The Binomial Setting

- 1. Each observation falls into one of just two categories, which for convenience we call "success" or "failure."
- 2. There is a fixed number of *n* observations.
- 3. The *n* observations are all **independent**. That is, knowing the outcome of one observation tells you nothing about the other observations.
- 4. The probability for success *p* is the same for each observation.

Binomial Distribution

The distribution of the count X of the successes in the binomial setting is the binomial distribution with parameters n and p. The parameter n is the number of observations, and p is the probability of a success on any one observation. The possible values of X are the whole numbers from 0 to n. As an abbreviation, we say that X is B(n, p).

Examples of Binomial Settings?

- Blood type is inherited. If both parents carry genes for O and A blood types, each child has probability 0.25 of getting two O genes and so of having blood type O. Different children inherently independently of each other. The number of O blood types among 5 children of these parents is the count X of successes in 5 independent observations with probability 0.25 of a success on each observation. So X has the binomial distribution with n = 5 and p = 0.25. We say that X is B(5, 0.25)
- 2. Deal 10 cards from a shuffled deck and count the number X of red cards. There are 10 observations, and each gives either a red or black card. A "success" is a red card. But the observations are not independent. If the first card is black, then there is a greater likelihood that the second card will be red. The count X DOES NOT have a binomial distribution.
- 3. An engineer chooses an SRS of 10 switches from a shipment of 10,000 switches. Suppose that (unknown to engineer) 10% of the switches in the shipment are bad. The engineer counts the number X of bad switches in the sample. *THIS IS NOT QUITE A BINOMIAL SETTING*. Just as removing one card in the previous example changes the makeup of the deck, removing one switch changes the proportion of bad switches remaining. So the state of the second switch is not independent of the first switch. BUT removing one switch from 10,000 changes the makeup of the remaining 9999 very little. In practice, the distribution of X is very close to the binomial distribution B(10, 0.1).

Sampling Distribution of a Count

Choose an SRS of size n from a population with a proportion p of successes. When the population is much larger than the sample, the count X of successes in the sample has approximately the binomial distribution with parameter n and p.

Example 4:

Engineers define reliability as the probability that an item will perform its function under specific conditions for a specified period of time. If an aircraft engine turbine has probability 0.999 of performing properly for an hour of flight, the number of turbines in a fleet of 350 engines that fly for an hour without failure has the B(350, 0.999) distribution. This binomial distribution is obtained assuming, as it seems reasonable, that the turbines fail independently of each other. A common cause of failure, such as sabotage, would destroy independence and make the binomial model inappropriate.

Binomial Formulas

We can find a formula doe the probability that a binomial random variable takes any value by adding probabilities for the different ways of getting exactly that many successes in n observations.

Example 5: Each child born to a particular set of parents has probability 0.25 of having blood type O. If these parents have 5 children, what is the probability that exactly 2 of them have type O blood? The count of children with type O blood is a binomial random variable X with n = 5 tries and probability p = 0.25 of a success on each try. We want P (X = 2).

Let's use "S" for success and "F" for failure.

Step 1:

Step 2:

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

The binomial coefficient counts the number of ways in which K successes can be distributed among n observations. The binomial probability P(X = k) is the count multiplied by the probability of any specific arrangement of the k successes.

Binomial Probability: If X has the binomial distribution with n observations and probability p of success on each observation, the possible values of X are 0, 1, 2, 3, ..., n. If k is any of these values,

Example 6:

The number X of switches that fail inspection in our previous example has approximately the binomial distribution with n = 10 and p = 0.1. The probability that no more than 1 switch fails is

Assignment: p. 516 8.1 to 8.6 p. 519-520 8.7 to 8.12

Finding Binomial Probabilities

Example 7: A quality engineer selects an SRS of 10 switches from a large shipment for detailed inspection. Unknown to the engineer, 10% of the switches in the shipment fail to meet the specifications. What is the probability that no more than 1 of the 10 switches in the sample fail inspection?

The count X of bad switches in the sample has approximately the B (10, 0.1) distribution. The distribution is strongly skewed. Although X can take any whole number value from 0 to 10, the probabilities of value larger than 5 are so small that they basically don't exist. We want to calculate:

When X is B (10, 0.1). The calculator command for this is **binompdf** (n,p,X). PDF stands for probability distribution function.

Type in both binompdf (10, 0.1, 0) and binompdf (10, 0.1, 1) and see what values you get.

Example 8: Corrine is a basketball player who makes 75% of her free throws. In a key game, Corrine shoots 12 free throws and makes 7 of them. Is it unusual for her to perform this poorly? To answer this question, assume free throws are independent with a probability of 0.75 on each shot. The number X of baskets (successes) in 12 attempts has the B(12, 0.75) distribution.

We want the probability of making a basket on at most 7 free throws.

Cumulative Binomial Probability: given a random variable X, the cdf of X calculates the sum of the probabilities for 0, 1, 2, ..., up to the value of X. That is, it calculates the probability of obtaining at most X successes in n trials.

Look back at the switches example, you could have just done binomialcdf (10, 0.1, 1) and it will automatically calculate the probability X = 1 and X = 0 and add them together.

Assignment: p. 523-524 8.13 to 8.18

Binomial Mean and Standard Deviation

Formula for mean and standard deviation of a binomial random variable:

Continuing with the switches in examples 6 and 7, the count X of bad switches is binomial with n = 10 and p = 0.1. The mean and standard deviation of the binomial distribution are

The Normal Approximation to Binomial Distributions

As the number of trials n gets larger, the binomial distribution gets close to a normal distribution. When n is large, we can use Normal probability calculations to approximate hard to calculate binomial probabilities.

Example 12: Are attitudes toward shopping changing? Sample surveys show that fewer people enjoy shopping than in the past. A survey asked a nationwide random sample of 2500 adults if the agreed or disagreed that "I like buying new clothes, but shopping is often frustrating and time consuming." The population that the poll wants to draw conclusions about is all US residents ages 18 and over. Suppose that in fact 60% of all adult US residents would say "Agree" if asked the same question. What is the probability that 1520 or more of the sample agree?

Because there are more than 218 million adults, we can take the responses to be independent. So the number that agree is a random variable X having the binomial distribution with n = 2500 and p = 0.16. To find the probability that at least 1520 of the people in the sample find shopping frustrating, we must add the binomial probabilities of all outcomes from X = 1520 to X = 2500. This isn't practical.

a. We can use our calculator to find the probability that X is less than or equal to 1519 and subtract from 1. What answer does this give?

b. We can also use a normal curve with the same mean and standard deviation as our binomial distribution. What would our mean and standard deviation have to be?

If we act as though the count X has the N(1500, 24.49) distribution, here is the probability we want, using table A:

How much does the Normal approximation differ from the result we got with the calculator?

Normal Approximation for Binomial Distributions: Suppose that a count X has the binomial distribution with n trials and success probability p. When n is large, the distribution of X is approximately normal, $N^{(np,\sqrt{np(1-p)})}$. As a rule of thumb, we will use the Normal approximation when n and p satisfy $np \ge 10$ and $n(1-p) \ge 10$.

Assignment: p. 529-530 8.19 to 8.24

Simulating Binomial Events

Example 14: Back to Corrine and her free throws. Remember that she has a 75% record with free throws. In our particular game, she made 7 of 12 free throws. We wanted to know how unusual it was for her to only make 7 of her 12 free throws. The last time we worked this problem, we calculated the binomial probability to be 0.1576. Now we are going to use the calculator to do a simulation of this situation. Let X = the number of hits in 12 attempts. The probability of success is 0.75. We will assign the number 1 to be success (hit) and 0 to be failure (miss).







In the long run, this random function will select 1 75% of the time and 0 25% of the time. If we repeated this process several times and counted the number of times Corrine made 7 or fewer shots, that would give an estimate of the probability that X is less than 7.

We can make this go faster by placing the entries in L1 and then find the sum of L1.



Press ENTER 10 times to simulate shooting 12 shots on 10 different occasions. How many numbers are there less than or equal to 7? This gives you an estimate of the proportion of times that she should only make 7 out of 12 shots. As you continue to increase your simulations, the proportion should get closer to 0.1576 (Law of Large Numbers). Simulate 50 games and see if you get a proportion close to 0.1576.

How could you make a binomial histogram on your calculator?

Make a histogram to represent the switches problem. B (10, .1)

Assignment: p. 534-535 8.25 to 8.28 8.1 Review Questions p. 536-538 8.31, 8.33, 8.34, 8.37, 8.38, 8.39

Geometric Distribution

The geometric Setting:

1. Each observation falls into one of just two categories, which we call success and failure.

- 2. Observations are all independent
- 3. Probability of success (p) is the same for each observation
- 4. The variable of interest is the number of trials required to obtain the first success.

Example 15: A game consists of rolling a single die. The event of interest is rolling a 3; this event is called a success. The random variable is defined as X = the number of trials until a 3 occurs. Is this a geometric setting?

Success =

Independence?

Probability of rolling a 3 on each roll =

Variable of interest?

Example 16: Suppose you repeatedly draw cards without replacement from a deck of 52 cards until you draw an ace. There are two categories of interest: ace = success, not ace = failure. Is this a geometric setting?

Success =

Independence?

Probability of drawing an ace =

Variable of interest?

Rule for calculating Geometric Probabilities

If X has a geometric distribution with probability p of success and (1 - p) of failure on each observation, the possible values of X are 1, 2, 3, ... If n is any one of these values, the probability that the first success occurs on the nth trial is

a.

b.

On your calculator

geometpdf (p, n) $(2^{nd}/Dist/D)$

geometcdf(p, n) $(2^{nd}/Dist/E)$

Example 17: Construct a probability distribution table for X = number of rolls of a die until a 3 occurs:

\mathbf{v} .	1	^	2	Λ	E	6	
A.:		2	.)	4		0	1
	-	_	U	-	U	<u> </u>	

P(X):

Find:
$$P(X = 3)$$
 $P(X \le 2)$

$$\mathsf{P}(1 \le \mathsf{X} \le 4) \qquad \qquad \mathsf{P}(\mathsf{X} > 6)$$

Mean and Standard deviation of Geometric Distribution

Mean (expected value):

Variance (take square root to get standard deviation):

Example 18: Glenn likes the game at the state fair where you toss a coin into a saucer. You win if the coin some to rest in the saucer without sliding off. Glenn has played this game many times and determined that on average 1 out of every times he plays he wins. He believes that his chances of winning are the same for each toss. Each play is independent. Let X be the number of tosses until a win. Glenn believes that this describes a geometric setting.

p =

Variance of X:

Simulating a Geometric Situation

Cheerios has advertised that they have a "free \$1 bill in every 20th box." Perform a simulation to determine the number of boxes of Cheerios that you would need to purchase in order to get one of the "free" dollar bills.

How many boxes did it take for you to "win"?

How many boxes would you expect to buy to win?

What is the variance and st. dev. in this situation?

Assignment: p. 543-544 8.41, 8.42 p. 550 8.45, 8.47, 8.49, 8.50 p. 552-553 8.52, 8.54

Chapter Review: p. 556-558 8.59, 8.60, 8.61, 8.64, 8.67