Chapter 11

Bernoulli Trials

11.1 The Binomial Distribution

In the previous chapter, we learned about i.i.d. trials. Recall that there are three ways we can have i.i.d. trials:

1. Our units are trials and we have decided to assume that they are i.i.d.

2. We have a finite population and we will select our sample of its members at random with replacement—the dumb form of random sampling. The result is that we have i.i.d. random variables which means the same thing as having i.i.d. trials.

3. We have a finite population and we have selected our sample of its members at random without replacement—the smart form of random sampling. If n/N—the ratio of sample size to population size—is 0.05 or smaller, then we get a good approximation if we treat our random variables as i.i.d.

In this chapter, we study a very important special case of i.i.d. trials, called **Bernoulli trials**. If each trial has exactly two possible outcomes, then we have Bernoulli trials. For convenient reference, I will now explicitly state the assumptions of Bernoulli trials.

Definition 11.1 (The assumptions of Bernoulli trials.) *If we have a collection of trials that satisfy the three conditions below, then we say that we have Bernoulli trials.*

1. *Each trial results in one of two possible outcomes, denoted success (S) or failure (F).*

2. *The probability of a success remains constant from trial-to-trial and is denoted by p. Write $q = 1 - p$ for the constant probability of a failure.*

3. *The trials are independent.*

We will use the method described on page 168 of Chapter 8 to assign the labels success and failure. When we are involved in mathematical arguments, it will be convenient to represent a success by the number 1 and a failure by the number 0. Finally,

We are not interested in either of the *trivial cases* in which $p = 0$ or $p = 1$. Thus, we restrict attention to situations in which $0 < p < 1$.

One reason that Bernoulli trials are so important, is that if we have Bernoulli trials, we can calculate probabilities of a great many events. Our first tool for calculation is the multiplication rule that we learned in Chapter 10. For example, suppose that we have $n = 5$ Bernoulli trials with $p = 0.70$. The probability that the Bernoulli trials yield four successes followed by a failure is:

$$P(SSSSF) = ppppq = (0.70)^4(0.30) = 0.0720.$$

Our next tool is extremely powerful and very useful in science. It is the **binomial probability distribution**. Suppose that we plan to perform/observe n Bernoulli trials. Let X denote the total number of successes in the n trials. The probability distribution of X is given by the following equation.

$$P(X = x) = \frac{n!}{x!(n-x)!}p^x q^{n-x}, \text{ for } x = 0, 1, \ldots, n. \tag{11.1}$$

Equation 11.1 is called the **binomial probability distribution** with **parameters** n and p; it is denoted by the Bin(n, p) distribution. I will illustrate the use of this equation below, a compilation of the Bin(5,0.60) distribution. I replace n with 5, p with 0.60 and q with $(1 - p) = 0.40$. Below, I will evaluate Equation 11.1 six times, for $x = 0, 1, \ldots, 5$: You should check a couple of the following computations to make sure you are comfortable using Equation 11.1, but you don't need to verify all of them.

$$P(X = 0) = \frac{5!}{0!5!}(0.60)^0(0.40)^5 = 1(1)(0.01024) = 0.01024.$$

$$P(X = 1) = \frac{5!}{1!4!}(0.60)^1(0.40)^4 = 5(0.60)(0.0256) = 0.07680.$$

$$P(X = 2) = \frac{5!}{2!3!}(0.60)^2(0.40)^3 = 10(0.36)(0.064) = 0.23040.$$

$$P(X = 3) = \frac{5!}{3!2!}(0.60)^3(0.40)^2 = 10(0.216)(0.16) = 0.34560.$$

$$P(X = 4) = \frac{5!}{4!1!}(0.60)^4(0.40)^1 = 5(0.1296)(0.40) = 0.25920.$$

$$P(X = 5) = \frac{5!}{5!0!}(0.60)^5(0.40)^0 = 1(0.07776)(1) = 0.07776.$$

Whenever probabilities for a random variable X are given by Equation 11.1 we say that X has a binomial probability (sampling) distribution with parameters n and p and write this as $X \sim$ Bin(n, p).

There are a number of difficulties that arise when one attempts to use the binomial probability distribution. The most obvious is that each trial needs to give a dichotomous response. Sometimes it is obvious that we have a dichotomy: For example, if my trials are shooting free throws or attempting golf putts, then the natural response is that a trial results in a make or miss. Other

times, the *natural response* might not be a dichotomy, but the response of interest is. For example, in my example of the American roulette wheel in Chapter 10, the *natural response* is the winning number, but if I like to bet on red, then the response of interest has possible values *red* and *not red*. Similarly, in the game of craps, I might be primarily interested in whether or not my *come out* results in a *pass line win*, a dichotomy.

Thus, let's suppose that we have a dichotomous response. The next difficulty is that in order to calculate probabilities, we need to know the numerical values of n and p. Almost always, n is known to the researcher and if it is unknown, we might be able to salvage something by using the Poisson distribution, which you will learn about in Chapter 13. There are situations in which p is known, including the following:

- Mendelian inheritance; my roulette example above, *assuming the wheel is fair*; and my craps example, *assuming both dice are fair and they behave independently*.

In other words, sometimes I feel that the phenomenon under study is sufficiently well understood that I feel comfortable in my belief that I know the numerical value of p. Obviously, in many situations I won't know the numerical value of p. For shooting free throws or attempting golf putts I usually won't know exactly how skilled the player/golfer is. When my interest is in a finite population, typically I won't know the composition of the population; hence, I won't know the value of p.

Before I explore how to deal with the difficulty of p being unknown, let's perform a few computations when p is known.

Example 11.1 (Mendelian inheritance with the 3:1 ratio.) *If you are unfamiliar with Mendelian inheritance, you can find an explanation of the 1:2:1, 3:1 and 9:3:3:1 ratios at:*

`http://en.wikipedia.org/wiki/Mendelian_inheritance`.

Each trial—offspring—will possess either the dominant (success) or recessive (failure) phenotype. Mendelian inheritance tells us that these will occur in the ratio 3:1, which means that $p = 3q = 3(1-p)$. *Solving for* p*, this equation becomes* $4p = 3$ *and, finally,* $p = 0.75$. *Suppose that we will observe* $n = 8$ *independent trials, each with* $p = 0.75$. *Let* X *denote the total number of these eight offspring who will possess the dominant phenotype. We can calculate probabilities for* X *by using Equation 11.1 with* $n = 8$ *and* $p = 0.75$*; i.e., by using the Bin(8,0.75) probability distribution.*

Suppose, for example, that I want to calculate $P(X \geq 6)$ for X having the Bin(8,0.75) distribution. If I want to do this by hand, I must first rewrite my event of interest:

$$P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8).$$

Next, I must evaluate Equation 11.1 *three times*; for $x = 6$, $x = 7$ and $x = 8$. Sorry, but this does not sound like fun!

Fortunately, there is a website that can help. Go to:

`http://stattrek.com/Tables/Binomial.aspx`

Table 11.1: The output, after rounding, from the binomial website for $p = 0.75$, $n = 8$ and $x = 6$.

Binomial Probability:	$P(X = 6) = 0.3115$
Cumulative Probability:	$P(X < 6) = 0.3215$
Cumulative Probability:	$P(X \leq 6) = 0.6329$
Cumulative Probability:	$P(X > 6) = 0.3671$
Cumulative Probability:	$P(X \geq 6) = 0.6785$

I will now show you how to use this website. The website requires you to enter three numbers:

- **Probability of success on a single trial**: Enter the value of p; for our current problem, I enter $p = 0.75$.
- **Number of trials**: Enter the value of n; for our current problem, I enter $n = 8$.
- **Number of successes (x)**: This is a bit tricky to explain explicitly, but once you see one example, you will understand how to do it. Because my event of interest, $(X \geq 6)$, involves the number 6, I enter $x = 6$.

After entering my values for p, n and x, I click on the *Calculate* box and obtain the output, rounded to four digits, printed in Table 11.1. The answer I want is the fifth entry in the list:

$$P(X \geq 6) = 0.6785.$$

Note that there is a great deal of redundancy in five answers in this list. Make sure you understand why the following identities are true:

- The third probability is the sum of the first two.
- The fifth probability is the sum of the first and the fourth.
- The sum of the second and the fifth probabilities equals 1.
- The sum of the third and the fourth probabilities equals 1.

For example, the third probability $P(X \leq 6)$ can be written as $P(X < 6) + P(X = 6)$. In the listing above, this becomes $0.6329 = 0.3115 + 0.3215$ which is correct except for round-off error.

The website is good for computing individual probabilities, but it is tedious to use it to generate an entire binomial distribution. For the latter objective, I use Minitab. In particular, with the help of Minitab, I obtained the Bin(8,0.75) distribution, displayed in Table 11.2. Literally, the first two columns of this table present the sampling distribution. It's easy to have the computer create the cumulative sums in the third and fourth columns, so I have included them. From the table we can find the five probabilities given by the website. For example,

$$P(X > 6) = P(X \geq 7) = 0.3671, \text{ from the fourth column.}$$

Table 11.2: The binomial distribution with $n = 8$ and $p = 0.75$.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
0	0.0000	0.0000	1.0000
1	0.0004	0.0004	1.0000
2	0.0038	0.0042	0.9996
3	0.0231	0.0273	0.9958
4	0.0865	0.1138	0.9727
5	0.2076	0.3215	0.8862
6	0.3115	0.6329	0.6785
7	0.2670	0.8999	0.3671
8	0.1001	1.0000	0.1001
Total	1.0000	—	—

11.1.1 Computational Difficulties

By trial-and-error, I discovered that if I go to Minitab and ask for the Bin(n,0.50) distribution with $n \geq 1023$, then I am given the following message:

> * ERROR * Completion of computation impossible.

If, however, $p = 0.50$ and $n \leq 1022$, then Minitab gives an answer. Similarly, Minitab reports its error message for Bin(n,0.60) if, and only if, $n \geq 1388$. The people who wrote Minitab are good programmers. I am not a very good programmer, but I *could* write a program that would handle at least some of the situations Minitab does not. How can this be? Well, as we will learn later, if n is large enough, then we can use a either a Normal curve or the Poisson distribution (see Chapter 13) to obtain good approximations to binomial probabilities. Thus, my inference is that the Minitab programmers were somewhat *casual* in writing their code because they knew that their users could opt for an approximation.

If you read through the exposition on the website

`http://stattrek.com/Tables/Binomial.aspx`,

near the bottom you will find the following:

> When the number of trials is greater than 1,000, the Binomial Calculator uses a Normal distribution to estimate the binomial probabilities.

Very soon I will give you the details of the Normal curve approximation.

Here is my advice for this course. You can trust the website's answers provided $n \leq 1000$. Do not use it for $n > 1000$ until you have read my discussion of the Normal approximation in the next section.

Figure 11.1: The Bin(100, 0.5) Distribution.

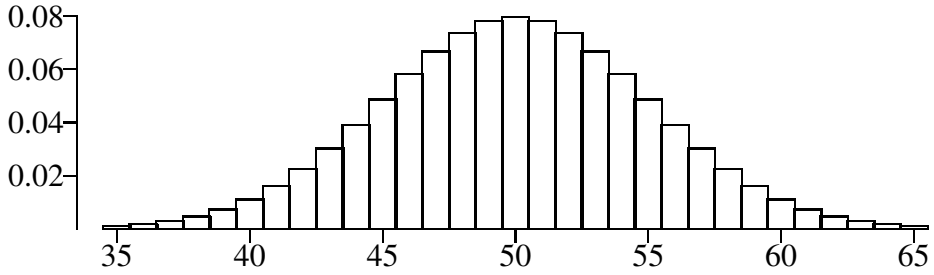


Figure 11.2: The Bin(100, 0.2) Distribution.

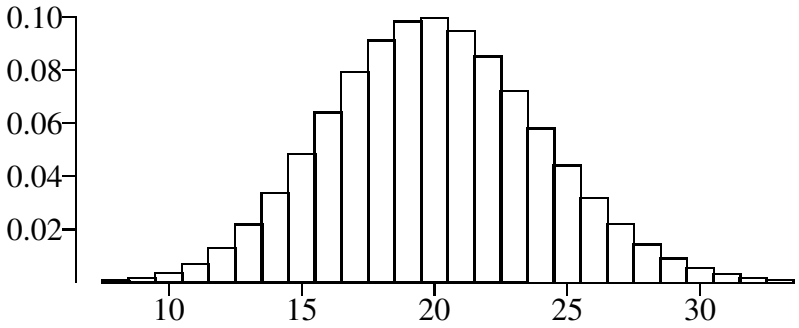


Figure 11.3: The Bin(25, 0.5) Distribution.

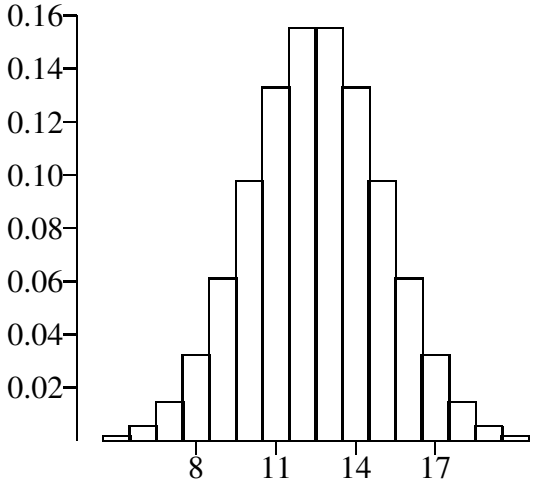
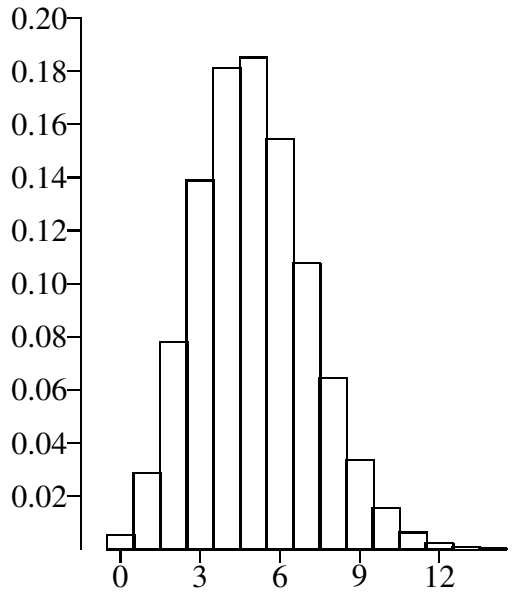


Figure 11.4: The Bin(50, 0.1) Distribution.



11.2 The Normal Curve Approximation to the Binomial

Recall that we learned how to draw a probability histogram on page 143 in Chapter 7. Figures 11.1–11.4 present **probability histograms** for several binomial probability distributions. Because $\delta = 1$ the area of each rectangle equals its height; thus, the probability of any integer value of x is the height of the rectangle centered at x.

As discussed in Chapter 7, a probability histogram allows us to 'see' a probability distribution. For example, for the four probability histograms that are presented above, the two with $p = 0.50$ are symmetric; the one with $n = 100$ and $p = 0.2$ is *almost* symmetric; and the one with $n = 50$ and $p = 0.1$ deviates a great deal from symmetry. Indeed, it can be shown that a binomial distribution is symmetric if, and only if, $p = 0.50$. Moreover, for $p \neq 0.5$, if both np and nq are *far from 0* then the binomial distribution is almost symmetric. A common guideline for *far from 0* is for both to be at least 25. We will return to this topic soon.

Below is a list of some other facts about binomial distributions.

1. The probability histogram for a binomial always has exactly one peak. The peak can be one or two rectangles wide, but never wider.

2. If np is an integer, then there is a one-rectangle wide peak located above np.

3. If np is not an integer, then the peak will occur either at the integer immediately below or above np; or, in some cases, at both of these integers.

4. If you move away from the peak in either direction, the heights of the rectangles become shorter. If the peak occurs at either 0 or n this fact is true in the one direction away from the

peak.

The following result is similar to Results 7.1–7.3 for the sum of ranks test in Chapter 7.

Result 11.1 (Mean and standard deviation of the binomial distribution.) *The mean and standard deviation of the Bin(n, p) distribution are:*

$$\mu = np \tag{11.2}$$

$$\sigma = \sqrt{npq} \tag{11.3}$$

Let's consider the Bin(100,0.50) distribution, pictured in Figure 11.1. From the above result, its mean and standard deviation are

$$\mu = np = 100(0.50) = 50 \text{ and } \sigma = \sqrt{npq} = \sqrt{100(0.50)(0.50)} = 5.$$

Suppose now that I want to compute $P(X \geq 55)$. I have three methods for obtaining this probability:

1. Because $n \leq 1000$ I can use the website

   `http://stattrek.com/Tables/Binomial.aspx`

   I go to the site, enter $p = 0.50$, $n = 100$ and $x = 55$; then I click on *Compute* and obtain the answer:

   $$P(X \geq 55) = 0.1841.$$

2. Because $n \leq 1022$, I can use Minitab. I did and obtained the answer 0.1841.

3. I can follow the method of Chapter 7 and obtain a Normal curve approximation. I go to the website:

   `http://davidmlane.com/hyperstat/z_table.html`

   I enter Mean $= 50$ and Sd $= 5$. Next to the option *Above* I enter 54.5—remember the continuity correction. The site tells me that the area under the N(50,5) curve to the right of 54.5 is 0.1841. To the nearest 0.0001, this approximation is exact!

Let me do another computational example. Consider the Bin(1200,0.60) distribution. I am interested in $P(X \leq 690)$. Again, I will try three methods for finding this probability.

1. Because $n \leq 1387$, I can use Minitab to find the exact probability. I did and obtained the answer 0.0414.

2. I calculate

   $$\mu = np = 1200(0.60) = 720 \text{ and } \sigma = \sqrt{npq} = \sqrt{1200(0.60)(0.40)} = 16.971.$$

   I go to the website:

`http://davidmlane.com/hyperstat/z_table.html`

I enter Mean $= 720$ and Sd $= 16.971$. Next to the option *Below* I enter 690.5—remember the continuity correction. The site tells me that the area under the N(720,16.971) curve to the left of 690.5 is 0.0411. If I trust Minitab, my approximate answer is too small by 0.0003. This is a very good approximation!

3. I can use the *binomial website*:

   `http://stattrek.com/Tables/Binomial.aspx`

   Because $n > 1000$, the website will give me an answer based on the Normal curve approximation. I entered $p = 0.60$, $n = 1200$, $x = 690$ and clicked on *Calculate*. The site gave me the answer 0.041. I would prefer the site to give me more digits of precision, but to the nearest 0.001, this answer agrees with my Normal curve approximation. And, of course, the binomial website is less work for me.

You might be wondering:

Is the Normal curve approximation always good?

I will give you a dramatic example of why the answer is **no!**

Let's consider the Bin(1001,0.001) distribution. I am choosing $n =$ 1,001 because this is the smallest value of n for which the *binomial website* uses the Normal curve to obtain approximate binomial probabilities. I am interested in $P(X = 0)$. Again, I will show you three methods.

1. Minitab gives me the exact probability, 0.3673.

2. I can calculate the exact probability using Equation 11.1:

$$P(X = 0) = \frac{1001!}{0!1001!}(0.001)^0(0.999)^{1001}.$$

   After canceling the factorials and noting that $(0.001)^0 = 1$, this answer reduces to

$$(0.999)^{1001}.$$

   Even the calculator on my cell phone can evaluate this! It obtains the correct answer, 0.3673.

3. I went to the *binomial website*, entered $p = 0.001$, $n =$ 1,001, $x = 0$ and clicked on *Calculate*. The website did not calculate an answer; instead, it printed the message:

   PROCESSING ERROR: Cannot complete computation.

   This is good; the program is smart enough not to use the Normal approximation. Sadly, however, I have bad news to report. Staying on the site, I calculated probabilities for $n =$

1,001 and $p = 0.5$; the site worked fine. Then I reentered $p = 0.001$ $x = 0$ and $n =$ 1,001; this time, the site gave me:

$$P(X = 0) = 0.025 \text{ and } P(X < 0) = 0.217.$$

Because X counts successes, it cannot be negative!

Thus, the kindest thing I can say is that for $n > 1000$ the site's behavior is erratic; do not use it unless you check that both np and nq equal or exceed 25.

In summary, here are two general guidelines I recommend you use.

1. **For the purpose of computing probabilities:** If $n \leq 1000$ use the binomial website. If $n > 1000$, $np \geq 25$ and $nq \geq 25$: you may use the binomial website or obtain the Normal curve approximation *by hand*.

2. **For the development of estimation and prediction intervals:** Use the Normal approximation to the binomial only if both np and nq are greater than or equal to 25.

Let me make a few comments on these guidelines. First, all statisticians agree that we need to consider the values of np and nq; not all would agree on my magic threshold of 25. Second, if $n > 1000$ and, say, $np < 25$ we can use the Poisson distribution to approximate the binomial; this material will be presented in Chapter 13.

Third, the second guideline is a bit odd. It implies, for example, that for $n = 50$ and $p = q = 0.50$ we **may use** the Normal curve approximation even though exact probabilities are readily available from the website. As you will learn in Chapter 12, being able to use the Normal curve approximation is very helpful for the development of general formulas.

11.3 Calculating Binomial Probabilities When p is Unknown

I could make this a very short section by simply remarking that if p is unknown then obviously neither the website, Minitab nor I can evaluate Equation 11.1. In addition, if p is unknown then we can calculate neither the mean nor standard deviation of the binomial, both of which are needed for the Normal curve approximation.

We do have this section, however, because I want to explore the idea of what if means to *know the value of* p. I am typing this in October, 2013. To date in his NBA (National Basketball Association) career, during the regular season Kobe Bryant has made 7,932 free throws out of 9,468 attempts. What can we do with these data in the context of what we have learned in this chapter?

I am always interested in computing probabilities; thus, when faced with a new situation I ask myself whether it is reasonable to assume a structure that will allow me to do so. Well, each free throw attempted by Bryant can be viewed as a trial, so I might assume that his 9,468 attempts were observations of i.i.d. trials. As a side note, let me state that years ago I was active in the *Statistics in Sports* section of the American Statistical Association. We had many vigorous debates—and many papers have been written—on the issue of whether the assumption of i.i.d. trials is reasonable in

sports in general, not just for free throws in basketball. In order to avoid a digression that could consume months, if not years, let's tentatively assume that we have i.i.d. trials for Bryant shooting free throws.

The next issue is the value of p for Bryant's trials. Bryant shooting a free throw was not as simplistic as, say, tossing a fair coin or casting a balanced die; nor was it as well-behaved as Mendelian inheritance. In short, there is no reason to believe that we **know** the value of p for Bryant or, indeed, any basketball player. But what do I mean by *know*? We know that p is strictly between 0 and 1. As any mathematician will tell you, *there are a lot of numbers between 0 and 1*! (The technical term is that there are an uncountable infinity of numbers between 0 and 1.) But we need to think more like a scientist and less like a mathematician. In particular, by scientist I mean someone who is—or strives to be—knowledgeable about basketball. A mathematician will (correctly) state that 0.7630948 and 0.7634392 are different numbers, but as a basketball fan, I don't see any *practical difference* between p equaling one or the other of these. In either situation, I would round to three digits and say, "The player's true ability, p, is that in the long-run he/she makes 76.3% of attempted free throws."

Bryant's data give us

$$\hat{p} = 7932/9468 = 0.838;$$

in words, during his career, to date, Bryant has made 83.8% of his free throws. As might be clear—and if not, we will revisit the topic in the next chapter—we may calculate the nearly certain interval (Formula 4.1 in Chapter 4) for p:

$$0.838 \pm 3\sqrt{\frac{0.838(0.162)}{9468}} = 0.838 \pm 0.011 = [0.827, 0.849].$$

Thus, given our assumption of i.i.d. trials, we don't know Bryant's p exactly, but every number in its nearly certain interval is quite close to his value of $\hat{p}$.

Thus, if I wanted to compute a probability for Bryant, I would be willing to use $p = 0.838$.

Let's consider the first $n = 50$ free throws that Bryant will attempt during the 2013–2014 NBA season. I am interested in the number of these free throws that he will make; call it X. Based on my discussion above I view X as having the Bin(50,0.838) distribution.

For example, I go to the website

`http://stattrek.com/Tables/Binomial.aspx`,

and enter $p = 0.838$, $n = 50$ and $x = 38$. I click on *Calculate* and obtain

$$P(X \geq 38) = 0.9482.$$

Thus, I believe that the probability that Kobe Bryant will make at least 38 of his first 50 free throws this season is just under 95%.

(**Optional enrichment for basketball fans.** It can be shown that the probability of the event I examine above, $(X \geq 38)$ is an increasing function of p. I found that for $p = 0.838$, this probability is 0.9482. If I use the lower bound of my nearly certain interval, $p = 0.827$, the website gives me 0.9201 for the probability of this event. If I use the upper bound of my nearly certain interval,

Table 11.3: The conference of the Super Bowl winner, by year. Year 1 denotes the 1966 season (game played in 1967) and year 47 denotes the 2012 season (game played in 2013). An NFC winner is labeled a success, denoted by 1, and an AFC winner is denoted by 0.

Year:	1	2	3	4	5	6	7	8	9	10	11	12
Winner:	**1**	**1**	0	0	0	**1**	0	0	0	0	0	**1**
Year:	13	14	15	16	17	18	19	20	21	22	23	24
Winner:	0	0	0	**1**	**1**	0	**1**	**1**	**1**	**1**	**1**	**1**
Year:	25	26	27	28	29	30	31	32	33	34	35	36
Winner:	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	**1**	0	0
Year:	37	38	39	40	41	42	43	44	45	46	47	
Winner:	**1**	0	0	0	0	**1**	0	**1**	**1**	**1**	0	

$p = 0.849$, the website gives me 0.9684 for the probability of this event. Thus, another approach is for me to say that I am nearly certain that the probability of the event $(X \geq 38)$ is between 0.9201 and 0.9684.)

11.4 Runs in Dichotomous Trials

There have been 47 Super Bowls played. Each game resulted in a winner from (using current names) the National Football Conference (NFC), which I will call a success, or the American Football Conference (AFC), which I will call a failure. The 47 outcomes, in time-order, are presented in Table 11.3. The outcome, *conference of the Super Bowl winner* is a dichotomous response. I am **not** assuming that they are the outcomes of 47 Bernoulli trials. In fact, later in this section I will conclude that they don't seem to be the result of observing Bernoulli trials. I am getting ahead of myself.

Our Super Bowl data contains 18 runs which are detailed (and implicitly defined) in Table 11.4. There is a great deal of information in this table. You don't need to study it exhaustively, but you should understand how it was constructed, which I will now explain.

The responses for years 1 and 2 are successes, but year 3 is a failure. Thus, the data begins with a run of successes of length 2 covering years 1 and 2. Next, the data has a run of failures of length 3 covering years 3–5. And so on.

In addition to noting that there are 18 runs in the Super Bowl data, it's difficult to miss the existence of the very long run of successes, 13, spanning years 19–31.

In general, any or all of of the following statistics may be used to investigate the issue of whether the data are the result of observing Bernoulli trials.

- The number of runs;

Table 11.4: The 18 runs for the data in Table 11.3.

Run:	1	2	3	4	5	6	7	8	9
Year(s):	1–2	3–5	6	7–11	12	13–15	16–17	18	19–31
Length:	2	3	1	5	1	3	2	1	13
Type:	S	F	S	F	S	F	S	F	S
Run:	10	11	12	13	14	15	16	17	18
Year(s):	32–33	34	35-36	37	38–41	42	43	44–46	47
Length:	2	1	2	1	4	1	1	3	1
Type:	F	S	F	S	F	S	F	S	F

- The length of the longest run of successes; and
- The length of the longest run of failures.

In particular:

- Does it appear that there is a constant probability of success from trial-to-trial?
- Do the trials appear to be independent?

This topic is frustrating for a number of reasons. In part because it is so frustrating, this topic typically is not presented in an introductory Statistics class. I will, however, discuss these issues in this section briefly for the following reasons.

1. I feel that I am doing you a disservice if I state assumptions without giving any idea of how to investigate their validity.
2. I feel that I am doing you a disservice if I present you with a *sanitized view* of Statistics; a view in which there are no controversies and no confusion about how to analyze data.
3. In my experience, this is one of the topics in Statistics that non-statisticians find very interesting. Especially the occurrence of long runs of successes (or failures) are interesting to people.

Let me now define some ideas rather formally. I plan to observe a total of n dichotomous trials. I want to investigate whether the dichotomous trials satisfy the assumptions of Bernoulli trials. I decide to pursue this by conducting a test of hypotheses. My null hypothesis is that the trials are Bernoulli trials. This is in line with the principle of Occam's Razor. Also, usually a researcher *wants* to have Bernoulli trials because Bernoulli trials allow the computation of many answers. Of course, there are some researchers—and I often find myself in this camp—who sometimes are hoping **not** to have Bernoulli trials because sometimes it is nice to live in a world that is a bit more complicated.

Please allow me a brief digression. Many textbooks state that the *alternative always represents what the researcher is trying to prove*. Often times, this is a valid view of the hypotheses, **but not always**. For example, in this current section, I must assume that the null is Bernoulli trials because otherwise there is no way to find a sampling distribution and so on; it doesn't really matter what I prefer to be true!

Wait a minute. You might be thinking—even though we don't yet have a test statistic:

> We need to be able to determine the sampling distribution of the test statistic **under the assumption that the null hypothesis is true**. This is hopeless! The null hypothesis does not specify the value of p for the Bernoulli trials; thus, it will be impossible to calculate a unique set of probabilities!

If these are your thoughts, then you are correct. Well, almost correct. The trick is that we use a *conditional* test statistic. Let me explain.

Let's look at the Super Bowl example again. The total number of trials in the data set is $n = 47$. Before collecting data I didn't know what the value of X, the total number of successes in the 47 trials, would be. After I collect the data I know that the observed value of X is $x = 25$. The trick is that I condition on X eventually being 25. Let me make the following points.

1. Most (nearly all?) statisticians feel fine about this conditioning; here is why. Given that $X = 25$—and therefore that the number of failures, $n - X = 47 - X$ is 22—what have I learned? I have learned that over the course of the data collection, neither conference had a huge advantage over the other. **But**, and this is the key point, knowing that $X = 25$ gives me no information about whether p is changing or whether the trials are independent. In other words, knowing that $X = 25$ gives me no information about whether the assumptions of Bernoulli trials are reasonable.

2. This point is a little more esoteric. I have always extolled you to remember: probabilities are calculated before we collect data. Conditioning on $X = 25$ and then calculating probabilities—as I will soon do—looks a lot like I am violating my directive. But I am not. Actually, before collecting data I can imagine 48 different computations, one for each of the 48 possible values of X (0, 1, 2, ..., 47). If I were to perform all of these computations, after I collect data I would find that my computations conditional on $X = x$ would be irrelevant for all $x \neq 25$. Why should I perform computations that I won't use?

I will now explain why conditioning is so useful mathematically for our problem. Consider the Super Bowl data again. Conditional on knowing $X = 25$, we know that the data will consist of an arrangement of 25 1's and 22 0's. The number of such arrangements is:

$$\frac{47!}{25!22!} = 1.483 \times 10^{13},$$

almost 15 trillion. It can be shown that, on the assumption that the null hypothesis is true, these arrangements are equally likely to occur, regardless of the value of p! Thus, if we choose our test statistic to be a function of the arrangement of 1's and 0's, then we can compute its sampling distribution **without knowing the value of** p.

Table 11.5: All possible arrangements of three 1's and two 0's. For each arrangement, the observed values of R, V and W are given, where R is the number of runs, V is the length of the longest run of successes, and W is the length of the longest run of failures. On the assumption that the data come from Bernoulli trials, conditional on obtaining three successes, the 10 arrangements are equally likely to occur.

Arrangement	r	v	w	Arrangement	r	v	w
11100	2	3	2	10011	3	2	2
11010	4	2	1	01110	3	3	1
11001	3	2	2	01101	4	2	1
10110	4	2	1	01011	4	2	1
10101	5	1	1	00111	2	3	2

Table 11.6: The sampling distributions of R, V and W for the 10 equally likely arrangements in Table 11.5.

r:	2	3	4	5	Total
$P(R = r)$:	0.2	0.3	0.4	0.1	1.0
v:	1	2	3		Total
$P(V = v)$	0.1	0.6	0.3		1.0
w:	1	2			Total
$P(W = w)$	0.6	0.4			1.0

Admittedly, 15 trillion is a lot of arrangements to visualize! To help make this result more concrete, I will do a quick example with $n = 5$ and conditioning on $X = 3$; see Table 11.5. I will now explain the information in this table.

With five trials, conditional on a total of three successes and two failures, the number of possible arrangements is:

$$\frac{5!}{3!2!} = \frac{5(4)}{2(1)} = 10;$$

the 10 arrangements are listed in the table. For each arrangement I determine the observed values of: the number of runs; the length of the longest run of successes; and the length of the longest run of failures. These test statistics are denoted by R, V and W, respectively. (I am sorry that V and W are not evocative of the successes and failures, but I feared confusion if I used S and F.) Once we have the results in Table 11.5, it is easy to obtain the sampling distributions of R, V and W. These sampling distributions are shown in Table 11.6.

We have seen that for a small number of trials it is possible to find the exact sampling distribution of any of the test statistics R, V and W. We will not find **exact** distributions for any practical example because there are too many arrangements to consider. It is quite easy to perform a sim-

ulation experiment for any of these test statistics; indeed, you might have noticed how similar the current situation is to selecting assignments for a CRD at random. (If you don't see the connection, no worries.)

There is a fancy math approximation for the sampling distribution of R and I will discuss it briefly in the following subsection.

11.4.1 The Runs Test

The null hypothesis is that the trials are Bernoulli trials. The test statistic is R. Pause for a moment. What is missing? Correct; I have not specified the alternative hypothesis. For the math level I want to maintain in this course, a careful presentation of the alternative is not possible. Instead, I will proceed by examples.

But first, I want to give you the formulas for the mean and standard deviation of the sampling distribution of R.

Result 11.2 (The mean and standard deviation of the sampling distribution of R.) *Conditional on the number of successes, x, and number of failures, $n - x$, in a sequence of n Bernoulli trials, the mean and standard deviation of the number of runs, R, are given by the equations below. First compute*

$$c = 2x(n - x); \text{ then} \tag{11.4}$$

$$\mu = 1 + \frac{c}{n}; \text{ and} \tag{11.5}$$

$$\sigma = \sqrt{\frac{c(c - n)}{n^2(n - 1)}}. \tag{11.6}$$

Recall my artificial small example—with $n = 5$, $x = 3$ and $n - x = 2$. First, I calculate $c = 2(3)(2) = 12$. The mean and standard deviation are:

$$\mu = 1 + 12/5 = 3.4 \text{ and } \sigma = \sqrt{\frac{12(7)}{5^2(4)}} = \sqrt{0.84} = 0.916.$$

(If you look at the 10 possible arrangements in Table 11.5, and sum the corresponding 10 values of r, you will find that the sum is 34 and the mean is $34/10 = 3.4$, in agreement with the above use of Equation 11.5.)

For my Super Bowl data, $n = 47$, $x = 25$ and $n - x = 22$. Thus, $c = 2(25)(22) = 1100$. The mean and standard deviation are:

$$\mu = 1 + 1100/47 = 24.404 \text{ and } \sigma = \sqrt{\frac{1100(1053)}{47^2(46)}} = \sqrt{11.399} = 3.376.$$

Note that the observed number of runs, 18, is smaller than the mean number under the null hypothesis. This will be relevant very soon.

I need to talk about the alternative, but first I need to present an artificial example. Imagine that I have $n = 50$ dichotomous trials with $x = n - x = 25$. From Result 11.2, $c = 2(25)(25) = 1250$. The mean and standard deviation are:

$$\mu = 1 + 1250/50 = 26 \text{ and } \sigma = \sqrt{\frac{1250(1200)}{50^2(49)}} = \sqrt{12.245} = 3.499.$$

Based on my intuition, there are two arrangements that clearly provide very strong evidence against Bernoulli trials. (I know. We are not supposed to talk about evidence against the null; please bear with me.) The first is a perfect alternating arrangement:

10101 01010 10101 01010 10101 01010 10101 01010 10101 01010

The second is 25 successes followed by 25 failures:

11111 11111 11111 11111 11111 00000 00000 00000 00000 00000

For the first of these arrangements, $R = 50$ and $V = W = 1$. For the second arrangement, $R = 2$ and $V = W = 25$. It is easy to see that the first arrangement gives the largest possible value for R—after all, there cannot be more runs than trials! Similarly, the first arrangement gives the smallest possible value for both V and W. Also, the second arrangement gives the smallest possible value of R and the largest possible value for both V and W. If we consider all possible arrangements it makes sense that large [small] values of R tend to be matched with small [large] values of both V and W. The important consequence of this tendency is that when I do talk about alternatives, the $<$ alternative for R will correspond to the $>$ alternative for V or W.

Note that everything I say about the first arrangement is also true for the arrangement:

01010 10101 01010 10101 01010 10101 01010 10101 01010 10101

Similarly, everything I say about the second arrangement is also true for the arrangement:

00000 00000 00000 00000 00000 11111 11111 11111 11111 11111

Let's suppose that I have convinced you that both of these arrangements—or, if you prefer all four—provide convincing evidence that we do **not** have Bernoulli trials. I now look at the question: Which assumption of Bernoulli trials is being violated? Here are two possible interpretations of the data in the second arrangement which, recall, consists of 25 successes followed by 25 failures:

1. There is almost—but not quite—perfect positive dependence. A success [failure] is almost always followed by a success [failure]. In fact, the only wrong prediction in the *predict the response to remain the same* paradigm occurs at trials 25 to 26.

2. The value of p is 1 for the first 25 trials and is 0 for last 25 trials.

Thus, we don't have Bernoulli trials, but is it because of dependence or because p changes? Note that I am talking about *what we know from the data*; it's possible that your knowledge of the science behind the study will lead you to discard one of my two explanations.

I will **not** give a careful statement of the alternative for the test based on R. Instead I will say that by $<$ I mean that the true model tends to give fewer runs than the (null) Bernoulli trials model. By $>$ I mean that the true model tends to give more runs than the (null) Bernoulli trials model. As usual, by $\neq$ I mean that the true model tends to give either fewer or more runs than the the (null) Bernoulli trials model. Below are the rules for finding the P-value.

Result 11.3 (The P-value for the runs test.) *In each of the equations below, r denotes the observed value of the test statistic R.*

1. *For the alternative $>$, the P-value equals*

$$P(R \geq r) \tag{11.7}$$

2. *For the alternative $<$, the P-value equals*

$$P(R \leq r) \tag{11.8}$$

3. *For the alternative $\neq$, the P-value is the smallest of the following three numbers: 1; twice the P-value for $>$; and twice the P-value for $<$.*

In the interest of intellectual honesty, I mention that there is some minor controversy over my rule for the P-value for $\neq$. But—as best I can tell—only a small proportion of professional statisticians care about it. In general, for many tests of hypotheses, the math arguments for the alternative $\neq$ can be messy. Also, in the examples in this course, we will use a Normal curve approximation, with the continuity correction of 0.5, to obtain approximate P-values for the runs test. I have not been able to locate a good general guideline for using the Normal curve approximation. Thus, I will (somewhat, ok, quite) arbitrarily state that you should use the Normal curve approximation only if the null standard deviation of R is 3 or larger. Thus, we would not use the Normal curve approximation if $n = 5$ and $x = 3$, because its standard deviation is 0.916. We can, however, use the Normal curve approximation for the Super Bowl data because its standard deviation, 3.376, is larger than 3.

I will now illustrate the use of Result 11.3 with the Super Bowl data. Recall that $r = 18$, $\mu = 24.404$ and $\sigma = 3.376$. Because R takes on integer values only, I will use the continuity correction. I go to the Normal curve website:

`http://davidmlane.com/hyperstat/z_table.html`,

enter 24.404 for the *Mean* and 3.376 for the *Sd*.

1. For $>$, I want $P(R \geq 18)$. I find that the area *above* 17.5 is 0.9796.

2. For $<$, I want $P(R \leq 18)$. I find that the area *below* 18.5 is 0.0402.

3. For $\neq$, the approximate P-value is $2(0.0402) = 0.0804$.

Table 11.7: The partial approximate sampling distribution of V, the longest run of successes, in 47 Bernoulli trials, conditional on a total of 25 successes.

v:	9	10	11	12	13	14
Rel. Freq. $(V \geq v)$	0.0355	0.0134	0.0049	0.0017	0.0004	0.0001

Let me note that I have been unable to find a website that allows us to enter our data for a runs test. As always, please let me know if you find one. Minitab will perform a runs test. It uses the Normal curve approximation, but it has two curious features:

1. Minitab does not use the continuity correction. Thus, its P-value for $\neq$ is 0.0578, substantially smaller than the answer—0.0804—obtained with the continuity correction.

2. Minitab gives the P-value only for the alternative $\neq$.

Regarding this second item; one could, of course, halve Minitab's answer—0.0578—to obtain 0.0289 as the approximate P-value for the alternative supported by the data, in this case $<$. I suspect that Minitab's creators are expressing the belief—which has merit—that when we check assumptions we should not be overly restrictive in our choice of alternative.

11.4.2 The Test Statistics V and W

The presentation in this section will be very brief. The null hypothesis is that the trials are Bernoulli trials. I will consider two possible test statistics: V [W] the length of the longest run of successes [failures]. The only alternative I will explore is $>$, that the true model tends to give larger values of V [W] than the (null) Bernoulli trials model. The P-value for the alternative $>$ and the test statistic V is

$$P(V \geq v), \tag{11.9}$$

where v is the observed value of V. The P-value for the alternative $>$ and the test statistic W is

$$P(W \geq w), \tag{11.10}$$

where w is the observed value of W.

I can obtain the exact sampling distributions of V and W only for very small values of n. I am unaware of the existence of any accurate fancy math approximation to either of these sampling distributions. Therefore, we will obtain approximations to these sampling distributions by using a computer simulation experiment.

I performed a simulation experiment with 10,000 reps on the Super Bowl data; my approximate distribution for V is in Table 11.7. Recall that the observed value of V for the Super Bowl data is $v = 13$. Thus, the approximate P-value for the alternative $>$ is 0.0004, a very small number. I am quite convinced that the Super Bowl data did not come from Bernoulli trials.

The observed value of W is 5 for the Super Bowl data; the AFC's longest run of victories occurred in years 7–11. I anticipated that the P-value for W would not be small; thus, I performed

Table 11.8: Katie's data on day 7. She obtained a total of 68 successes and 32 failures. The observed values of the various test statistics are $r = 42$, $v = 19$ and $w = 4$.

SS F SS F S FFF SSSSS FF SSSSSSSS F S F SS F SS F SS FF SSSSS FF SS FF SS F SSS F S FFFF SSSSSSSSSSSSSSSSSS FF SSS F SS FF S F SS F SS F S F

a simulation experiment with only 1,000 reps. The relative frequency of occurrence of $W \geq 5$ was 0.422; thus, this is my approximate P-value.

I end this section with a small piece of a large study. Katie Voigt was a starting shooting guard on the women's basketball team at the University of Wisconsin–Madison. A few years later she was kind enough to collect a large amount of data and allow me to analyze it. For each of 20 days, after warming-up, Katie would attempt 100 shots from a favorite spot behind the three-point line. In the following example I will tell you about Katie's data on day 7. The last Practice Problem will look at her data from day 3 and the last Homework Problem will look at her data from day 8.

Example 11.2 (Day 7 of Katie's Study of Shooting.) *On day 7, Katie made 68 of her 100 attempts. The data are presented in Table 11.8. I do* **not** *want you to verify any of the numbers in the caption of the table. Indeed, the table's construction is* **not** *conducive to verifying calculations. Rather, I want you to* look at *the data and form an impression of it.*

I will now find the approximate P-values for Katie's day 7 data, for each of our three test statistics. I will use the alternative $<$ for R for illustration because a small value of R tends to go with large values of V and W, as I discussed earlier.

First, I need to use Result 11.2 to find the (null) mean and standard deviation of R. First, $c = 2x(n - x) = 2(68)(32) = 4{,}352$. The mean and the standard deviation are:

$$\mu = 1 + c/n = 1 + (4352/100) = 44.52 \text{ and } \sigma = \sqrt{\frac{c(c-n)}{n^2(n-1)}} = \sqrt{\frac{4352(4252)}{100^2(99)}} = 4.323.$$

For the alternative $<$ the P-value equals $P(R \leq r) = P(R \leq 42)$. I go to the website:

`http://davidmlane.com/hyperstat/z_table.html`

I enter Mean $= 44.52$ and Sd $= 4.323$. Next to the option *Below* I enter 42.5—remember the continuity correction. The site tells me that the approximate P-value is 0.3202. Thus, the runs test detects only weak evidence in support of the alternative.

I performed two simulation experiments; with 10,000 reps for V and 1,000 reps for W. I obtained the following relative frequencies:

$$\text{Rel. Freq. } (V \geq 19) = 0.0074 \text{ and Rel. Freq. } (W \geq 4) = 0.525$$

Thus, V is very sensitive to the evidence in the data, while W is not.

11.5 Summary

If the response is a dichotomy, then i.i.d. trials are called Bernoulli trials. One possible response is denoted a success and the other is a failure. The probability of a success on any particular trial is denoted by p and the probability of a failure on any particular trial is denoted by $q = 1 - p$. We restrict attention to p's that satisfy $0 < p < 1$; i.e., we are not interested in the cases in which $p = 0$ or $p = 1$.

Let X denote the total number of successes in n Bernoulli trials. The sampling (probability) distribution of X is given by Equation 11.1:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

The above is really a *family* of equations, because we get a different equation by changing n and/or p. It is referred to as a binomial distribution with parameters n and p and a particular member of this family is denoted by the Bin(n, p) distribution.

Except for very small values of n it is tedious to calculate a binomial probability by hand. If $n \leq$ 1,000 the website:

`http://stattrek.com/Tables/Binomial.aspx`

provides exact probabilities. If $n >$ 1,000 and both np and nq are 25 or larger, then we can obtain good approximations to the binomial by using the Normal curve with

$$\mu = np \text{ and } \sigma = \sqrt{npq}.$$

As you learned in Chapter 7, the website:

`http://davidmlane.com/hyperstat/z_table.html`

can be used to obtain areas under a Normal curve—remember to use the continuity correction.

Instead of obtaining the Normal approximation *by hand*—by which I mean calculating the values of μ and σ by hand—if $n >$ 1,000 and both np and nq are 25 or larger, then the website

`http://stattrek.com/Tables/Binomial.aspx`

will do the work for you. **But be careful!** This website **does not always check the values of** np and nq and if either of these is small, then the approximate answer from the website can be very bad.

Stating the obvious, we cannot compute a binomial probability—exact or approximate—without knowledge of the numerical value of p. In Section 11.3, I argue that if I have data from a huge number of Bernoulli trials—huge in this context means, to me, approximately 10,000 or more, provided p is not too close to either 0 or 1—then I am willing to replace p by the proportion of successes in the data, $\hat{p}$. I will return to this topic in the next chapter.

It is more than a bit unsatisfactory to be forced to simply assume or fail to assume that one has Bernoulli trials. One might, for example, want to perform a test of hypotheses with the null hypothesis that the assumptions of Bernoulli trials are correct. In Section 11.4, I provide an approach

to this test of hypotheses with three possible test statistics: the number or runs, R; the length of the longest run of successes, V; and the length of the longest run of failures, W.

Apart from the issue of a test of hypotheses, sports fans have long been interested in long runs of successes; for example, Joe DiMaggio's record of hitting safely in 56 consecutive major league baseball games and Micheal Williams's record of 97 consecutive free throws made during NBA games. Knowing the sampling distribution of V—in addition to its use in finding a P-value—can be used to gauge how *remarkable* a record might be.

11.6 Practice Problems

1. Suppose that we have Bernoulli trials with $p = 0.8$.

   (a) Calculate the probability that the first four trials yield: two successes, then a failure and then a success.

   (b) Calculate, by hand, the probability that the first six trials yield a total of five or more successes.

   (c) Use the website:

   `http://stattrek.com/Tables/Binomial.aspx`

   to verify your answer to (b).

2. Let X denote the total number of successes in $n = 900$ Bernoulli trials for which $p = 0.7$.

   (a) Use the website

   `http://stattrek.com/Tables/Binomial.aspx`

   to find the exact value of: $P(X \geq 600)$; and $P(X \leq 645)$.

   (b) Calculate the mean and standard deviation of X.

   (c) Find the Normal curve approximation of: $P(X \geq 600)$; and $P(X \leq 645)$; remember to use the continuity correction.

   (d) Compare your answers in (a) and (c)

3. I programmed Minitab to generate 40 Bernoulli trials with $p = 0.6$. Below is the result I was given. (The spaces have no meaning; I typed them to make the counting in part (a) easier.)

   1 00 1111 00 1 00 1 0 1 0 1111111 00 1 00 1 00 111111 0 1 0

   (a) Determine the observed values of X, $n - X$, R, V and W.

   (b) Conditional on the observed value of X calculate the mean and standard deviation of R

   (c) Use the Normal approximation to compute the P-value for the test statistic R and the alternative $\neq$.

   (d) I performed a simulation experiment with 1,000 reps and obtained the following frequencies:

   $$\text{Frequency } (V \geq v) = 305 \text{ and Frequency } (W \geq w) = 887,$$

   where the values of v and w were determined in part (a). Comment on these results.

4. I examined the outcomes of the last 100 World Series in baseball, for the years 1912–2012—there was no Series in 1994. The American League team won (a success) 58 Series. The observed values of our three test statistics are $r = 58$, $v = 7$ and $w = 4$.

Table 11.9: The partial approximate sampling distribution of V, the longest run of successes, in 100 Bernoulli trials, conditional on a total of 66 successes.

v:	11	12	13	14
Relative Freq. $(V \geq v)$	0.2491	0.1536	0.0951	0.0587

Table 11.10: The partial approximate sampling distribution of W, the longest run of failures, in 100 Bernoulli trials, conditional on a total of 34 failures.

w:	7	8	9	10
Relative Freq. $(W \geq w)$	0.0219	0.0072	0.0019	0.0008

(a) Calculate the mean and standard deviation of R, conditional on $X = 58$ successes.

(b) Use the Normal curve approximation to find an approximate P-value for the runs test with alternative $>$.

(c) A simulation experiment with 1,000 reps yielded:

$$\text{Frequency } (V \geq 7) = 595 \text{ and Frequency } (W \geq 4) = 875.$$

Comment on the meaning of these results.

5. Recall day 7 of Katie's study of shooting, Example 11.2 and the description immediately before it. In this problem we will analyze Katie's data from day 3.

Here are some important statistics from Katie's 100 trials on day 3:

$$x = 66, r = 31, v = 12 \text{ and } w = 9.$$

(a) Use the Normal curve approximation to obtain the P-value for the runs test and the alternative $<$. Comment.

(b) I performed a simulation study with 10,000 reps. Partial results for the test statistic V are in Table 11.9. Use this information to find the approximate P-value for the test statistic V and the alternative $>$.

(c) I performed a simulation study with 10,000 reps. Partial results for the test statistic W are in Table 11.10. Use this information to find the approximate P-value for the test statistic W and the alternative $>$.

11.7 Solutions to Practice Problems

1. (a) The probability of the sequence SSFS is obtained by using the multiplication rule:

$$P(SSFS) = ppqp = (0.8)^3(0.2) = 0.1024.$$

   (b) The probability of interest is

$$P(X \geq 5) = P(X = 5) + P(X = 6) = \frac{6!}{5!1!}(0.8)^5(0.2) + (0.8)^6 =$$

$$0.3932 + 0.2621 = 0.6553.$$

   (c) I entered the values $p = 0.8$, $n = 6$ and $x = 5$; then I clicked on *Calculate* and obtained the answer:

$$P(X \geq 5) = 0.6554,$$

   which equals my answer, except for (my) round-off error.

2. (a) I entered the values $p = 0.7$, $n = 900$ and $x = 600$; then I clicked on *Calculate* and obtained the answer:

$$P(X \geq 600) = 0.9861.$$

   I changed my x to 645, then I clicked on *Calculate* and obtained the answer:

$$P(X \leq 645) = 0.8705.$$

   (b) The values are:

$$\mu = 900(0.7) = 630 \text{ and } \sigma = \sqrt{900(0.7)(0.3)} = 13.75.$$

   (c) For $P(X \geq 600)$: The area under the N(630,13.75) curve to the right of 599.5 is 0.9867.

   For $P(X \leq 645)$: The area under the N(630,13.75) curve to the left of 645.5 is 0.8702.

   (d) The first approximation is too large by 0.0006; the second is too small by 0.0003. Both approximations are quite good.

3. (a) By counting, I obtain:

$$X = 24, n - X = 16, R = 20, V = 7 \text{ and } W = 2.$$

   (b) First, $c = 2(24)(16) = 768$. Thus,

$$\mu = 1 + 768/40 = 20.2 \text{ and } \sigma = \sqrt{\frac{768(728)}{40^2(39)}} = 2.993.$$

(c) Note that the approximating Normal curve has mean equal to 20.2. For the alternative $<$, we need the area under the Normal curve to the left of 20.5. This area must be larger than 0.5 because $20.5 > 20.2$. For the alternative $>$, we need the area under the Normal curve to the right of 19.5. This area must be larger than 0.5 because $19.5 < 20.2$. The P-value for $\neq$ is the minimum of three numbers: 1 and two numbers that are both larger than 1. Hence, the P-value is 1.

(d) The smallest possible value of W is 1 and we obtained $W = 2$. Thus, it is no surprise that its approximate P-value is huge, 0.887. A run of seven successes seems, to my intuition, to be pretty long, but the relatively large approximate P-value, 0.305, means that the data's support for $>$ is not very strong.

4. (a) First, $c = 2(58)(42) = 4{,}872$. Thus,

$$\mu = 1 + 4872/100 = 49.72 \text{ and } \sigma = \sqrt{\frac{4872(4772)}{100^2(99)}} = 4.846.$$

(b) I need the area under the N(49.72,4.846) curve to the right of 57.5. This area equals 0.0542. The large number of runs is not quite statistically significant, but it indicates that the data set includes a great deal of switching between 0's and 1's.

(c) Because of the large amount of switching back-and-forth noted in part (b), I don't expect the lengths of runs will be very noteworthy. Indeed, my approximate P-value for $>$ for V [W] is 0.595 [0.875]. Thus, neither test statistic provides much evidence in support of the alternative.

5. (a) I need to obtain the values of μ and σ. First, $c = 2(66)(34) = 4{,}488$. Thus,

$$\mu = 1 + 4488/100 = 45.88 \text{ and } \sigma = \sqrt{\frac{4488(4388)}{100^2(99)}} = 4.460.$$

I need the area under the N(45.88,4.460) curve to the left of 31.5. This area is 0.0006, a very small P-value. Katie obtained many fewer runs than expected under the null hypothesis.

(b) The approximate P-value for the alternative $>$ and the test statistic V is 0.1536. There is evidence in support of the alternative, but it is not convincing.

(c) The approximate P-value for the alternative $>$ and the test statistic W is 0.0019. There is very strong evidence in support of the alternative; nine is a very long run of failures in a sequence of 100 trials, of which 66 are successes.

Table 11.11: The partial approximate sampling distribution of V, the longest run of successes, in 100 Bernoulli trials, conditional on a total of 59 successes.

v:	12	13	14	15
Relative Freq. $(V \geq v)$	0.0457	0.0242	0.0134	0.0064

Table 11.12: The partial approximate sampling distribution of W, the longest run of failures, in 100 Bernoulli trials, conditional on a total of 41 failures.

w:	3	4	5	6	7	8
Relative Freq. $(W \geq w)$	0.9982	0.8559	0.4890	0.2062	0.0770	0.0274

11.8 Homework Problems

1. We plan to observe $n = 600$ Bernoulli trials with $p = 0.40$. Let X denote the total number of successes in the 600 trials.

   (a) Calculate the probability that the first four trials yield: a success, two failures and a success, in that order.

   (b) Calculate the probability that the first four trials yield a total of exactly two successes.

   (c) Use the website

   `http://stattrek.com/Tables/Binomial.aspx`

   To obtain:

   $$P(X \leq 255), P(X \leq 230) \text{ and } P(231 \leq X \leq 255).$$

   (d) Use the Normal curve to obtain approximate answers for the three probabilities of interest in part (c).

2. Mimic what I did in Tables 11.5 and 11.6 for the 15 possible arrangements of four 1's and two 0's.

3. Recall day 7 of Katie's study of shooting, Example 11.2 and the description immediately before it. In this problem we will analyze Katie's data from day 8.

   Here are some important statistics from Katie's 100 trials on day 8:

   $$x = 59, r = 42, v = 13 \text{ and } w = 4.$$

   (a) Use the Normal curve approximation to obtain the P-value for the runs test and the alternative $<$. Comment.

(b) I performed a simulation study with 10,000 reps. Partial results for the test statistic V are in Table 11.11. Use this information to find the approximate P-value for the test statistic V and the alternative $>$.

(c) I performed a simulation study with 10,000 reps. Partial results for the test statistic W are in Table 11.12. Use this information to find the approximate P-value for the test statistic W and the alternative $>$.