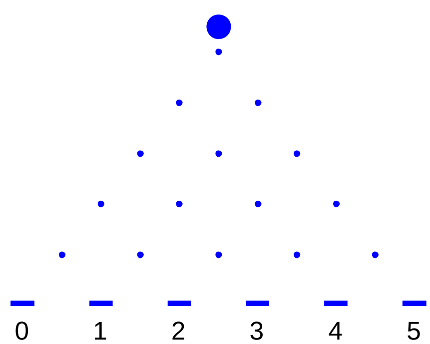
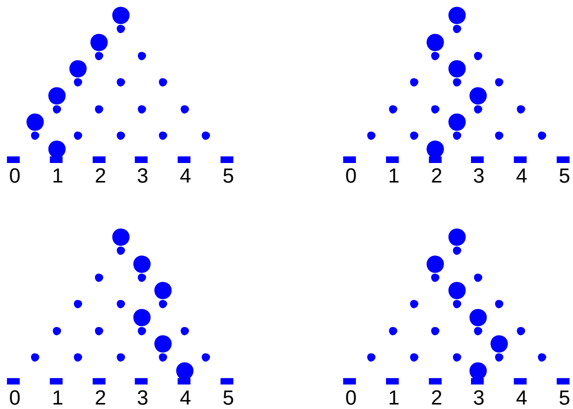
TOPIC P7: COIN-TOSSING DISTRIBUTIONS

SPOTLIGHT: A GALTON BOARD

A Galton board is a physical device for simulating a special type of random experiment that we describe in this chapter. It was named after the famous scientist Sir Francis Galton who lived from 1822 to 1911. Galton is noted for a wide range of achievements in the areas of meteorology, genetics, psychology, and statistics. The Galton board consists of a set of pegs laid out in the configuration shown in the below figure – one peg is in the top row, two pegs are in the second row, three pegs in the third row, and so on. A ball is placed above the top peg. When the ball is dropped and hits a peg, it is equally likely to fall left or right. We are interested in the location of the ball after striking five pegs – as shown in the diagram, the ball can land in locations 0, 1, 2, 3, 4, or 5.



The below figure shows the path of four balls that fall through a Galton board. The chances of falling in the locations follow a special probability distribution that has a strong connection with a simple coin-tossing experiment.



PREVIEW

Consider the following random experiment. You take a quarter and flip it ten times, recording the number of heads you get. There are four special characteristics of this simple coin-tossing experiment.

1. You are doing the same thing (flip the coin) ten times. We will call an individual coin flip a *trial*, and so our experiment consists of ten identical trials.
2. On each trial, there are two possible outcomes, heads or tails.
3. In addition, the probability of flipping heads *on any trial* is ½.
4. The results of different trials are independent. This means that the probability of heads, say, on the fourth flip, does not depend on what happened on the first three flips.

We are interested in the number of heads we get – we will refer to this number as X. In particular, we are interested in the probability of getting five heads, or Prob(X = 5).

In this topic, we will see that this binomial probability model applies to many different random phenomena in the real world. We discuss probability computations for the binomial and closely related negative binomial models and illustrate the usefulness of these models in representing the variation in real-life experiments.

In this topic, the learning objectives are to:

- Understand and recognize binomial and negative binomial experiments.

- Understand how to compute and apply binomial and negative binomial probabilities to real-life problems.
- Understand how real coin tossing can be differentiated from fake coin tossing where students are imagining flips of heads and tails.

NCTM Standards

✓In Grades 9-12, all students should understand the concepts of sample space and probability distribution and construct sample spaces and distributions in simple cases

✓In Grades 9-12, all students should compute and interpret the expected value of random variables in simple cases.

PROBABILITIES OF A COIN-TOSSING EXPERIMENT

Let's return to our experiment where a quarter is flipped ten times, recording X, the number of heads. We are interested in the probability of flipping exactly five heads, that is, Prob(X = 5). To compute this probability, we first have to think of possible outcomes in this experiment. Suppose we record if each flip is heads (H) or tails (T). Then one possible outcome when we make ten flips is

Trial	1	2	3	4	5	6	7	8	9	10
Result	H	H	T	T	H	T	T	H	H	T

Another possible outcome is TTHHTHTHHH. The sample space consists of all possible ordered listings of ten letters, where each letter is either an H or a T.

Next, consider computing the probability of a single outcome of ten flips such as the HHTTHHTHHT sequence shown above. We can write the probability of this outcome as

Prob("H on toss 1" AND "H on toss 2" AND "T on toss 3" AND … AND "T on toss 10")

Using the fact that outcomes on different trials are independent, this probability can be written as the product

Prob(H on toss 1) × Prob(H on toss 2) × Prob(T on toss 3) × … × Prob(T on toss 10).

Since the probability of heads (or tails) on a given trial is ½, we have

$$\text{Prob(HHTTHHTTHT)} = \left(\frac{1}{2}\right)\times\left(\frac{1}{2}\right)\times\cdots\times\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{10}.$$

Actually, the probability of any outcome (sequence of ten letters with H's or T's) in this experiment is equal to $\left(\frac{1}{2}\right)^{10}$.

Let's return to our original question – what is the probability that we get exactly five heads? If we think of the individual outcomes of the ten trials, then we'll see that there are many ways to get five heads. For example, we could observe

HHHHHTTTTT or HHHHTTTTTH or HHHTTTTTHH

In each of the three outcomes, note that the number of heads is five.

How many outcomes (like the ones shown above) will result in exactly five heads? As before, we label the outcomes of the individual flips by the trial number:

Trial	1	2	3	4	5	6	7	8	9	10
Outcome	_	_	_	_	_	_	_	_	_	_

If we observe five heads, then we wish to place five H's in the ten slots above. In the outcome HHHHHTTTTT, the heads occur in trials 1, 2, 3, 4, 5, and in the outcome HHHTTTTTHH, the heads occur in trials 1, 2, 3, 9, and 10. If we observe exactly 5 heads, then we must choose five numbers from the possible trial numbers 1, 2, …, 10 to place the five H's. There are ${}_{10}C_5$ ways of choosing these trial numbers. (The order in which we choose the trial numbers is not important.) Since there are ${}_{10}C_5$ ways of getting exactly five heads, and each outcome has probability $\left(\frac{1}{2}\right)^{10}$, we see that

$$\text{Prob}(X = 5) = {}_{10}C_5\left(\frac{1}{2}\right)^{10} = 0.246.$$

From a basic property of probabilities, we see that the Prob(five heads *are not* tossed) = 1- 0.246 = 0.754. It is interesting to note that although we *expect* to get five heads when we flip a coin ten times, it is actually much more likely *not* to flip five heads than to flip five heads.

PRACTICE: COIN-TOSSING EXPERIMENTS

Suppose you flip a coin five times.

1. Write down all possible sequences of five coin flips.

2. Next to each outcome of five flips, write down the value of X = the number of heads observed.

3. Using the work in 1 and 2, find the probability distribution for X. Put your probabilities in the below table.

X	0	1	2	3	4	5
P(X)						

4. By using counting arguments, how many sequences of five flips will contain exactly 2 heads? Check that this number agrees with your computation in part 3.

5. If you flip a coin 20 times, how many outcomes will result in exactly 10 heads? What is the probability of 10 heads?

BINOMIAL EXPERIMENTS

Although the coin tossing experiment described above seems pretty artificial, many random experiments share the same basic properties as coin tossing. Consider the following *binomial experiment*:

1. We are repeating the same basic task or trial many times – let the number of trials be denoted by n.

2. On each trial, there are two possible outcomes, which we will call "success" or "failure". (We could call the two outcomes "black" and "white", or "0" or "1", but they are usually called success and failure.)

3. The probability of a success, denoted by p, is the same for each trial.

4. The results of outcomes from different trials are independent.

Here are some examples of binomial experiments.

A sample survey. Suppose the Gallup organization is interested in estimating the proportion of adults in the United States who use the popular auction web site EBay. They take a random sample of 100 adults and 45 say that they use EBay. In this story, we see that

1. The results of this survey can be considered to be a sequence of 100 trials where one trial is asking a particular adult if he or she uses EBay.
2. There are two possible responses to the survey question – either the adult says "yes" (he or she uses EBay) or "no" (he or she doesn't use EBay).
3. Suppose the proportion of all adults that use EBay is p. Then the probability that the adult says "yes" will be p.
4. If the sampling is done randomly, then the chance that one person says "yes" will not depend on the answers of the people who were previously asked. This means that the responses of different adults to the question can be regarded as independent events.

A baseball hitter's performance during a game. Suppose you are going to a baseball game and your favorite player comes to bat five times during the game. This particular player is a pretty good hitter and his batting average is about .300. You are interested in the number of hits he will get in the game. This can also be considered a binomial experiment:

1. The player will come to bat five times – these five at-bats can be considered the five trials of the experiment ($n = 5$).
2. At each at-bat, there are two outcomes of interest – either the player gets a hit or he doesn't get a hit.
3. Since the player's batting average is .300, the probability that he will get a hit in a single at-bat is $p = .300$.
4. It is reasonable to assume that the results of the different at-bats are independent. That means that the chance that the player will get a hit in his fifth at-bat will be unrelated to his performance in the first four at-bats. (This is a debatable assumption, especially if you believe that a player can have a hot-hand.)

Sampling without replacement. Suppose a committee of four will be chosen at random from a group of five women and five men. You are interested in the number of women that will be in the committee. Is this a binomial experiment?

1. If we think of selecting this committee one person at a time, then we can think this experiment as four trials (corresponding to selecting the four people).
2. On each trial, there are two possible outcomes – either we select a woman or a man.

At this point, things are looking good – this may be a binomial experiment. But …

3. Is the probability of choosing a woman the same for each trial? For the first pick, the chance of picking a woman is 5/10. But once this first person has been chosen, the probability of choosing a woman is not 5/10 – it will be either 4/9 or 5/9 depending on the outcome of the first trial. So the probability of a "success" is not the same for all trials, so this violates the third property of a binomial experiment.
4. Likewise, in this experiment, the outcomes of the trials are not independent. The probability of choosing a woman on the fourth trial is dependent on who was selected in the first three trials, so again the binomial assumption is violated.

PRACTICE: BINOMIAL EXPERIMENTS

Are each of the following binomial experiments? If so, indicate what is a “success” and give values of n and p.

1. From weather records, you know that 60% of the days in your town will be sunny. You record the number of sunny days for ten randomly selected days in the year.

2. You keep flipping a coin until you observe two heads.

3. You have a box of 10 mittens of which six are red. You select four mittens from the box and count the number of red.

4. A company knows from experience that 10% of the products they sell will need some repair during the 90-day warranty period. They survey 12 consumers who have purchased this product and record the number who need repair within the warranty period.

5. A car dealer passes out of consumer satisfaction survey. Twenty people are surveyed of which 10 are satisfied, 5 are not satisfied, and 5 have no opinion.

ACTIVITY: COIN FLIPPING: IS IT REAL OR FAKE?

1. (Fake coin tossing.) Pretend to flip a coin 200 times – put your results (H or T for each toss) in the boxes below.

Pretend Coin Flips																			

X	Y	Z

2. (Real coin tossing) Now flip a quarter 200 times – put your results (H or T) in the boxes below.

Real Coin Flips																			

X	Y	Z

3. For a given sequence of coin tosses, we define a *run* as a consecutive sequence of heads or tails. So for example, if we observe the sequence

TTHHTHHHTTTTTHHHTTT

we observe a run of two tails, a run of two heads, a run of one tail, a run of three heads, and so on. Here the length of the longest run is 5, since the longest run is TTTTT and the length of this run is 5. We define the *number of switches* as the number of changes from H to T, or from T to H. I count a total number of six switches in the above sequence.

For each row of 20 tosses in the above two tables, compute X = the number of heads, Y = the length of the longest run of heads or tails, Z = the number of switches.

4. Collect the values of X, Y, and Z from all students in the class for the fake coin flips and the real coin flips. Place the data in the boxes below

Fake coins – values of X	Fake coins – values of Y	Fake coins – values of Z

(number of heads)	(longest run)	(number of switches)
Real coins – values of X (number of heads)	Real coins – values of Y (longest run)	Real coins – values of Z (number of switches)

5. Compare the number of heads for the fake coin flips and the real coin flips by constructing parallel dotplots. By looking at the two histograms and calculating suitable summary statistics, explain how the numbers of heads for the real coins look different from the number of heads for the fake coins.

6. Do the same comparison using the longest run variable. Which dataset tends to have "long" runs – the fake coins or the real coins?

7. Repeat the comparison using the number of switches variable. Do you notice any differences between the histogram for the number of switches for the real coins and the number of switches for the fake coins?

TECHNOLOGY ACTIVITY: SIMULATED COIN FLIPPING

This activity simulates coin flipping on Fathom. We'll consider an imaginary coin where the chance of heads on a single flip is p, which could be different from ½. We'll toss this coin 20 times – each time we will keep track of the number of heads and the length of the longest run of heads.

1. Open up the Fathom document "coin_tossing.ftm". You'll see a

- slider where you can change p, the probability of getting heads on a single flip

- the results of 20 random coin flips

2. Select the Collection and simulate 20 flips by typing Apple (or Control) – Y. For each simulation, record the number of heads, and the length of the longest run.

Simulation	Number of heads	Length of longest run
1		
2		
3		
4		
5		

3. This Fathom program has been set up to repeat this experiment (of flipping 20 coins) 1000 times – for each experiment, we record the number of heads and the length of the longest run.

To do this, select the Measures from Collection and type Apple-Y. This collection contains the number of heads and length of longest run for these 1000 experiments.

4. First look at the Number of Heads for the 1000 experiments.

(a) Construct a histogram of the number of heads.

(b) Construct a count or frequency table of the number of heads.

(c) What is the most likely number of heads you will get?

(d) What is the probability you will flip exactly 10 heads?

(e) What is the probability you will flip 15 or more heads?

5. Now look at the Length of the Longest Run for the 1000 experiments.

(a) Construct a histogram and count table.

(b) What is the most likely length of the longest run?

(c) What is the probability the longest run is 6 or more?

6. Above, we assumed that $p = .5$ (we were flipping a fair coin). Now using the slider, change p to .3. Redo the simulation of 1000 experiments.

Answer the same questions as in 4 and 5.

ACTIVITY: IS A PROFESSIONAL ATHLETE STREAKY?

Sports fans are often interested in streaky performances of athletes during a game. For example, the following table gives the results of 40 shots taken by Kobe Bryant (basketball player who plays for the Lakers) during a 2002 professional basketball game.

ALL OF KOBE BRYANT SHOTS for Lakers 96, Warriors 89

11/15/2002 STAPLES Center, Los Angeles, CA

1st Period	3rd Period
(8:06) Bryant Jump Shot: MADE	(11:15) Bryant Turnaround Jump: MADE
(7:33) Bryant Jump Shot: MADE	(10:31) Bryant Jump Shot: MISSED
(6:46) Bryant Jump Shot: MADE	(8:31) Bryant Layup Shot: MISSED
(5:50) Bryant Jump Shot: MADE	(7:04) Bryant Jump Shot: MISSED
(5:16) Bryant Jump Shot: MISSED	(6:33) Bryant Jump Shot: MISSED
(4:55) Bryant Dunk Shot: MADE	(4:23) Bryant Driving Finger Roll: MADE
(4:21) Bryant Turnaround Jump: MISSED	(2:06) Bryant Dunk Shot: MADE
(3:15) Bryant Jump Shot: MISSED	(0:57) Bryant Jump Shot: MISSED
(3:05) Bryant Jump Shot: MADE	
(0:39) Bryant Jump Shot: MISSED	4th Period
(0:01) Bryant Jump Shot: MISSED	(8:19) Bryant Turnaround Jump: MADE
	(6:54) Bryant Slam Dunk Shot: MADE
2nd Period	(6:24) Bryant Jump Shot: MADE
(10:24) Bryant Jump Shot: MADE	(5:48) Bryant Jump Shot: MISSED
(9:41) Bryant Jump Shot: MISSED	(5:10) Bryant Jump Shot: MISSED
(5:20) Bryant Jump Shot: MADE	(4:12) Bryant Jump Shot: MISSED
(4:38) Bryant Jump Shot: MADE	(1:57) Bryant Jump Shot: MADE
(4:03) Bryant Fade Away: MISSED	(0:47) Bryant Fade Away: MISSED
(1:09) Bryant Driving Layup: MADE	(0:18) Bryant Jump Shot: MISSED
(0:34) Bryant Jump Shot: MISSED	
(0:02) Bryant Turnaround Jump: MISSED	1st Overtime (4)
	(4:26) Bryant Turnaround Jump: MISSED
	(2:08) Bryant Jump Shot: MISSED
	(1:31) Bryant Turnaround Jump: MADE
	(0:34) Bryant Turnaround Jump: MISSED

1. For Kobe's data, compute the length of the longest run of makes or misses.
2. Do you think this value is unusually small or large? Why?
3. One way of deciding if Kobe's longest run of makes or misses is unusual is to compare this value with the longest run of heads or tails in 40 flips of a fair coin. (We use a fair coin since Kobe's probability of making a particular shot is approximately .5.) Below we have simulated 20 sequences of forty coin flips. For each sequence, compute Y = the longest run of heads or tails and record this value on the right.

Y = longest run

```
seq 1   HTHTHHHTTTTTHHTHTHTTHHHTTTHTHTTHHTTHTTTH
seq 2   TTHTHHTTTTTTTHTHHTHTTHTHHTHTTHTTTTHHHTHT
seq 3   TTHHTHHTHTHHTHTHTTTHTTTHHHHTHHHHHTHTHTTH
seq 4   THHTTHHHHTTTTHHTTTTHHTTTTHHTHTHHHHHTHTHH
seq 5   HHTTTTHHHHHHTHTHHHTHTHHHTTHHTTHHHHTTHHTT
seq 6   THTHHHTTHHTTTTTTHHHHHHTHHTHTHTTHTHTHHHHH
seq 7   HHHHHTHHHHTTTHHTHTHTHHTHTTHHHHTTHTHTHHTT
seq 8   HTHHTTTHHTTTHTTTHTHHHHTTHTHTHTTHHTTHTTHH
seq 9   TTHHHHHHHTTHHTTTTTHHTHHTHHHTTHHTHTTHHHHT
seq 10  HTHTTTTHTHTTHTHTHTHTTHHHTHTTTHHHHHHTTTTT
seq 11  HTTTTTHTHHTTTHTHHHHHHHHHTHHTTTHTHTTTHHTH
seq 12  TTHHHTHTHHHTHHHTHTHHHTHHTTTHHHHHHTTTTTHT
seq 13  THHTTHTHTHHTHHHTHHHTTHTTHTTTHTTHTTHHHTHH
seq 14  TTTTHHHHTTHTTTHTHTHTHTTTHTTHTTHTTHTTHTTH
seq 15  THHHTHTTTTHTTTTTTHHHHHHTHTTHTHHHHHTTTHTH
seq 16  THHHTTTHHTHHHTHTHTTHHHHHHHTTHTHTHHHHHTTH
seq 17  HHTHHTTTTHHHHTTHTTTTHHHHHTTTHHTHHTHHHHHT
seq 18  THHTTTTTTHHHTTHTTHHHTTTHHHTHTTTHTTHTHHTH
seq 19  THHTTTHTTHTTHHTHHTTTTHHHHTTHTTTTHHHTTHHT
seq 20  HTTTHHTHTTTTTHTHHHTHTHTTTHTHHTTTHTTHTTTH
```

4. Construct a dotplot of the 20 values of the longest run Y. Indicate the length of Kobe's longest run by a vertical line placed on the dotplot. Based on this graph, would

you say that Kobe's longest run is unusual relative to the distribution of the longest run for 40 coin flips? Explain.

BINOMIAL COMPUTATIONS

A binomial experiment is defined by two numbers

n = the number of trials, and p = probability of a "success" on a single trial.

If we recognize an experiment as being binomial, then all we need to know is n and p to determine probabilities for the number of successes X.

Using the same argument as we made in the coin-tossing example, one can show that the probability of k successes in a binomial experiment is given by

$$P(X = k) = {}_nC_k\, p^k (1-p)^{n-k}, \quad k = 0, \ldots, n.$$

Let's illustrate using this formula for a few examples.

Baseball example (revisited). Remember our baseball player with a true batting average of .300 is coming to bat five times during a game. What is the probability that he gets exactly two hits?

We showed earlier that this was a binomial experiment. Since the player has five opportunities, the number of trials is n = 5. If we regard a success as getting a hit, the probability of success on a single trial is p = 0.3. The random variable X is the number of hits of the player during this game.

Using the formula, the probability of exactly two hits is

$$P(X = 2) = {}_5C_2 (0.3)^2 (1-0.3)^{5-2} = 0.3087$$

What is the probability that the player gets at least one hit? To do this problem, we first construct the collection of binomial probabilities for n = 5 trials and probability of success p = 0.3. The table below shows all possible values of X (0, 1, 2, 3, 4, 5) and the associated probability that can be found using the binomial formula.

X	P(X)
0	0.168
1	0.360
2	0.309
3	0.132
4	0.029
5	0.002

We are interested in the probability that the player gets at least one hit or Prob(X >= 1). "At least one hit" means that X can be 1, 2, 3, 4, or 5. To find this we simply sum the probabilities of X between 1 and 5:

$$\text{Prob}(X>=1) = P(X = 1, 2, 3, 4, 5) = 0.360 + 0.309 + 0.132 + 0.029 + 0.002 = 0.832.$$

There is a simpler way of doing this computation using the complement property of probabilities. We note that if the player doesn't get at least one hit, then he was hitless in the game (that is, X = 0). Using the complement property

$$\text{Prob}(X>=1) = 1 - \text{Prob}(X = 0) = 1 - 0.168 = 0.832.$$

PRACTICE: BINOMIAL COMPUTATIONS

Suppose a student takes a six-question true/false test. He or she guesses at each question.

1. Explain why this is a binomial experiment and give values of n and p.
2. Write down the formula expression for the probability that the student gets exactly three correct.
3. The probability distribution for X = the number correct is shown below. Using this table, find the probability the student gets at least 3 correct.

X	P(X)
0	0.016
1	0.094
2	0.234
3	0.312
4	0.234
5	0.094
6	0.016

4. Find the probability the student gets fewer than 2 correct.
5. Suppose the student gets all of the questions correct. Is it reasonable to assume that the student is really guessing at all the questions?

MEAN AND STANDARD DEVIATION OF A BINOMIAL

There are simple formula for the mean and variance for a binomial random variable. First let X_1 denote the result of the first binomial trial where

$$X_1 = \begin{cases} 1, & if \ \ we\ observe\ a\ success \\ 0, & if \ \ we\ observe\ a\ \ failure \end{cases}$$

In the exercises, you will be asked to show that the mean and variance of X_1 are given by

$$E(X_1) = p, \quad Var(X_1) = p(1-p).$$

If $X_1, \ldots, X_n$ represent the results of the n binomial trials, then the binomial random variable X can be written as

$$X = X_1 + \cdots + X_n$$

Using this representation, the mean and variance of X are given by

$$E(X) = E(X_1) + \cdots + E(X_n), \ \ Var(X) = Var(X_1) + \cdots + Var(X_n).$$

(the result about the variance is a consequence of the fact that the results of different trials of a binomial experiment are independent). Using this result and the previous result on the mean and variance of an individual trial outcome, we obtain

$$E(X) = p + \cdots + p = np$$

$$Var(X) = p(1-p) + \cdots + p(1-p) = np(1-p).$$

To illustrate these formulas, recall the first example where X denoted the number of heads when a fair coin is flipped 10 times. Here the number of trials and probability of success are given by n = 10 and p = .5. The expected number of heads would be

$$E(X) = 10\ (.5) = 5$$

and the variance of the number of heads would be

$$Var(X) = 10\ (.5)\ (1- .5) = 2.5.$$

PRACTICE: MEAN AND STANDARD DEVIATION OF A BINOMIAL

1. Suppose you have a box with 15 black beads and 5 white beads. You select six beads from the box with replacement and count the number of white beads. Explain why this is a binomial experiment and give values of n and p.
2. If X is the number of white beads you select, find the mean and standard deviation of X.
3. You should notice that E(X) is not a whole number. Is it reasonable to say that E(X) is the most likely value of X? If not, give an alternative interpretation to E(X).
4. Suppose you flip a fair coin 10 times and count the number of heads X. In a second experiment, you flip the coin 20 times and count the number of heads Y. In which experiment, do you expect to get the larger number of heads? In which experiment will you see the greater spread of values? Explain.

NEGATIVE BINOMIAL EXPERIMENTS

The 2004 baseball season was exciting since particular players had the opportunity to break single-season records. Let's focus on Ichiro Susuki of the Seattle Mariners who had the opportunity to break the season record for the most hits that was set by George Sisler in 1920. Sisler's record was 257 hits and Susuki had 255 hits before the Mariners' game on September 30. Was it likely that Susuki would tie Sisler's record during this particular game?

We can approximate this process as a coin-tossing experiment. When Susuki comes to bat, there are two relevant outcomes: either he will get a hit, or he will get an out. (We are ignoring other batting plays such as a walk or sacrifice bunt that don't result in a hit or an out.) Assume the probability that he gets a hit on a single at-bat is p = .372 (his 2004 batting average) and we can assume (for simplicity) that the outcomes on different at-bats are independent.

Susuki needs two more hits to tie the record. How many at-bats will it take him to get two hits?

This is not a binomial experiment since the number of trials is not fixed. Instead the number of successes (hits) is fixed in advanced and the number of trials to achieve this is random. Consider

Y = number of at-bats to get two hits.

We are interested in probabilities about the number of bats Y.

It should be obvious that Y has be at least 2 (he needs at least 2 at-bats to get 2 hits), but Y could be 3, 4, 5, etc. Let's find the probability that Y = 5.

First we know that the 2^{nd} hit must have occurred in the fifth trial (since Y=5). Also we know that there must have been one hit and three outs in the first four trials – there are ${}_4C_1$ ways of arranging the H's and the O's in these trials.

$$\underbrace{\_\_\quad \_\_\quad \_\_\quad \_\_}_{\text{H, 3 O's}}\quad \underline{\text{H}}$$

Also the probability of each possible outcome is $p^2(1-p)^3$, where p is the probability of a hit. So the probability that it takes 5 trials to observe 2 hits is

$$P(Y=5) = {}_4C_1\ p^2(1-p)^3.$$

Since p = .372 in this case, we get

$$P(Y=5) = {}_4C_1\ .372^2(1-.372)^3 = .1371$$

A general *negative binomial* experiment can be described as follows:

- We have a sequence of independent trials where each trial can be a success (S) or a failure.
- The probability of a success on a single trial is p.
- We continue the experiment until we observe r successes, and Y = number of trials we observe.

The probability that it takes us y trials to observe r successes is

$$P(Y=y) = {}_{(y-1)}C_{(r-1)}\ p^r(1-p)^{y-r},\ y = r, r+1, r+1, \ldots$$

Let's use this formula in our baseball example where r = 2 and p = .372. The table below gives the probabilities for the number of at-bats y = 2, 3, …, 9.

Y	P(y)
2	.1384
3	.1738
4	.1637
5	.1371

6	.1076
7	.0811
8	.0594
9	.0426

Note that it is most likely that Ichiro will only need three at-bats to get his two additional hits, but the probability of three at-bats is only 17%.   Actually each of the values 2, 3, 4, 5, and 6 have probabilities exceeding 10%. There is a significant probability that Ichiro will take a large number of bats – by adding the probabilities in the table, we see that the probability that Y is at most 9 is .904, so the probability that Y exceeds 9 is 1 - .904 = .096.

For a negative binomial experiment where Y is the number of trials needed to observe r successes, one can show that the mean value is

$$E(Y) = \frac{r}{p}.$$

For our baseball example, r = 2 and p = .372, so the expected number of at-bats to get two hits would be E(Y) = 2/.372 = 5.4.   It is interesting to note that although Y = 3 is the most probable value, Ichiro would average over 5 at-bats to get 2 hits in many repetitions of this random experiment.

PRACTICE: NEGATIVE BINOMIAL EXPERIMENTS

Suppose a candy box contains a large number of candies of which 30% are peppermint. You select candies from the box with replacement until you choose a peppermint.

1. Explain why this is a negative binomial experiment and give values of r and p.
2. If Y represents the number of draws until you choose a peppermint, find the probability that Y is equal to 2.
3. Find P(Y <= 3).
4. How many draws do you expect to make before you choose a peppermint?

5. I simulated this experiment 20 times and I've listed the values of Y for these twenty experiments. Use this data to approximate P(Y <=3) and E(Y) and compare your answers with your answers to parts 3 and 4.

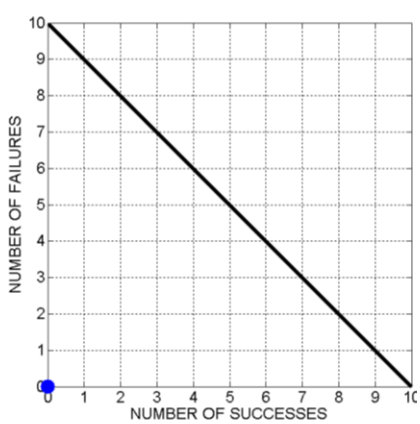
```
3   3   2   3   3  10   1  11   1   1   2   5   1   1   4   4   1   4   3   1
```

ACTIVITY – GRAPHING BINOMIAL AND NEGATIVE EXPERIMENTS

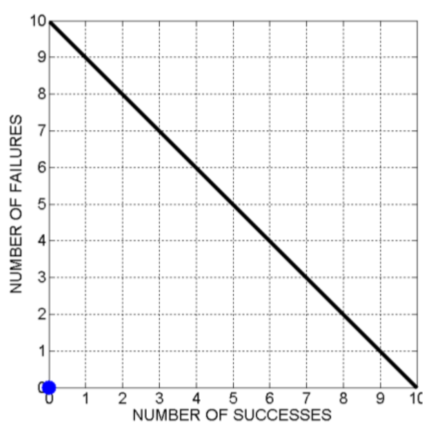
Suppose a basketball player makes a free throw shot with probability 0.7. We can simulate this shot using a ten-sided die **– if the die roll is between 1 and 7, she makes the shot; otherwise she misses the shot.** (If a ten-sided die is not available, then a table of random digits can be used.)

Using the die, simulate the shooting of 10 free throws for four games. Record your data on the graph. Start at the (0, 0) point (where there is a big dot). If the player makes the shot, draw a line one unit to the right; if she misses, draw a line one unit up. When you are done shooting 10 shots, your line should be on the dark diagonal line.

Game 1: Number of successes: _____

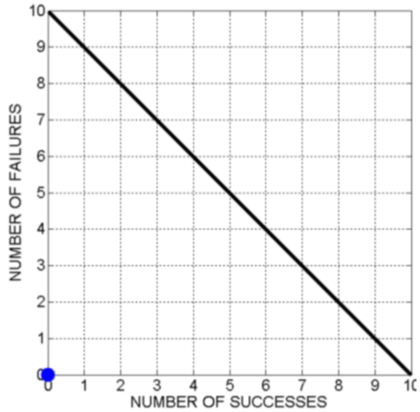


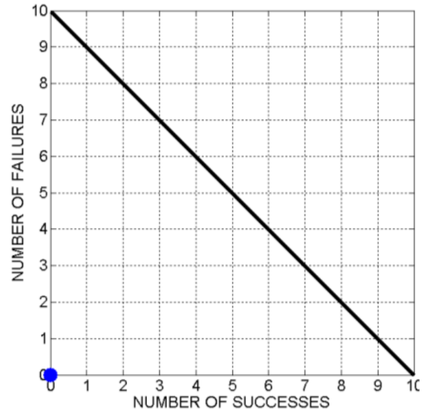
Game 2: Number of successes: _____



Game 3: Number of successes: _____

Game 4: Number of successes: _____





1. Record the number of makes (successes) for each game – put your answers in the blanks.

2. Collect the number of makes from all students in. Construct a suitable graph of these data.

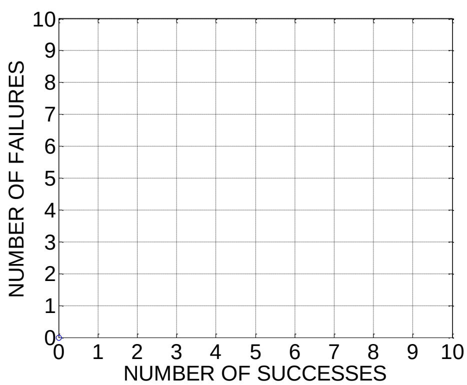
3. What is the most likely number of successes during a game of 10 shots?

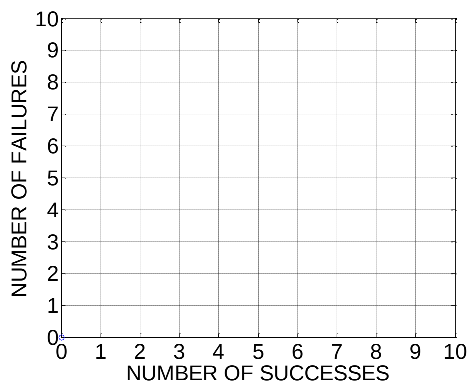
4. Find the probability she makes at least half her shots.

5. Find the probability she makes all of her shots.

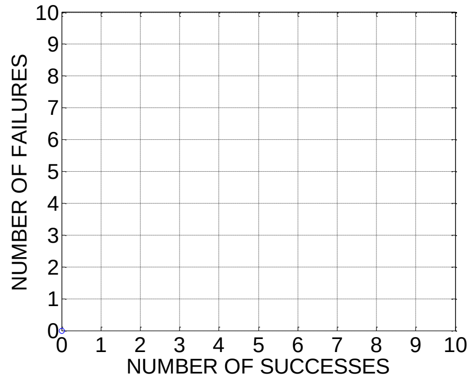
6. Suppose that the shooter continues to shoot free throws until she misses three shots. Use this same diagram to record the results of the individual shots until the experiment is completed. For each experiment, record Y = the total shots taken.

Game 1: Total shots = ______ Game 2: Total shots = ______





Game 1: Total shots = ______ Game 2: Total shots = ______



7. Collect the number of makes from all students in. Construct a suitable graph of these data.
8. What is the most likely number of shots taken?
9. What is the probability the woman will take at least 8 shots?

WRAP-UP

In this topic, we were introduced to the binomial experiment which represents a popular type of random experiment that resembles coin tossing. The experiment is a sequence of trials where there are two possible outcomes on each trial, the probability of a success is the same for each trial, and outcomes from different trials are independent. The focus is on the number of successes X that has a binomial distribution with parameters n, the number of trials, and p, the probability of success. A probability formula to compute P(X = k) was derived, and simple expressions were presented for the mean and standard deviation of X. The negative binomial experiment is a similar coin-tossing experiment where one continues sampling until one observes r successes and the random variable is Y, the number of trials.

EXERCISES

1. **Binomial Experiments**

Is each random process described below a binomial experiment? If it is, give values of n and p. Otherwise, explain why it is not binomial.

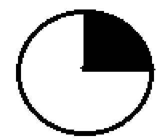
a. Roll a die 20 times and count the number of sixes you roll.

b. There is a room of 10 women and 10 men – you choose five people from the room without replacement and count the number of women you choose.

c. Same process as (b) but you sample with replacement instead of without replacement.

d. You flip a coin repeatedly until you observe 3 heads.

e. The spinner shown to the right is spun 50 times – you count the number of spins in the black region.



2. **Binomial and Negative Binomial Experiments**

Each of the random processes below is a binomial experiment, a negative binomial experiment, or neither. If the process is binomial, give values of n and p, and if the process is negative binomial, give values of r and p.

a. Suppose that 30% of students at a college regularly commute to school. You sample 15 students and record the number of commuters.

b. Same scenario as part a. You continue to sample students until you find two commuters and record the number of students sampled.

c. Suppose that a restaurant offers apple and orange juice. From past experience, the restaurant knows that 30% of the breakfast customers order apple juice, 50% order orange juice, and 20% order no juice. One morning, the restaurant has 30 customers and the numbers ordering apple juice, orange juice, and no juice are recorded.

d. Same scenario as part b. The restaurant only records the number ordering orange juice out of the first 30 customers.

e. Same scenario as part b. The restaurant counts the number of customers that order breakfast until exactly three order apple juice.

f. Same scenario as part b. Suppose that from past experience, the restaurant knows that 40% of the breakfast bills will exceed $10. Of the first 30 breakfast bills, the number of bills exceeding $10 is observed.

3. **Shooting Free Throws**

Suppose that Michael Jordan makes 80% of his free throws. Assume he takes 10 free shots during one game.

a. What is the most likely number of shots he will make?

b. Find the probability that he makes at least 8 shots.

c. Find the probability he makes more than 5 shots.

4. **Purchasing Audio CDs**

Suppose you know that 20% of the audio cd's sold in China are defective. You travel to China and you purchase 20 cd's on your trip.

a. What is the probability that at least one cd in your purchase is defective?

b. What is the probability that between 4 and 7 cd's are defective?

c. Compute the "average" number of defectives in your purchase.

5. **Rolling Five Dice**

Suppose you roll five dice and count the number of 1's you get.

a. Find the probability you roll exactly two 1's. [Do an exact calculation.]

b. Find the probability all the dice are 1's. [Do an exact calculation.]

c. Find the probability you roll at least two 1's. [Do an exact calculation.]

6. **Choosing Socks from a Drawer**

Suppose a drawer contains 10 socks, of which 4 are brown. I select 5 socks from the drawer with replacement.

a. Find the probability two of the five selected are brown.

b. Find the probability I choose more brown than non-brown.

c. How many brown socks do I expect to select?

d. Does the answer to part a change if we select socks from the drawer without replacement? Explain.

7. **Choosing Socks from a Drawer**

Suppose that I select socks from the drawer with replacement until I see two that are brown.

a. Find the probability that it takes me four selections.

b. Find the probability it takes more than 2 selections.

c. How many selections do I expect to make?

8. **Sampling Voters**

In your local town, suppose that 60% of the residents are supportive of a school levy that will be on the ballot in the next election. You take a random sample of 15 residents

a. Find the probability that a majority of the sample support the levy.

b. How many residents in the sample do you expect will support the levy?

c. If you sample the residents one at a time, find the probability that it will take you five residents to find three that support the levy.

9. **Taking a True/False Test**

Suppose you take a true/false test with twenty questions and you guess at the answers.

a. Find the probability you pass the test assuming that passing is 60% or higher correct.

b. Find the probability you get a B or higher where B is 80% correct.

c. If you get an 80% on this test, is it reasonable to assume that you were guessing? Explain.

10. **Bernoulli Experiment**

Let X_1 denote the result of one binomial trial, where $X_1 = 1$ if you observe a success and $X_1 = 0$ if you observe a failure. Find the mean and variance of X_1.

11. **Rolling a Die**

Suppose we roll a die until we observe a 6. This is a special case of a negative binomial experiment where r = 1 and p = 1/6. When we are interested in the number of trials until the first success, this is a geometric experiment and Y is a geometric random variable.

a. Find the probability that it takes you 4 rolls to get a 6.

b. Find the probability that it takes you more than 2 rolls to get a 6.

c. How many rolls do you need, on average, to get a 6?

12. **Heights of Male Freshmen**

Suppose that one third of male freshmen entering a college are over 6 feet tall. Four men are randomly assigned to a dorm room. Let X denote the number of men in this room that are under 6 feet tall. (You can ignore the fact that the actual sampling of men is done without replacement.)

a. Assuming X has a binomial distribution, what is a "success" and give values of n and p.

b. What is the most likely value of X? What is the probability of this value?

c. Find the probability that at least three men in this room will be under 6 feet tall.

13. **Basketball Shooting**

Suppose a basketball player is practicing shots from the free-throw line. She hasn't been playing for a while and she becomes more skillful in making shots as she is practicing. Let X represent the number of shots she makes in 50 attempts. Explain why the binomial distribution should not be used in finding probabilities about X.

14. **Collecting Posters from Cereal Boxes**

Suppose that a cereal box contains one of four posters and you are interested in collecting a complete set. You first purchase one box of cereal and find poster #1.

a. Let X_2 denote the number of boxes you need to purchase to find a different poster than #1. Find the expected value of X_2.

b. Once you have found your second poster, say #2, let X_3 denote the number of boxes you need to find a different poster than #1 or #2. Find the expected value of X_3.

c. Once you have collected posters #1, #2, #3, let X_4 denote the number of boxes you need to purchase to get poster #4. Find the expected value of X_4.

d. How many posters do you need, on average, to get a complete set of four?

15. **Baseball Hitting**

In baseball, it is important for a batter to get "on-base" and batters are rated in terms of their on-base percentage. In the 2004 baseball season, Bobby Abreu of the

Philadelphia Phillies had 705 "plate appearances" or opportunities to bat. Suppose we divide his plate appearances into groups of five – we record the number of times Abreu was on-base for plate appearances 1 through 5, for 6 through 10, for 11 through 15, and so on. If we let X denote the number of times on-base for five plate appearances, then we observe the following counts for X:

X	0	1	2	3	4	5	TOTAL
Count	10	29	44	40	15	3	141

To help understand this table, note that the count for X=1 is 29 – this means there were 29 periods where Abreu was on-base exactly one time. The count for X=2 is 44 – this means that for 44 periods Abreu was on-base two times.

Since each outcome is either a success or failure, where success is getting on-base, one wonders if the variation in these data can be explained by a binomial distribution.

X	0	1	2	3	4	5	TOTAL
P(X)							
Expected Count							

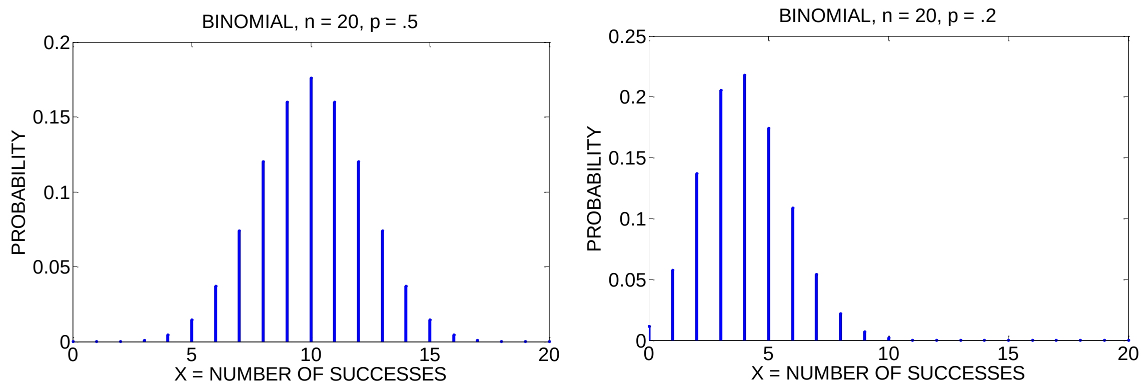
a. Find the probabilities for a binomial distribution with n=5 and p=.443. (This value of p is Abreu's on-base rate for the entire 2004 baseball season.) Place these probabilities in the P(X) row of the table.

b. Multiply the probabilities you found in part (a) by 141, the number of periods in the 2004 season. Place these numbers in the Expected Count row of the table. These represent the expected number of times Abreu would have 0, 1, 2, .., 5 times on-base if the probabilities followed a binomial distribution.

c. Compare the expected counts with the actual observed counts in the first table. Does a binomial distribution provide a good description of these data?

16. **Graphs of Binomial Distributions**

The below figures show the binomial distributions with n = 20 and p = .5 (left) and n = 20 and p = .2 (right).



Recall the 68 rule from Topic P6 that said that if a probability distribution is approximately bell-shaped, then approximately 68% of the probability falls within one standard deviation of the mean.

a. For the binomial distribution with n = 20 and p = .5, find the mean μ and standard deviation σ and compute the interval $(\mu - \sigma, \mu + \sigma)$.

b. Find the exact probability that X falls in the interval $(\mu - \sigma, \mu + \sigma)$.

c. Repeat parts a and b for the binomial distribution n = 20 and p = .2.

d. For which distribution was the 68% rule more accurate? Does that make sense based on the shapes of the two distributions?

17. **Guessing on a Test**

Students in a statistics class were given a five-question baseball trivia quiz. On each question, the students had to choose one of two possible answers. The number correct X was recorded for each student – a count table of the values of X is shown below.

X = number correct	Count	Probability	Expected
0	0		
1	3		
2	4		
3	7		

4	6		
5	1		

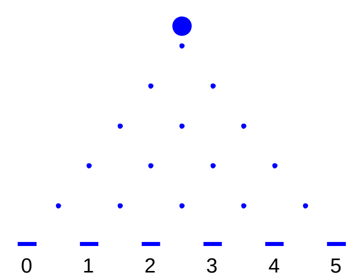
a. Suppose the students know little about baseball and so they are guessing on each question. If this is true, find the probability distribution of the number correct X.
b. Using this distribution, find the probability of each value of X and place these probabilities in the above table.
c. By multiplying these probabilities by the number of students (21), find the expected number of students for each value of X.
d. Compare your expected counts with the actual counts – does a binomial distribution seem like a reasonable assumption in this example?

18. **Playing Roulette**

Suppose you play the game roulette 20 times. Each game, you place a Trio Bet on three numbers and you win with probability 3/38.
a. Find the probability you win the game exactly two times.
b. Find the probability that you are winless in the 20 games.
c. Find the probability you win at least once.
d. How many games do you expect to win?

19. **The Galton Board**

Consider the Galton board described in the Spotlight at the beginning of this topic. A ball is placed above the first peg and dropped. When it strikes a peg, it is equally likely to fall left or right. The location at the bottom X is equal to the number of times that the ball falls right.



a. Explain why X has a binomial distribution and give the values of n and p.

b. Find P(X = 2).

c. Find the probability the ball falls to the right of the location "1".

d. Suppose that we change the experiment so that the probability of falling right is equal to 1/4. Explain how this changes the binomial experiment and find P(X = 2).

20. **Drug Testing**

In a *New York Times* article "Facing Questions, Rodriguez Raises More" (February 21, 2008), Major League Baseball is said to have a drug-testing policy where 600 tests are randomly given to a group of 1200 professional ballplayers. Alex Rodriguez claimed one season that he received five random tests.

a. If every player is equally likely to receive a single random blood test, what is the probability that Rodriguez gets tested?

b. If X represents the number of tests administered to Rodriguez among the 600 test, then explain why X has a binomial distribution and give the values of n and p.

c. Compute the probability that Rodriguez receives exactly one test.

d. Recall Rodriguez's claim that he received five random tests. Compute the probability of this event.

e. You should find the probability computed in part d to be very small. If Rodriguez is indeed telling the truth, what do you think about the randomness of the drug-testing policy?