

## Binomial Distribution (a very useful discrete one) Exercises

<u>Pages</u>	<u>Suggested Reading</u>
164 – 166	Section 4.5 (you can stop reading once it hits the “Notation” part on page 166)

<u>Pages</u>	<u>Problems</u>
186 – 193	<p>(<b>Section 4.15</b>) 7, 9, <a href="#">12</a>, 15, <a href="#">20</a> (in this one, don't actually <b>DO</b> the problem...but, explain why it's <b>not</b> binomial. Also, <a href="#">here's a tad bit extra</a>, for those of you who are interested in going above and beyond. 😊),  <a href="#">22</a>, 33-35, 41</p> <ul style="list-style-type: none"> <li>• If anyone needs a derivation of the formula for the binomial, <a href="#">here she is!</a></li> <li>• Here's a video that shows you how to <a href="#">use Excel to create a binomial distribution</a> (also shows you how to average it and create a histogram of the distribution).</li> </ul>

Hepatitis C (“Hep C”) is a virus affecting the liver, whose symptoms include inflammation, cirrhosis, and all sorts of other nastiness. According to the World Health Organization (WHO), 4.1% of people have Hep C. In humans, it is discovered via a routine blood test. In screening for Hep C, some health care providers, to save time and money, combine blood samples from 5 patients to test, and

- if the combined sample comes back negative, it means that all 5 folks are not infected with Hep C.
- if the combined sample comes back positive, it means that at least one of the people in the combined sample has Hep C. Then, each of the 5 must be individually tested to see who is infected.

Suppose 5 randomly selected folks have their blood taken, and their blood is placed into a combined sample. This combined sample is then tested for Hep C.

**E1.** Find the chance the combined sample comes back negative, to the nearest percent.

**E2.** Find the chance the combined sample comes back positive, to the nearest percent.

**E3.** Suppose a Hep C test (whether done on a single blood sample, or 5) costs around \$100. If you sample 100 random people, approximately how much is saved by using combined samples instead of individual testing (assuming you want to individually ID **all** those infected with Hep C)?

According to a 2010 CDC report, approximately 85% of Americans have health insurance. Suppose we randomly select 10 Americans.

**E4.** Does this sampling represent independent or dependent sampling?

**E5.** Find  $P(\text{all of them have health insurance})$ .

**E6.** Find  $P(\text{at least one of them does not have health insurance})$ .

**E7.** Find  $P(\text{half of them have health insurance})$ .

**E8.** Find the most likely number of them that has health insurance.

**E9.** Find the average number of those 10 that have health insurance.

**E10.** Find the standard deviation around that average.

**E11.** 95% of the time, at least how many people (out of 10) are insured?

**E12.** What would be an **unusual** number of them to have health insurance?

**E13.** Refer back to the Chuck – A – Luck game from the previous lesson. The number of 5’s gotten per three rolls of a die is actually a binomial random variable ( $n = 3$ ,  $p = 1/6$ ,  $x = 0,1,2,3$ ). Use this fact to create the T – table distribution for the number of 5’s gotten per three rolls in Chuck – A – Luck. How does your T – table compare to the one we got in class using Excel? Which values of the random variable are outliers? Why?

**E14.** Think back to our lovely chickens. Remember that, with 4 birds purchased, we average 3.6 girls (with  $p = 0.9$  for a sexed pullet). How many would we have to buy to **average** at least 4 girls? If you buy that many birds, what’s the chance you get at least 4 girls?

**Answers.**

**E1.**  $\approx 81\%$

**E2.**  $\approx 19\%$

**E3.** Let’s do the “individual test” case first...it’s the easier of the two, because the answer is just \$10,000 (100 people who each cost \$100).

Slightly trickier is the “groups of 5” cost. You’ll have to shell out \$2000 to test 20 groups of 5, right (20 groups times \$100)? So, we’re \$2000 in the hole.

From your answer in E1, you know that you’ll get a negative test result about 81% of the time. So, 16 of those 20 groups (roughly 81%) will test negative, and they’re done (they’re all clean).

However, the other 4 groups (roughly 19%) will test positively, so each of those 20 people will have to be tested individually, which is \$2000 more (20 people times \$100). Combining that cost with the initial out of \$2000, and you’ve spent \$4000 total to find those infected.

So, you spend about 60% less (\$6000) to test for Hep C in groups. Seems like a non – brainer to use combined testing all of the time...right?

Actually, no. Can you think of why it might not make financial sense to combine test? (hint: look at the rate of infection of Hep C).

In case you want to see a video of the solution of this one (some students mentioned they’d like that), [I made one for you!](#)

**E4.** Since the population is so vast and the sample so small, it’s essentially independent (remember this idea?)

**E5.** About 20%.

**E6.** About 80% (the complement of the last one).

**E7.** About 0.8% (I used the program)

**E8.** 9 (with probability 34.7%)

**E9.** 8.5

**E10.** 1.129

**E11.** Take a look at your DISTFILL L1 / L2 screen (the probability distribution...that’s it at right). Can you see how the probabilities of 0 through 3 people are so small as to basically be 0? Good. That means that chance of getting at most 3 uninsured people (out of 10) is, essentially, 0%. Make sense? (if not, look at #9 again). Now,  $P(4 \text{ insured}) \approx 0.1\%$ ,  $P(5 \text{ insured}) \approx 0.8\%$  and  $P(6 \text{ insured}) \approx 4\%$ , so the chance of getting at MOST 6 insured out of 10 is about 4.9% (either 0,1,2,3,4,5 OR 6 insured). Thus, there’s about a 95% chance of getting 7 or more (or, “at least 7”) insured folks out of ten!

L1	L2	L3	1
0	5.8E-9	-----	
1	3.3E-7		
2	8.3E-6		
3	1.3E-4		
4	.00125		
5	.00849		
6	.0401		
7	.12983		
8	.2759		
9	.34743		
10	.19687		
	-----		
L1(12) =			

**E12.** Whaddaya say?

**E13.** Close, I bet! And I’d use the “5%” outlier ID rule, since the data aren’t bell – shaped.

**E14.** To average 4 girls, you’d have to buy 5 birds (since  $\mu = np$ , if  $\mu = 4$  and  $p = 0.9$ ,  $n \approx 4.4$ . You’d have to round up to 5 to get the average you want). I get a 92% chance that you’d get at least 4 girls out of 5.

# Binomial Distribution (a very useful discrete one) Quizzes

## Quiz 1.

In December 2014 in Bend, 132 out of 1639 homes for sale were in foreclosure. We're going to use this population<sup>1</sup> to analyze the probabilities associated with sampling from this population. We'll do an experiment (as many folks looking for houses do) where we define the random variable  $X$  as the number of houses (out of a certain number,  $n$ , chosen) that are in foreclosure.

1. Suppose four homes for sale in Bend were randomly selected in December 2014.
  - a. **(1 point)** Start by completing the distribution below (use the Excel sheet to get the probabilities you need to at least 3 places):

<b>X = number of houses (out of 4) in foreclosure</b>	<b>P(X)</b>
<b>0</b>	
<b>1</b>	
<b>2</b>	
<b>3</b>	
<b>4</b>	

- b. **(2 points)** Why is it OK to use the binomial, even though we most likely sampled without replacement?
    - c. **(2 points)** What's the average number of Bend homes in foreclosure (out of 4)?
  2. **(2 points)** Suppose 10 homes for sale in Bend were randomly selected in December 2014. What's the average number of Bend homes in foreclosure (out of 10)?
  3. **(3 points)** Suppose  $n$  homes were selected from the population of Bend for sale homes. About how many homes would be in foreclosure? You'll have to answer as an expression that involves the variable " $n$ ". Assume that  $n$  is small enough to not trigger illegal use of the binomial (remember talking about that in #1 b?).
  4. **(extra 2 points)** How many homes would you have to select to be 99% certain that at least one of them was in foreclosure? Please explain *exactly* what you did!

---

<sup>1</sup> I used Zillow to get these values – therefore, as far as I know, I have all the possible “for sale” values - therefore, a population.

## Quiz 2.







Suppose you have an alarm clock that functions properly 50% of the time. That's right: flip a coin. If it comes up heads, it works. Tails, it doesn't (there's the binomial part). Now imagine that you have a very important meeting the next day, so you decide to purchase an additional alarm clock (also with 50% reliability). You know that you will awaken if **at least one** of the alarm clocks works.

1. **(2 points)** With only these two alarm clocks (each with 50% chance of working properly) what is the probability that you'll wake up?
  2. **(2 points)** OK, so that last probability wasn't high enough for you. You decide, instead, to buy a third alarm clock. With only these three alarm clocks (each with 50% chance of working properly) what is the probability that you'll wake up?
  3. **(6 points) (w)** Now you're frustrated, because you'd like to be at least 99% certain that you'll wake up. How many alarm clocks would you need to use to be 99% certain that you'll wake up? Explain exactly what you did to arrive at this answer.
  4. **(extra 1 point)** Refer back to question 1. Suppose that one of the two clocks has a 50% reliability, but you found a better one with 75% reliability. What is the chance that you'll wake up using just these two clocks (this is like buying baby chicks from two different suppliers with different sexing rates)?
- .....

### Quiz 3.

When playing a game of rock – paper – scissors, the probability of a player winning in any given game is  $\frac{1}{3}$  (see the chart below for verification). If you play exactly 5 games,

1. (1 point) what is the probability that you win exactly one game? For all of these questions, assume that you (and your opponent) are randomly deciding what to throw each round. Also, ties are not “do overs” – like in futbol, they’re a legitimate way to end a game.
2. (2 points) what is the probability that you win most of them?
3. (3 points) what is the probability that you win *or* tie most of them?
4. (2 points) what is the probability that you win *all* of them?
5. (2 points) what is the probability that you *lose* all of them?

		What your opponent throws...		
				
What you throw...		Tie.	You lose.	You win.
		You win.	Tie.	You lose.
		You lose.	You win.	Tie.

---

## Quiz 4.

I read an [article in Rolling Stone](#) that spoke of Sarah Palin's powerful political influence:

*"[Palin] throws her political weight behind candidates who deny the reality and risks of global warming. (More than half of the 64 candidates she endorsed in the midterm elections won.)"*

I thought that statistic had a bias: intending to prove to me that Palin must be some kind of political catalyst. So (as I tend to do), I dug deeper. As it turns out, exactly 33 of her 64 endorsements won. More than half? Yep. Significantly more than half? We shall see. 😊

What we're going to do in this quiz is use the idea of a probability distribution (in particular, the binomial) to approximate the average number of correct endorsements that I (or anyone, for that matter) could have gotten by simply **guessing** at names. You see, I don't believe that Sarah Palin's "33 out of 64" is all that impressive...in fact, I think that a good chunk of people could have done just as well if not better by guessing. But, I'd like to test that belief with some probability.

So, here's what I did...I looked at each and every one of the 64 races in which Palin endorsed an entrant:

- In 59 of those, there were either a) only two parties represented, or b) predominantly two parties represented (that is, the remaining parties added up to less than 5% of the popular vote, so, statistically, and unfortunately, they didn't matter in the election). Therefore, the chance of my guessing any of these correct is, basically, a coin flip (50%).
- In the remaining 5, three parties were all fairly well – represented (Dems, Repubs, and a 3<sup>rd</sup> party that earned a considerable chunk of the popular vote). I would need a three - sided coin to guess, wouldn't I? A die would do, I suppose...but, no matter what I actually **use**, the chance of me getting each of these correct is 1/3.

Now, in order to use the binomial to approximate this experiment, I need a fixed number of trials (got it! 64), and a probability of success. That probability needs to be weighted, doesn't it?

$$p = \frac{5(1/3) + 59(1/2)}{64} \approx 0.487$$

(hopefully that makes sense...if I were truly guessing with two people per race, it would be 50% in each case. It's close, but a **tad** lower because of the small number of 3 – people races, which would make it harder to guess at those outcomes)

Let's do it!

1. (1 point) What's the average number of correct guesses I should get out of 64 (to the nearest whole number)? Remember to use the Excel Calculator for this!
2. (1 point) What's the standard deviation around that average, to the nearest whole number?
3. (1 point) Take a screen shot of the graph accompanying the probability distribution and include it. Pretty bell – shaped, huh?
4. (2 points...one for each blank) 95% of the time, if Palin were **guessing**, she'd have gotten between \_\_\_\_\_ and \_\_\_\_\_ candidates correct.
5. (3 points) Based on your previous 3 answers, does it appear that guessing 33 correct out of 64 is unusual? Why or why not?
6. (2 points) At **least** how many endorsements would Sarah Palin have had to have gotten correct in order for me to have believed she was doing "better than guessing?" Use your answer from Number 4 to help!

## Quiz 5.

In class, we were mentioning the definition of a “combination” (also lovingly called the “binomial coefficient” by yours truly). If you remember, we were calculating these wonderful little coefficients by counting. Literally. Chickens, to be exact.

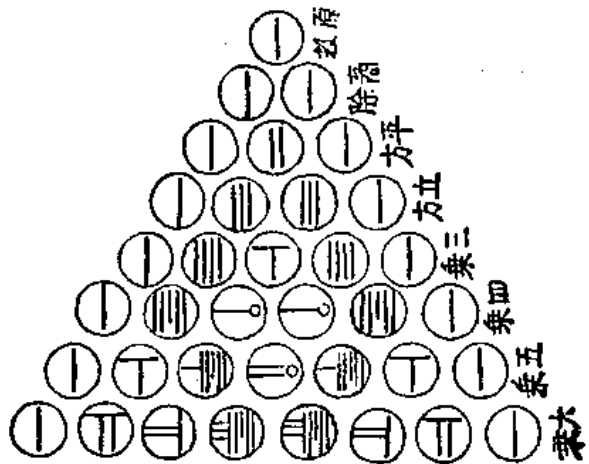
$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Take, for example, the calculation of the “6” that was the coefficient for the number of ways to have two hens and two roosters in 4 birds. Mathematically, there is a formula for a combination. There it is, at left. It represents the number of unordered ways to select  $r$  things from a group of  $n$  (in the case of chickens,  $n$  would be the total number of birds (4), and  $r$  would be the number of the birds that would be female (2).

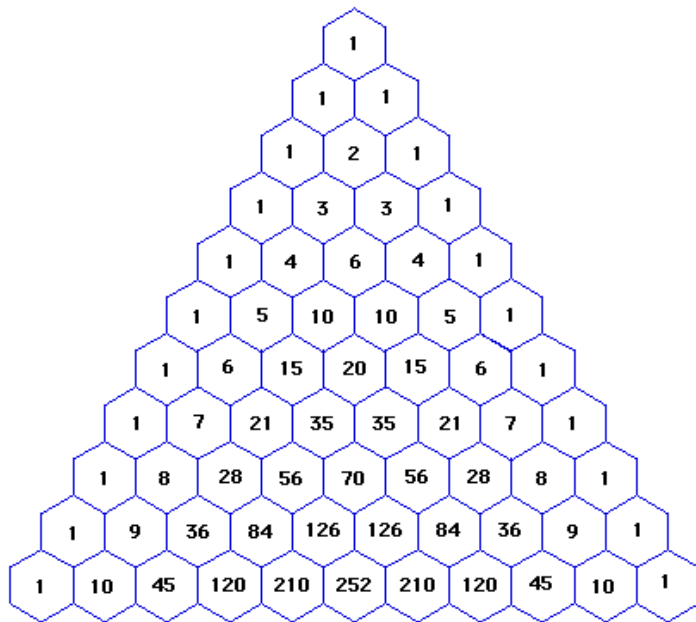
So, numerically, you’d have

$$\binom{4}{2} = \frac{4!}{(4-2)!2!} = \frac{24}{4} = 6$$

And that’s great...except, with technology, we don’t even need to ever worry about manual calculation of these things. Ever! So long as you understand what’s going on above, we’re good to go.



So...what I’d like to do now is show you a really neat connection between that formula and a nifty little mathematical construct called Pascal’s Triangle. Pascal’s Triangle is an arrangement of numbers that wasn’t discovered by Pascal. It was actually discovered well before. Above is a picture from Murai Chuzen's Sampo Doshi-mon (1781, Japan) of Pascal's Triangle. “Konichiwa!” you might say. “But we don’t read Japanese.” Right! So, let’s take a look at the Triangle in a non – picture language, shall we?



What you have at right are the first eleven rows of Pascal’s Triangle. You start with “row zero” (the 1 at the very top). Row 1 is the two 1’s beneath. Each successive row is found as follows: the first element and last elements are always 1’s, but every other element in the row is the sum of the two elements diagonally above it. For example, in row 3, the 3 is the sum of the 1 and 2 above it. Also, in row nine, the 84 is the sum of the 28 and 56 above.

**(2 points) Assuming this pattern repeats (it does...forever), what would the values in the next (11<sup>th</sup>) row be?**

Pretty neat, huh? “Yeah, sure...it’s neat. But what the heck *good* is it?” Well, you’re about to see its beauty *as well as* its usefulness. Look at the following:

$$\binom{0}{0} = 1$$

$$\binom{1}{0} = 1 \quad \binom{1}{1} = 1$$

$$\binom{2}{0} = 1 \quad \binom{2}{1} = 2 \quad \binom{2}{2} = 1$$

$$\binom{3}{0} = 1 \quad \binom{3}{1} = 3 \quad \binom{3}{2} = 3 \quad \binom{3}{3} = 1$$

$$\binom{4}{0} = 1 \quad \binom{4}{1} = 4 \quad \binom{4}{2} = 6 \quad \binom{4}{3} = 4 \quad \binom{4}{4} = 1$$

row number, and  $r$  is the position along that row. If you stare at that long enough, and compare it to the actual triangle, you will see where it applies.

Let's discover another neat result... check out the sums of each row's values. I've done a couple at right.

**(2 points)** List row 6's sum (remember...the top line is row ZERO).

**(2 points)** List row 11's sum.

**(2 points)** What is the sum of the values in row  $n$ ?

Back before instant technology, Pascal's Triangle was **the** best way to calculate combinations. However, since the advent of pocket calculators, its place in mainstream probability classes has waned a bit. But...that doesn't mean its beauty is any less relevant. Check this out:

✓ if you color all of the even values in the Triangle, you get a fractal<sup>2</sup>.

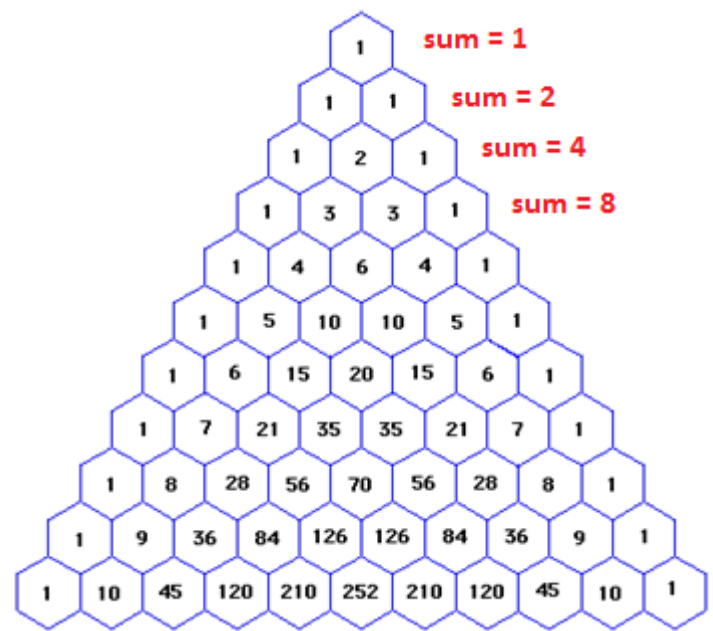
✓ The expansion of the binomial  $(x + y)^n$  has coefficients from row  $n$  in the triangle. For example,  $(x + y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6$  ...and if you think I distributed all of that out, you're giving me too much credit. ☺ Just look at the numbers in row 6.

**(2 points)** There's a cool result called the "hockey Stick" as it relates to Pascal's Triangle. Do a little Googling, and find out what it is! Source, please!

Rad! Totally, totally rad!

Isn't that **rad**? The first 5 rows of Pascal's Triangle (row 0 to row 4) are represented by all possible combinations with  $n$  from 0 to 4. Ha! And it continues infinitely! Row 5 contains the values 1,5,10,10,5 and 1. If we calculate the values of  $\binom{5}{x}$  for  $x = 0,1,2,3,4,5$ , we will arrive at those six values. I say again...HA!

The triangle is based on the identity  $\binom{n-1}{r-1} + \binom{n-1}{r} = \binom{n}{r}$ . This is what you all figured out in class; the "add the two numbers diagonally above to get the number centrally below." In that equation,  $n$  is the



Gorgeous, isn't it? By the way, here's a proof of why  $\binom{n-1}{r-1} + \binom{n-1}{r} = \binom{n}{r}$ :

$$\binom{n-1}{r-1} + \binom{n-1}{r} = \frac{(n-1)!}{(r-1)!(n-r)!} + \frac{(n-1)!}{r!(n-1-r)!} = \frac{r(n-1)!}{r!(n-r)!} + \frac{(n-r)(n-1)!}{r!(n-r)!} = \frac{(r+n-r)(n-1)!}{r!(n-r)!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$