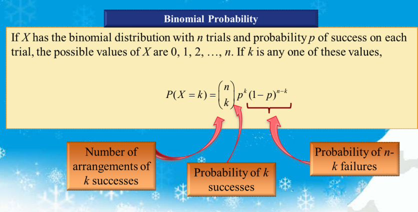
The number of ways of arranging k successes among n observations is given by the **binomial coefficient**

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

for $k = 0, 1, 2, \ldots, n$ where $n! = n(n-1)(n-2)\bullet\ldots\bullet(3)(2)(1)$ 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.

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



Example 3: Each child of a particular pair of parents has probability 0.25 of having type O blood. Suppose the parents have 5 children.

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

We want to find $P(X = 3)$.

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

There is about a 9% chance that exactly 3 of the 5 children have type O blood.

b) Should the parents be surprised if more than 3 of their children have type O blood? Justify your answer.

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

$$P(X>3)=P(X=4)+P(X=5)=\binom{5}{4}(0.25)^4(0.75)^1+\binom{5}{5}(0.25)^5(0.75)^0$$
$$=5(0.25)^4(0.75)^1+1(0.25)^5=0.01465+0.00098=0.01563$$

Because there's only about a 1.5% chance of having more than 3 children with type O blood, the parents should definitely be surprised if this happens.

How to find binomial probabilities

Step 1: State the distribution and the values of interest. Specify a binomial distribution with the number of trials n, success probability p, and the values of the variable clearly identified.

Step 2: Perform calculations—show your work! Do one of the following:

(i) Use the binomial probability formula to find the desired probability; or

(ii) use the binompdf or binomcdf command and label each of the inputs.

Step 3: Answer the question.

Example 4: A local fast-food restaurant is running a "Draw a three, get it free" lunch promotion. After each customer orders, a touch-screen display shows the message "Press here to win a free lunch." A computer program then simulates one card being drawn from a standard deck. If the chosen card is a 3, the customer's order is free. Otherwise, the customer must pay the bill.

a) All 12 players on a school's basketball team place individual orders at the restaurant. What is the probability that exactly 2 of them win a free lunch?

Step 1: State the distribution and the values of interest. Let X = the number of players who win a free lunch. There are 12 independent trials. $B(12, 4/52)$.

We want to find $P(X = 2)$.

Perform calculations—show your work!

Using technology: The command binompdf(trials: 12, *p*: (4/52), *x* value: 2) = 0.1754.

Step 3: Answer the question. There is about a 17.5% probability that exactly 2 players will win a free lunch.

b) If 250 customers place lunch orders on the first day of the promotion, what's the probability that fewer than 10 win a free lunch?

Step 1: State the distribution and the values of interest. Let Y = the number of customers who win a free lunch. There are 250 independent trials. $B(250, 4/52)$

We want to find $P(Y < 10)$.

Step 2: Perform calculations—show your work! The values of Y that interest us are 0 1 2 3 4 5 6 7 8 9 10 11 12 … 250

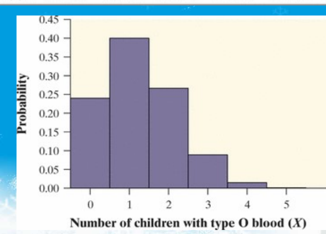
To use the binomial formula, we would have to add up the probabilities for $Y = 0$, $Y = 1$,…, $Y = 9$. That's too much work! The better option is to use technology:

$P(Y < 10) = P(Y \leq 9)$ = binomcdf(trials: 250, *p*: (4/52), *x* value: 9) = 0.00613.

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.

Value x_i:	0	1	2	3	4	5
Probability p_i:	0.23730	0.39551	0.26367	0.08789	0.01465	0.00098



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: $\mu_X = \sum x_i p_i = (0)(0.23730) + (1)(0.39551) + \cdots + (5)(0.00098) = 1.25$

So the expected number of children with type O blood in families like this one with 5 children is 1.25.

Spread: The variance of X is $\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i = (0 - 1.25)^2(0.23730) + (1 - 1.25)^2(0.39551) + \cdots + (5 - 1.25)^2(0.00098) = 0.9375$

So the standard deviation of X is $\sigma_X = \sqrt{0.9375} = 0.968$

The number of children with type O blood will typically differ from the mean by about 0.968 in families like this one with 5 children.

Value x_i:	0	1	2	3	4	5
Probability p_i:	0.23730	0.39551	0.26367	0.08789	0.01465	0.00098

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.

Ch 6.3 Day #2

Mean and Standard Deviation of a Binomial Random Variable

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

$$\mu = np$$

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

Note: ***These formulas work ONLY for binomial distributions. They can't be used for other distributions!***

Example 5: Mr. Bullard's AP Statistics class has 21 students. If we assume that the students in his class could *not* tell tap water from bottled water, then each one was basically guessing, with a 1/3 chance of being correct. Let X = the number of students who correctly identified the cup containing the different type of water.

a) Explain why X is a binomial random variable.

Binary: on each trial, "success" = correct identification; "failure" = incorrect identification.

Independent: one student's result should have no effect on any other student's result.

Number: there are $n = 21$ trials.

Success: each student has a $p = 1/3$ chance of guessing correctly.

Since X is counting the number of successes in this binomial setting, it is a binomial random variable.

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

The mean of X is $\mu_X = np = 21\left(\frac{1}{3}\right) = 7$

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

The standard deviation of X is $\sigma_X = \sqrt{np(1-p)} = \sqrt{21\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)} = 2.16$

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.

c) Of the 21 students in the class, 13 made correct identifications. Are you convinced that Mr. Bullard's students can tell bottled water from tap water? Justify your answer.

The class's result corresponds to $X = 13$, a value that's nearly 3 standard deviations above the mean. How likely is it that 13 or more of Mr. Bullard's students would guess correctly? It's $P(X \geq 13) = 1 - P(X \leq 12)$. Using the calculator's binomcdf (trials: 21, p:1/3, x value:12) command: $P(X \geq 13) = 1 - 0.9932 = 0.0068$.

The students have less than a 1% chance of getting so many right if they are all just guessing. This is strong evidence that some of the students in the class can tell bottled water from tap water.

Binomial Distributions in Statistical Sampling

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

Example 6: A supplier inspects an SRS of 10 flash drives from a shipment of 10,000 flash drives. Suppose that (unknown to the supplier) 2% of the flash drives in the shipment are defective. Count the number X of bad flash drives in the sample.

This is not quite a binomial setting. Removing 1 flash drive changes the proportion of bad flash drives remaining in the shipment. The conditional probability that the second flash drive chosen is bad changes when we know whether the first is good or bad. But removing 1 flash drive from a shipment of 10,000 changes the makeup of the remaining 9999 flash drives very little. The distribution of X is very close to the binomial distribution with $n = 10$ and $p = 0.02$. To illustrate this, let's compute the probability that none of the 10 flash drives is defective. Using the binomial distribution, it's $P(X=0) = \binom{10}{0}(0.02)^0(0.98)^{10} = 0.8171$

The actual probability of getting no defective flash drives is

$$P(\text{no defectives}) = \frac{9800}{10000} \times \frac{9799}{9999} \times \frac{9798}{9998} \times \ldots \times \frac{9791}{9991} = 0.8170$$

Those two probabilities are quite close!

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 $n \leq \frac{1}{10}N$.

Example 7: An airline has just finished training 25 first officers—15 male and 10 female—to become captains. Unfortunately, only eight captain positions are available right now. Airline managers decide to use a lottery to determine which pilots will fill the available positions. Of the 8 captains chosen, 5 are female and 3 are male. Explain why the probability that 5 female pilots are chosen in a fair lottery is *not*

$$P(X=5) = \binom{8}{5}(0.40)^5(0.60)^3 = 0.124$$

(The correct probability is 0.106.)

The managers are sampling without replacement when they do the lottery. There's a 0.40 chance that the first pilot selected for a captain position is female. Once that person is chosen, the probability that the next pilot selected will be female is no longer 0.40. The binomial formula assumes that the conditional probability of success stays constant at 0.40 throughout the eight trials of this chance process. This calculation will be approximately correct if the success probability doesn't change too much—as long as we don't sample more than 10% of the population. In this case, managers are sampling 8 out of 25 pilots—almost 1/3 of the population. That explains why the binomial probability is off by about 17% (0.018/0.106) from the correct answer.

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.

A **geometric setting** arises when we perform independent trials of the same chance process and record the number of trials it takes to get one success. On each trial, the probability p of success must be the same.

Binary? The possible outcomes of each trial can be classified as "success" or "failure."

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.

Trials? The goal is to count the number of trials until the first success occurs.

Success? On each trial, the probability p of success must be the same.

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 a **geometric distribution**.

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!*

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,

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

Lucky Day Game
Your teacher is planning to give you 10 problems for homework. As an alternative, you can agree to play the Lucky Day Game. Here's how it works. A student will be selected at random from your class and asked to pick a day of the week (for instance, Thursday). Then your teacher will use technology to randomly choose a day of the week as the "lucky day." If the student picks the correct day, the class will have only one homework problem. If the student picks the wrong day, your teacher will select another student from the class at random. The chosen student will pick a day of the week and your teacher will use technology to choose a "lucky day." If this student gets it right, the class will have two homework problems. The game continues until a student correctly guesses the lucky day. Your teacher will assign a number of homework problems that is equal to the total number of guesses made by members of your class. Are you ready to play the Lucky Day Game?

Example 8: The random variable of interest in this game is Y = the number of picks it takes to correctly match the lucky day. Each pick is one trial of the chance process.

a) Let's check the conditions for a geometric setting:

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.

b) Find the probability that the class receives exactly 10 homework problems as a result of playing the Lucky Day Game.

$P(Y = 10) = (6/7)^9(1/7) = 0.0357.$

c) Find $P(Y < 10)$ and interpret this value in context.

$P(Y < 10) = P(Y = 1) + P(Y = 2) + P(Y = 3) + \ldots + P(Y = 9) = 1/7 + (6/7)(1/7) + (6/7)^2(1/7) + \ldots + (6/7)^8(1/7) = 0.7503$. There's about a 75% chance that the class will get less homework by playing the Lucky Day Game.

Mean (Expected Value) of a Geometric Random Variable
If Y is a geometric random variable with probability of success p on each trial, then its **mean** (expected value) is $E(Y) = \mu_Y = \frac{1}{p}$.

That is, the expected number of trials required to get the first success is $\frac{1}{p}$.