Module 5.3: A Workbook on Bernoulli's Binomial Distribution Formula

Gregory V. Bard

October 11, 2020

Some Review

Before we start, we should review what we learned in the previous module. Recall the repetition formulas:

- If you have n independent repetitions of an attempt of event A, and if each attempt succeeds with probability p, then the probability of all n attempts succeeding is p^n.
- If you have n independent repetitions of an attempt of event A, and if each attempt succeeds with probability p, then the probability of all n attempts failing is $(1-p)^n$.

Now if I have n repetitions of an event, the last two bullets give me the probability of n success and 0 failures, or 0 successes and n failures. What about situations in between these two extremes?

Bernoulli's Binomial Distribution Formula

If you have n independent repetitions of an attempt of event A, and if each attempt succeeds with probability p (and therefore fails with probability $1-p$), the probability that A actually happens x times (and therefore fails to happen $n-x$ times), is given by

$$C_{n,x}(p^x)(1-p)^{n-x}$$

Easy-to-Remember Alternative: We can rename $1-p=q$, which we've seen before in Module 3.5: "The Square Root of npq Rule." We can also rename $n-x=y$, so that y is the number of failures. If we do that, then we get the following:

If you have n independent repetitions of an attempt of event A, and if each attempt succeeds with probability p and fails with probability q, the probability that A actually happens x times, and fails to happen y times, is given by

$$C_{n,x}(p^x)(q^y)$$

The second form is less common, but perhaps easier to remember. Note, $n=x+y$.

The Two Extreme Situations: Something cool happens for the two extreme situations. If $x=0$ we have

$$C_{n,x}(p^x)(1-p)^{n-x} = C_{n,0}(p^0)(1-p)^{n-0} = (1)(1)(1-p)^n = (1-p)^n$$

and if $x=n$ we have

$$C_{n,x}(p^x)(1-p)^{n-x} = C_{n,n}(p^n)(1-p)^0 = (1)(p^n)(1) = p^n$$

which means that the repetition formulas are just "special cases" of Bernoulli's Binomial Distribution Formula. On the other hand, using the repetition formulas can be vastly easier than using the Binomial Distribution Formula.

Using Sage to Save Time

The "Binomial Distribution Formula" helper is a Sage Applet that I wrote. It will save you enormous amounts of time by performing the tedious calculations for you. There is a link to it on `www.discrete-math-hub.com` right below the link to this workbook.

You definitely will want to use this applet to save you time while doing the homework problems. By the way, the code for the applet is freely available when you click on the applet. If you know some Python and some Sage, then you might learn something more by reading the code, but only after completing this module.

A Quick Cautionary Note about Hand Calculators

Bernoulli's Binomial Distribution Formula is particularly sensitive to rounding error. Once again, I remind you that you must never round in the middle of a math problem. You can only round at the end of a math problem. In Fall 2019, the question about the Bernoulli Binomial Distribution Formula on a midterm in Math-270: *Discrete Mathematics* was a bit of a bloodbath because about half the class decided to round at an intermediate stage, causing incorrect final answers.

Questions

1. A student has a multiple choice test of ten questions with five options each. Because the student has not studied, he has to guess. However, he is not entirely clueless, so he has a 40% chance of getting each question correct. That's in contrast to the 20% represented by blind guessing, for a 5-option multiple-choice question. What is the probability distribution of his performance? (i.e. the probability of getting x correct and $10 - x$ wrong, for $x \in \{0, 1, 2, \ldots, 10\}$.)

2. We return to the online company (from the previous module's Problem 9) with 8 servers. Each server has 99% up-time. The servers are scattered in hosting centers around the planet, so independence is a very fair assumption. What does the Binomial Distribution Formula give as the probability distribution for how many servers are up versus down at any given moment? (For example, we might want to know the probability of 6-up and 2-down.)

3. Suppose there is a biometric system which scans your palm and looks really cool, therefore the CEO says that it should be adopted at your computing company's server center. (It will impress the investors when they come to visit.) However, the reliability isn't all that great. If a legitimate user attempts to scan in, there is only a 90% chance that it will recognize the user. Moreover, when a non-user attempts to scan in, there is a 1% chance that they will get in anyway. These probabilities are not acceptable. Nonetheless, we can greatly improve the performance by reprogramming the device to scan three times, and taking the "majority vote" of the three scans.

   Here is the Binomial Distribution Table for a legitimate user:

   ```
   0 successes and 3 failures occurs with probability 0.000999999999999999
   1 successes and 2 failures occurs with probability 0.0270000000000000
   2 successes and 1 failures occurs with probability 0.243000000000000
   3 successes and 0 failures occurs with probability 0.729000000000000
   Grand Total: 1.00000000000000
   ```

   Here is the Binomial Distribution Table for a non-user:

   ```
   0 successes and 3 failures occurs with probability 0.970299000000000
   1 successes and 2 failures occurs with probability 0.0294030000000000
   2 successes and 1 failures occurs with probability 0.000297000000000000
   ```

```
3 successes and 0 failures occurs with probability 1.00000000000000e-6
Grand Total: 1.00000000000000
```

Using the "majority vote" of the three scans, what is the probability that a legitimate user is refused entry? granted entry? What is the probability that a non-user is refused entry? granted entry?

4. Now we return to the oil-drilling problems (the previous module's Problems 15 and 16). Recall that in a particular Alaskan oil field, the chance of striking oil after drilling is 5%. Find the probability distribution for the number of oil strikes, for x strikes with $x \in \{0, 1, 2, \ldots, 15\}$. Assume that the boss went with your suggestion of 59 drillings.

Background: Suppose that it is known that an altimeter on a particular model of commercial aircraft breaks with probability 1 in 800 on any given flight. An aircraft design firm has a decision to make: should they have five altimeters, taking the common output of 3 out of 5 to be the trustworthy altitude? Or should they have seven altimeters, taking the common output of 4 out of 7? Or should they have three altimeters, taking the common output of 2 out of 3? We will examine this situation in detail, over the course of questions 5–8. This is a serious matter. On the 20th of November, 1993, Avioimpex Flight 110 crashed with a loss of 116 lives, primarily because they were 2300 feet too high when they attempted to land. They didn't notice this visually, because it was during a blizzard.

5. To help analyze the matter of the altimeters mentioned above, make the Binomial Distribution Table for 3 altimeters. Do this by hand, just so that you know that you understand how to use the formula. Then, compute the probability of 2 or 3 altimeters being broken using that table.

6. Continuing with the previous problem, suppose that an intern gives you the Binomial Distribution Table for 5 altimeters. However, he spilled coffee on it, and a coffee stain has wiped out two entries, as shown. Find the two missing entries.

```
Probability 0 up and 5 down: 3.05175781250000e-15
Probability 1 up and 4 down: 1.21917724609375e-11
Probability 2 up and 3 down: 1.94824523925781e-8
Probability 3 up and 2 down: coffee stain
Probability 4 up and 1 down: coffee stain
Probability 5 up and 0 down: 0.993765605480954
Grand Total: 1.00000000000000
```

Furthermore, tell me the probability that the majority of the altimeters are broken.

7. Tell me why it was necessary, in the previous problem, that the coffee stain knock out two entries? Why would the math problem be broken (that is to say, far too easy) if I only had the coffee stain wipe out one entry of the Binomial Distribution Table?

8. Now let's examine the Binomial Distribution Table for seven altimeters. Compute the probability that the majority of the altimeters are broken.

```
0 successes and 7 failures occurs with probability 4.76837158203054e-21
1 successes and 6 failures occurs with probability 2.66695022582974e-17
2 successes and 5 failures occurs with probability 6.39267969131402e-14
3 successes and 4 failures occurs with probability 8.51291845560001e-11
4 successes and 3 failures occurs with probability 6.80182184602456e-8
5 successes and 2 failures occurs with probability 0.0000326079339298424
6 successes and 1 failures occurs with probability 0.00868457973664822
7 successes and 0 failures occurs with probability 0.991282744226010
Grand Total: 1.00000000000000
```

Note: I have added some analysis about the altimeter problems, from a systems engineering perspective in the answers section, after the answer to the previous problem.

9. I'm currently writing a paper[1] analyzing the infamous absence policy of Whole Foods, the expensive grocery store. Their policy is that if you are absent more than 5 times in any six month sliding window, you get fired. In six months, there are 130 working days. Suppose an employee is forced to call into work sick, for example, to take care of a child, with probability 3% on any given day. An ordinary person might think that they will be okay, because $(0.03)(130) = 3.9$ but it turns out that isn't exactly true.

   Here is the relevant Binomial Distribution Table. Compute the probability that this employee is fired in the first six months. Then, using the repetition formula, compute the probability that the person survives two years, taking it as four attempts of surviving for six months.

```
0 absences and 130 presences occurs with probability 0.0190688918050422
1 absences and 129 presences occurs with probability 0.0766687402470768
2 absences and 128 presences occurs with probability 0.152942280802158
3 absences and 127 presences occurs with probability 0.201820741677075
4 absences and 126 presences occurs with probability 0.198179645822076
5 absences and 125 presences occurs with probability 0.154457538393288
6 absences and 124 presences occurs with probability 0.0995216097894898
7 absences and 123 presences occurs with probability 0.0545243576460828
8 absences and 122 presences occurs with probability 0.0259271752208821
9 absences and 121 presences occurs with probability 0.0108698122919162
10 absences and 120 presences occurs with probability 0.00406777511542844
11 absences and 119 presences occurs with probability 0.00137244521232825
12 absences and 118 presences occurs with probability 0.000420930361513045
```

Closing Example: Next, I'd like to share some ideas that are not just a funny story, but an illustration of why the Binomial Distribution Formula depends strongly on independence. As I described in Module 3.1: "A Formal Introduction to Probability Theory," I once filled in for a colleague who was teaching Math-123: *Finite & Financial Mathematics*. There were 30-something students on the roster for each of two sections, so let's estimate with 33. As you might remember, only 5 students showed up to class for the first section, and only 1 student showed up to class for the second section. On the other hand, it was a Friday in April.

I don't know what his usual attendance was, but typically one might expect 75%, 70%, 65%, or 60%, roughly speaking. Surely it would be harsh to say that a business student is absent with probability 50%. Yet, even with probability 50%, let's see what happens. I have made a Binomial Distribution Table below, with 33 students, and 50% probability.

```
0 successes and 33 failures occurs with probability 1.16415321826935e-10
1 successes and 32 failures occurs with probability 3.84170562028885e-9
2 successes and 31 failures occurs with probability 6.14672899246216e-8
3 successes and 30 failures occurs with probability 6.35161995887756e-7
4 successes and 29 failures occurs with probability 4.76371496915817e-6
5 successes and 28 failures occurs with probability 0.0000276295468211174
6 successes and 27 failures occurs with probability 0.000128937885165215
7 successes and 26 failures occurs with probability 0.000497331842780113
8 successes and 25 failures occurs with probability 0.00161632848903537
9 successes and 24 failures occurs with probability 0.00448980135843158
```

[1] On September 23rd, 2020, this paper got accepted for presentation at the Joint Mathematics Meetings of the Mathematical Association of America and the American Mathematical Society, in January of 2021, which will be held online.

```
10 successes and 23 failures occurs with probability 0.0107755232602358
11 successes and 22 failures occurs with probability 0.0225306395441294
12 successes and 21 failures occurs with probability 0.0413061724975705
13 successes and 20 failures occurs with probability 0.0667253555729985
14 successes and 19 failures occurs with probability 0.0953219365328550
15 successes and 18 failures occurs with probability 0.120741119608283
16 successes and 17 failures occurs with probability 0.135833759559318
17 successes and 16 failures occurs with probability 0.135833759559318
18 successes and 15 failures occurs with probability 0.120741119608283
19 successes and 14 failures occurs with probability 0.0953219365328550
20 successes and 13 failures occurs with probability 0.0667253555729985
21 successes and 12 failures occurs with probability 0.0413061724975705
22 successes and 11 failures occurs with probability 0.0225306395441294
23 successes and 10 failures occurs with probability 0.0107755232602358
24 successes and 9 failures occurs with probability 0.00448980135843158
25 successes and 8 failures occurs with probability 0.00161632848903537
26 successes and 7 failures occurs with probability 0.000497331842780113
27 successes and 6 failures occurs with probability 0.000128937885165215
28 successes and 5 failures occurs with probability 0.0000276295468211174
29 successes and 4 failures occurs with probability 4.76371496915817e-6
30 successes and 3 failures occurs with probability 6.35161995887756e-7
31 successes and 2 failures occurs with probability 6.14672899246216e-8
32 successes and 1 failures occurs with probability 3.84170562028885e-9
33 successes and 0 failures occurs with probability 1.16415321826935e-10
Grand Total: 1
```

As you can see, the probability of 1 student being present, while 32 are absent, is very small ($3.84170 \cdots 10^{-9}$) and based on the Binomial Distribution Formula, we are shocked that only 1 student came to class for that second section. It is much less likely than "a one in a million" event. Moreover, if we replace the 50% with perhaps 60%, 65%, 70%, or 75%, then that probability of $3.84170 \cdots 10^{-9}$ will become even smaller.

Nonetheless, it was a sunny Friday afternoon in April, after a long winter. The Bernoulli Random Variable inside each student that decides "present" or "absent" is not independent, in the following sense: the sunshine and warm temperatures affected all 33 students. They all have the temptation, induced by the warmth and the light of the sun, affecting them all at the same time. This destroys independence.

(This is kind of like my note at the end of the answer to Problem #3 of the previous module, where a problem in the wave soldering machine caused both the screen and the battery to be damaged simultaneously, much more often than we would otherwise expect.)

Answers

1. According to Sage:

```
0 successes and 10 failures occurs with probability 0.00604661760000000
1 successes and 9 failures occurs with probability 0.0403107840000000
2 successes and 8 failures occurs with probability 0.120932352000000
3 successes and 7 failures occurs with probability 0.214990848000000
4 successes and 6 failures occurs with probability 0.250822656000000
5 successes and 5 failures occurs with probability 0.200658124800000
```

```
6 successes and 4 failures occurs with probability 0.111476736000000
7 successes and 3 failures occurs with probability 0.0424673280000000
8 successes and 2 failures occurs with probability 0.0106168320000000
9 successes and 1 failures occurs with probability 0.00157286400000000
10 successes and 0 failures occurs with probability 0.000104857600000000
Grand Total: 1.00000000000000
```

He's most likely to get 30%, 40%, or 50%, with an intermediate chance of 20% or 60%.

2. According to Sage,

```
0 successes and 8 failures occurs with probability 1.00000000000001e-16
1 successes and 7 failures occurs with probability 7.92000000000005e-14
2 successes and 6 failures occurs with probability 2.74428000000001e-11
3 successes and 5 failures occurs with probability 5.43367440000002e-9
4 successes and 4 failures occurs with probability 6.72417207000002e-7
5 successes and 3 failures occurs with probability 0.0000532554427944001
6 successes and 2 failures occurs with probability 0.00263614441832280
7 successes and 1 failures occurs with probability 0.0745652278325593
8 successes and 0 failures occurs with probability 0.922744694427920
Grand Total: 1.00000000000000
```

Notice that the 8 success and 0 failures matches our previous result of $0.922744\cdots$.

3. A legitimate user will be granted entry with probability

$$0.729 + 0.243 = 0.972$$

whereas a non-user will be granted entry with probability

$$0.000297 + 10^{-6} = 0.000298$$

A non-user will be denied entry with probability

$$0.970299 + 0.029403 = 0.999702$$

whereas a legitimate user will be denied entry with probability

$$0.027 + 0.0009\overline{9} = 0.028$$

As you can see, the product is greatly improved by this change. In fact, it is improved by a surprisingly large magnitude. Also, I should mention that you can switch a "probability that the user is granted entry" into a "probability that the user is denied entry" by using the complement principle.

4. According to Sage,

```
0 successes and 59 failures occurs with probability 0.0484945252494231
1 successes and 58 failures occurs with probability 0.150588262616630
2 successes and 57 failures occurs with probability 0.229845242941172
3 successes and 56 failures occurs with probability 0.229845242941172
4 successes and 55 failures occurs with probability 0.169359652693495
5 successes and 54 failures occurs with probability 0.0980503252436023
```

```
6 successes and 53 failures occurs with probability 0.0464448909048643
7 successes and 52 failures occurs with probability 0.0185081144207354
8 successes and 51 failures occurs with probability 0.00633172335446211
9 successes and 50 failures occurs with probability 0.00188840871975186
10 successes and 49 failures occurs with probability 0.000496949663092594
11 successes and 48 failures occurs with probability 0.000116509729624579
12 successes and 47 failures occurs with probability 0.0000245283641314904
13 successes and 46 failures occurs with probability 4.66734054323907e-6
14 successes and 45 failures occurs with probability 8.07134078906004e-7
15 successes and 44 failures occurs with probability 1.27442222985159e-7
Grand Total: 1.00000000000000
```

As you can see, getting 15 strikes has probability around 1 in 8 million, so we don't really need to continue our tabulations past 15 strikes.

5. Here is the Binomial Distribution Table for three altimeters, computed by Sage.

```
Probability 0 successes and 3 failures: 1.95312500000000e-9
Probability 1 successes and 2 failures: 4.68164062500000e-6
Probability 2 successes and 1 failures: 0.00374063085937500
Probability 3 successes and 0 failures: 0.996254685546875
Grand Total: 1.00000000000000
```

The probability of having an emergency (because 2 or 3 out of the 3 altimeters are broken) is given by

$$1.95312\cdots \times 10^{-9} + 4.68164\cdots \times 10^{-6} = 4.68359\cdots \times 10^{-6}$$

6. Here is the Binomial Distribution Table for five altimeters

```
Probability 0 successes and 5 failures: 3.05175781250000e-15
Probability 1 successes and 4 failures: 1.21917724609375e-11
Probability 2 successes and 3 failures: 1.94824523925781e-8
Probability 3 successes and 2 failures: 0.0000155664794616699
Probability 4 successes and 1 failures: 0.00621880854493713
Probability 5 successes and 0 failures: 0.993765605480954
Grand Total: 1.00000000000000
```

Note, you should check the answer by seeing that it all adds to 1.

The probability that a majority of the altimeters are not operating turns out to be 1.94946×10^{-8}.

7. If I had given you a table with only one stain missing, the problem would be far too easy. That's because all the probabilities add to 1. If you were to add up all the given probabilities, then the missing entry must be one minus that sum. This would not test your knowledge of the Binomial Distribution Formula.

8. For operating with seven altimeters, we have a probability of $8.51931\cdots \times 10^{-11}$ that the majority of the altimeters would be broken.

Analysis: Here is a summary of what we learned from the altimeter problems.

- Best 2 out of 3: $4.68359\cdots \times 10^{-6}$, or 1 in $213,511$.
- Best 3 out of 5: $1.94946\cdots \times 10^{-8}$, or 1 in $51,296,200$.
- Best 4 out of 7: $8.51931\cdots \times 10^{-11}$, or 1 in $11,738,000,000$.

Clearly, these probabilities are not roughly the same—they are shockingly different. Normally, when human life is at stake, a probability of 1 in one million is called for. That rules out the 2-out-of-3 model. However, for a decent sized airliner, you could lose 150–250 lives at once. So the 2-out-of-3 model is completely unacceptable. On the other hand, 1 in 11 billion probability is overkill in many ways. Even if the only thing that the airplane designer cared about was the safety of human lives, the money that you spend on the extra altimeters is money that you are not spending on the tires, the fuel pumps, the on board emergency oxygen tanks, the radar, *et cetera*.... So it seems that the 3-out-of-5 model is the way to go.

9. The probability that the employee (who is absent 3% of work days) will survive the first six months is given by

$$0.0190688\cdots + 0.0766687\cdots + 0.152942\cdots + 0.201820\cdots + 0.198179\cdots + 0.154457\cdots = 0.803137\cdots$$

so the probability that they will be fired is given by

$$1 - 0.803137\cdots = 0.196862\cdots$$

and the probability that they survive two years is given by

$$(0.803137\cdots)^4 = 0.416064\cdots$$

This policy disadvantages mothers, especially single mothers, who might have to call in as absent to work in order to take care of a sick child. To read more about this, I recommend the article by Jon Graef, "Fired Whole Foods Worker Rhiannon Broschat-Salguero: 'I'm Not Going To Choose My Job Over My Son,"' published in the *The Chicagoist*, on February 9th, 2014.

`http://chicagoist.com/2014/02/09/fired_whole_foods_worker_rhiannon_b.php`

Interestingly, because it is a sliding window, our method actually underestimates the probability of getting fired during the first two years. Consider an employee who gets only two absences during the first 6 months. If those occur late in the window, and are followed by four absences during the next 6 months, then it is probable that some sliding window of length 182 days or less includes all six absences, and the employee would get terminated. An employee is actually more likely to be fired than these computations indicate.

Last but not least, it has been mentioned to me (but I have not confirmed) that this policy has been abandoned by The Whole Foods company. In a certain sense, by abandoning the policy, they have admitted that the policy was a bad idea.