DISCRETE RANDOM VARIABLES
EXCEL LAB #3

Overview

This lab is written for Excel 2011, which is available to students in the library. The notation => can be read as "go to" or "click on." This notation will most often be used when navigating the menu or toolbars in Excel. To indicate a command or icon that you might click on or search for in Excel, **bold** will be used. Likewise, anything that you are to type into Excel will be **bolded** in the instructions. Please do not enter such text as bolded text unless the instructions ask you to do so.

Tutorial

1. Open Excel. Name the left-most worksheet, current called Sheet 1, to **Binomial**.

2. Before starting work with discrete random variables, we will format the worksheet a bit.

    1. Set the font and font size to Arial 10 point font for columns A through J.

    2. Set the width of column A to 3, and left justify all of the column A cells: (hint) **right click on column A** => **Column Width** => **3** => **OK** => **left click on column A** => **Home** tab => **Alignment** box => **click on the left justify (horizontally) icon** (first on the left in the second row).

    3. Set the width of columns B – D to 10, and center justify these columns.

    4. Set the numeric format of cells C7 through D31 to include 6 digits after the decimal point: (hint) **left click on cell C7 and drag through cell D31** => **right click on the shaded area** => **Format Cells** => **Number** (tab) => **Number** (category) => **6** (decimal places) => **OK**.

    5. Set the width of column F to 20, and left justify all of the column F cells.

    6. Set the width of column G to 10, and right justify all of the column G cells.

    7. Set the numeric format of column G to percentages with 2 digits after the decimal point.

In this worksheet we are going to create the probability distribution function (pdf) and cumulative distribution function (cdf) for a binomial distribution with 24 draws and then use the pdf and cdf to answer some probability questions. Recall that in the instructions items you will type into the spreadsheet are in bold but you are only to type them in as bold if the instructions explicitly say to do so. If the instructions are underlined, however, then have your worksheet include the underlining.

3. In cell A1 enter **<u>Binomial PDF and CDF</u>** (bold and underlined).

4. In cell B3 enter **<u>n</u>** (underline, no bold). Recall that n is the number of trials in a binomial distribution. In cell C3 enter **24**. By doing this, if we wanted to change the number of trials in the distribution, we would only need to change cell C3.

5. In cell B4 enter <u>π</u> (this is the Greek letter pi, underlined, no bold): **click on cell B4** => **Insert** tab => **Symbols** box (usually top, far right) => **Symbol** => scroll to find lower-case π among the lower-case Greek letters => click on the symbol => **Insert** => **Close**. (Hint: if you can't easily scroll through all of the symbols and find the Greek section and then find π, you can also enter the character code of 03C0.) For a binomial distribution, π is the (independent) probability that each trial will produce a success. In cell C4 enter 0.4 as the probability of a success.

6. In cell B6 enter **<u>x</u>**. Below <u>x</u> we are going to list the number of "successes" from our binomial distribution. As there are 24 draws in our distribution, we can have 0, 1, 2, …, 23, or 24 successes. In cell B7 enter **0**. In cell B8 enter **=1+B7**. A **1** should appear in cell B8, but the formula appears in the textbox. Copy and paste cell B8 into cells B9 – B31. Cells B7 through B31 should now have the numbers 0 through 24.

7. In cell C6 enter **<u>p(x)</u>**. Here <u>p(x)</u> represents the pdf of the distribution. That is, <u>p(x)</u> = Pr(X = x). To calculate binomial probabilities, we will use Excel's **BINOMDIST** function. In cell C7 enter **=BINOMDIST(B7,C3,C4,FALSE)**. Excel's **BINOMDIST** function requires four pieces of information:
    1. The number of successes, x, which for us is listed in column B.
    2. The number of trials, n, which for us is listed in cell C3 which is why we must put dollar signs before the C and before the 3.
    3. The probability of a success, π, which for us is listed in cell C4 which is why we must put dollar signs before the C and before the 4.
    4. Whether we want Excel to report the pdf of the distribution (which is indicated by FALSE) or the cdf of the distribution (which is indicated by TRUE).

8. Copy and paste cell C7 into cells C8 through C31. To check, you should find that Pr(X = 8) equals 0.135977 in cell C15.

9. In cell D6 enter **<u>F(x)</u>**. Here <u>F(x)</u> represents the cdf of the distribution. That is, <u>F(x)</u> = Pr(X ≤ x). To calculate the cumulative probabilities for the binomial, we will again use the **BINOMDIST** function. In cell D7 enter **=BINOMDIST(B7,C3,C4,TRUE)**.

    Notice that this is the same formula we entered into cell C7 to calculate the pdf except the last parameter is set to TRUE rather than to FALSE. By setting this value to TRUE, Excel calculates the cdf rather than the pdf. Copy and paste cell D7 into cells D8 through D31. To check, you should find that Pr(X ≤ 8) equals 0.327922 in cell D15.

One of the more useful things about creating the pdf and cdf this way is that we can change the probability of a success in cell C4, and all of the probabilities will change automatically. For example, when $\pi = 0.4$, with 24 draws, the probability of having exactly 6 successes is $p(6) = 0.055991$ and the probability of having at most 6 successes is $F(6) = 0.095961$, which are reported in cells C13 and D13 respectively. Now change cell C4 to 0.2, and notice that the values in cells C7 through D31 automatically change. Now X is distributed binomial with 24 draws, each with a probability of success of 0.2. In this case, the probability of having exactly 6 successes is $p(6) = 0.155179$ and the probability of having at most 6 successes $F(6) = 0.811071$, which are still reported in cells C13 and D13 respectively. Notice that because of the lower probability of success, more of the probability has been shifted to this lower end of the distribution.

10. We will now use the pdf for our random variable, $X \sim Bin(0.2, 24)$ to answer some probability questions. Make sure you are using $\pi = 0.2$ (i.e., double-check that the value in cell C5 in your worksheet is 0.2).
    1. In cell F7 enter **Q1. Pr(X=5)?**
    2. In cell F8 enter **Q2. Pr(X=3 or X=4)?**
    3. In cell F9 enter **Q3. Pr(X>3)?**
    4. In cell F10 enter **Q4. Pr(X<6)?**
    5. In cell F11 enter **Q5. Pr(3<X<8)?**

11. In cells G7 through G11 we want to answer these three questions using cells C7 through D31 so that, if we changed the probability of success, the answers would then change automatically.
    1. In cell G7 enter **=C12**.
    2. In cell G8 enter **=C10+C11**.
    3. In cell G9 enter **=SUM(C11:C31)**, which calculates $p(4) + p(5) + \ldots p(24)$.
    4. In cell G10 enter **=SUM(C7:C12)**, which calculates $p(0) + p(1) + \ldots + p(5)$. Alternatively, notice that the Q4 is asking a cdf question, $Pr(X < 6)$. Thus, in cell G10, we could have simply entered **=D12**. But if you do this, be sure to enter the correct cell. As the question is $Pr(X < 6)$, the correct cell corresponds to $F(5)$, which is D12. Had the question been $Pr(X \leq 6)$, then the correct cell would correspond to $F(6)$, which is D13.
    5. In cell G11 enter **=SUM(C11:C14)**, which calculates $p(4) + p(5) + p(6) + p(7)$.

Part of your worksheet should now look like the following.

Binomial PDF and CDF

n	24
π	0.2

x	p(x)	F(x)		
0	0.004722	0.004722	Q1. Pr(X=5)?	19.60%
1	0.028334	0.033057	Q2. Pr(X=3 or X=4)?	34.54%
2	0.081461	0.114517	Q3. Pr(X>3)?	73.61%
3	0.149345	0.263862	Q4. Pr(X<6)?	65.59%
4	0.196015	0.459877	Q5. Pr(3<X< 8)?	64.70%

12. Now suppose the probability of success changes to 0.32. Make this change by entering 0.32 into cell C4, and the probabilities in the pdf and cdf change automatically, as do the answers to the questions. Following this change, the top part of your worksheet will look like the following.

Binomial PDF and CDF

n	24
π	0.32

x	p(x)	F(x)		
0	0.000096	0.000096	Q1. Pr(X=5)?	9.37%
1	0.001079	0.001175	Q2. Pr(X=3 or X=4)?	6.99%
2	0.005840	0.007015	Q3. Pr(X>3)?	97.28%
3	0.020153	0.027168	Q4. Pr(X<6)?	17.07%
4	0.049791	0.076959	Q5. Pr(3<X< 8)?	45.22%

13. Save this file as YourName_Lab3.xlsx.

14. We are going to repeat the previous tutorial for the Poisson distribution. In the same file, rename the second worksheet, currently called Sheet 2, to **Poisson**. Then:

   1. Set the font and font size to Arial 10 point font for columns A through J.
   2. Set the width of column A to 3, and left justify all of the column A cells.
   3. Set the width of columns B – D to 10, and center justify these columns.
   4. Set the numeric format of cells C7 through D47 to include 6 digits after the decimal point. (This is the only difference from task #2 above. Here we need through row 47 not 31, because of the Poisson distribution.)
   5. Set the width of column F to 20, and left justify all of the column F cells.
   6. Set the width of column G to 10, and right justify all of the column G cells.
   7. Set the numeric format of column G to percentages with 2 digits after the decimal point.

In this worksheet we are going to create the probability distribution function (pdf) and cumulative distribution function (cdf) for a Poisson distribution and then use the pdf and cdf to answer some probability questions. One issue/problem that we will encounter is that the possible values for x for a Poisson are 0, 1, 2, … to infinity. We obviously cannot (and do not want to) consider thousands of potential x values. As we will see, this will not be a huge problem.

15. In cell A1 enter **Poisson PDF and CDF** (bold and underlined).

16. In cell B3 enter **λ** (this is the Greek letter lambda, underlined, no bold). To enter this symbol in Excel: **click on cell B3** => **Insert** tab => **Symbols** box (usually top, far right) => **Symbol** => scroll to find lower-case λ among the lower-case Greek letters => click on the symbol => **Insert** => **Close**. (Hint: if you can't easily scroll through all of the symbols and find the Greek section and then find π, you can also enter the character code of 03BB.)

17. For a Poisson distribution, λ is the expected number of "arrivals" during the given length of time. To begin, we will expect 5 arrivals in any given hour. In cell C3 enter **5** as the expected number of arrivals in an hour.

18. In cell B6 enter **x**. Below x we are going to list the number of "successes" from our Poisson distribution, up to 40. In cell B7 enter **0**. In cell B8 enter **=1+B7**. Copy and paste cell B8 into cells B9 – B47.

19. In cell C6 enter **p(x)**. Here p(x) represents the pdf of the distribution. That is, p(x) = Pr(X = x). To calculate Poisson probabilities, we will use Excel's **POISSON** function. In cell C7 enter **=POISSON(B7,C3,FALSE)**. Excel's **POISSON** function requires the following three pieces of information:
    1. The number of successes, x, which for us is listed in column B.
    2. The expected number of arrivals in the given time period, λ, which for us is listed in cell C3 (which is also why we must put dollar signs before the C and the 3).
    3. Whether we want Excel to produce the pdf of the distribution, which is indicated by FALSE, or the cdf of the distribution, which is indicated by TRUE.

20. Copy and paste cell C7 into cells C8 through C47. To check, you should find that Pr(X = 8) = 0.065278 in cell C15. Notice too why we don't need to worry (too much) about there potentially being a large number of arrivals. The probability of having 19 arrivals, p(19) is 0.000001, and any number greater than 19 is even less than this.

21. In cell D6 enter **F(x)**. Here F(x) represents the cdf of the distribution. That is, F(x) = Pr(X ≤ x). To calculate the cumulative probabilities for the binomial, we will again use the **POISSON** function. In cell D7 enter **=POISSON(B7,C3,TRUE)**.

    Notice that this is the same formula we entered into cell C7 to calculate the pdf except the last parameter is set to TRUE rather than to FALSE. By setting this value to TRUE, Excel calculates the cdf rather than the pdf. Copy and paste cell D7 into cells D8 through D47. To check, you should find that Pr(X ≤ 8) = 0.931906 in cell D15. You should also see that cell D7 equals cell C7 as Pr(X = 0) = Pr(X ≤ 0) for a Poisson and F(x) = 1.000000 for all x ≥ 20 which is attributable to rounding.

One of the more useful things about creating the pdf and cdf in the way that we have is that we can change the expected value of arrivals in cell C3 and all of the probabilities will change automatically. For example, when λ = 5, the probability of having exactly 6 arrivals is p(6) = 0.146223 and the probability of having at most 6 successes is F(6) = 0.762183, which are reported in cells C13 and D13 respectively. Now change cell C3 to 8, and notice that the values in cells C7 through D47 automatically change. In this case, the probability of having exactly 6 successes is p(6) = 0.122138 and the probability of having at most 6 successes is F(6) = 0.313374, which are still reported in cells C13 and D13 respectively. Notice that because of the higher expected number of arrivals, less of the probability has been shifted to this lower end of the distribution.

22. We will now use the pdf for our random variable, X ~ Pois(8) to answer some probability questions. Make sure you are using λ = 8 in cell C3.
    1. In cell F7 enter **Q1. Pr(X=5)?**
    2. In cell F8 enter **Q2. Pr(X=3 or X=4)?**
    3. In cell F9 enter **Q3. Pr(X>3)?**
    4. In cell F10 enter **Q4. Pr(X<6)?**
    5. In cell F11 enter **Q5. Pr(3<X<8)?**

5

23. In cells G7 through G11 we want to answer these three questions using cells C7 through D47 so if we changed the probability of success, the answers would change automatically.
    1. In cell G7 enter **=C12**.
    2. In cell G8 enter **=C10+C11**.
    3. In cell G9 enter **=SUM(C11:C47)**.
    4. In cell G10 enter **=SUM(C7:C12)**.
    5. In cell G11 enter **=SUM(C11:C14)**.

Part of your worksheet should now look like the following.

Poisson PDF and CDF

λ   8

x	p(x)	F(x)		
0	0.000335	0.000335	Q1. Pr(X=5)?	9.16%
1	0.002684	0.003019	Q2. Pr(X=3 or X=4)?	8.59%
2	0.010735	0.013754	Q3. Pr(X>3)?	95.76%
3	0.028626	0.042380	Q4. Pr(X<6)?	19.12%
4	0.057252	0.099632	Q5. Pr(3< X < 8)?	41.06%

Now suppose the expected number of arrivals increases to 12. Make this change by entering 12 into cell C3, and the probabilities in the pdf and cdf change automatically, as do the answers to the questions. Following this change, the top part of your worksheet will look like the following.

Poisson PDF and CDF

λ   12

x	p(x)	F(x)		
0	0.000006	0.000006	Q1. Pr(X=5)?	1.27%
1	0.000074	0.000080	Q2. Pr(X=3 or X=4)?	0.71%
2	0.000442	0.000522	Q3. Pr(X>3)?	99.77%
3	0.001770	0.002292	Q4. Pr(X<6)?	2.03%
4	0.005309	0.007600	Q5. Pr(3< X < 8)?	8.72%

24. We are going to repeat the previous tutorial for the Hypergeometric distribution. Rename the third worksheet, currently called Sheet 3, to **Hypergeometric**. Then:

    1. Set the font and font size to Arial 10 point font for columns A through J.
    2. Set the width of column A to 3, and left justify all of the column A cells.
    3. Set the width of columns B – D to 10, and center justify these columns.
    4. Set the numeric format of cells C7 through D15 to include 6 digits after the decimal point.
    5. Set the width of column F to 20, and left justify all of the column F cells.
    6. Set the width of column G to 10, and right justify all of the column G cells.
    7. Set the numeric format of column G to percentages with 2 digits after the decimal point.

In this worksheet we are going to create the probability distribution function (pdf) and cumulative distribution function (cdf) for a hypergeometric distribution and then use the pdf and cdf to answer some probability questions.

25. In cell A1 enter **Hypergeometric PDF and CDF** (bold and underlined).

26. In cell B3 enter **N**. In cell C3 enter **20**. Recall that for a hypergeometric distribution, N is the total number of items that can be selected, i.e., the total number of items "in the bin."

27. In cell B4 enter **r**. Recall that for a hypergeometric distribution, r is the total number of "successes" in the bin. In cell C4 enter **12**. So, of our 20 items, 12 are successes.

28. In cell B5 enter **n**. Recall that for a hypergeometric distribution, n is the number of draws that will be made. In cell C5 enter **8**. So we are modeling the random variable X where X $\sim$ Hyper(20,12,8).

29. In cell B6 enter **x**. Below x we are going to list the number of "successes" from our hypergeometric distribution. As there are 8 draws in the distribution, we can have 0, 1, 2, …, 7, or 8 successes. In cell B7 enter **0**. In cell B8 enter **=1+B7**. Copy and paste cell B8 into cells B9 – B15.

30. In cell C6 enter **p(x)**. Here p(x) represents the pdf of the distribution. That is, p(x) = Pr(X = x). To calculate hypergeometric probabilities, we will use Excel's **HYPGEOMDIST** function. In cell C7 enter **=HYPGEOMDIST(B7,C5,C4,C3)**.

    Excel's **HYPGEOMDIST** function requires four pieces of information:
    1. The number of successes, x, which for us is listed in column B.
    2. The number of draws, n, which for us is listed in cell C5 which is why we must put dollar signs before the C and before the 5.
    3. The number of successes in the bin, r, which for us is listed in cell C4 which is why we must put dollar signs before the C and before the 4.
    4. The total number of elements in the bin, N, which for us is listed in cell C3 which is why we must put dollar signs before the C and before the 3.
    5. Copy and paste cell C7 into cells C8 through C15. You should have Pr(X = 8) = 0.003930 (cell C15).

    Notice that unlike the previous distribution function we used, Excel does not give the user the option of pdf vs. cdf when using **HYPGEOMDIST**. Rather, Excel will always return the pdf, and we are left to calculate the cdf ourselves.

31. In cell D6 enter **F(x)**. Here F(x) represents the cdf of the distribution. That is, F(x) = Pr(X ≤ x). To calculate the cumulative probabilities for the hypergeometric, we must enter the formula ourselves. In cell D7 enter **=SUM(C7:C7)**. This will add up all values starting in cell C7 through cell C7, but when it is copied and pasted, it will sum up the values starting in cell C7 and going through some further C cell down the list when copied. Copy and paste cell D7 into cells D8 through D15. To check, you should find that Pr(X ≤ 6) = 0.945773 in cell D13. You should also see that cell D8 = 1.000000 as the greatest possible value for X is 8.

One of the more useful things about creating the pdf and cdf in the way that we have is that we can change the parameters (at least r and n) and all of the probabilities will change automatically. For example, when N = 20, r = 12 and n = 8, the probability of having exactly 6 successes is p(6) = 0.205382 and the probability of having at most 6 successes is F(6) = 0.945773, both of which are reported in cells C13 and D13 respectively. Now change cell C4 to 10, so that half (rather than 60 percent) of the items in the bin are successes. Notice that the values in cells C7 – D15 automatically change. Now X ~ Hyper(20,10,8). In this case, the probability of having exactly 6 successes is p(6) = 0.075018 and the probability of having at most 6 successes is F(6) = 0.990117, which are still reported in cells C13 and D13 respectively.

A word of caution regarding the hypergeometric distribution, however. In class it was mentioned that the hypergeometric distribution is defined under certain parameter restrictions, specifically that $r > n$ and that $N – r > n$.. The easiest way to think of those restrictions is that, with each draw, regardless of what has been drawn previously, it must still be possible to draw a success or a failure. For example, $X \sim$ Hyper(20,14,10) does not fit the requirements because after 7 draws, all of the remaining possible draws could be successes. This consideration isn't usually an issue when working Hypergeometric problems in class or on a test, but it can be an issue when using Excel simply because it is easy to enter parameter numbers that are not consistent with the distribution.

32. We will now use the pdf for our random variable, X ~ Hyper(20,10,8). Make sure you are using r=10 in cell C4
    1. In cell F7 enter **Q1. Pr(X=5)?**
    2. In cell F8 enter **Q2. Pr(X=3 or X=4)?**
    3. In cell F9 enter **Q3. Pr(X>3)?**
    4. In cell F10 enter **Q4. Pr(X<6)?**
    5. In cell F11 enter **Q5. Pr(3<X<8)?**

33. In cells G7 through G11 we want to answer these three questions using cells C7 through D15 so if we changed the probability of success, the answers would change automatically.
    1. In cell G7 enter **=C12**.
    2. In cell G8 enter **=C10+C11**.
    3. For the previous two distributions, we entered **=SUM(C11:C15)** in cell G9. We can enter that here too, but we could encounter a problem. If the number of draws is lower than 8, then Excel will "crash" as there cannot be 8 successes. To avoid this future problem, instead use the complement rule:
       Pr(X>3) = 1 – Pr(X≤3), so in cell G9 enter **=1-SUM(C7:C10)**.
    4. In cell G10 enter **=SUM(C7:C12)**.
    5. In cell G11 enter **=SUM(C11:C14)**.

Part of your worksheet should now look like the following.

Hypergeometric PDF and CDF

N	20
r	10
n	8

x	p(x)	F(x)		
0	0.000357	0.000357	Q1. Pr(X=5)?	24.01%
1	0.009526	0.009883	Q2. Pr(X=3 or X=4)?	59.01%
2	0.075018	0.084901	Q3. Pr(X>3)?	67.50%
3	0.240057	0.324958	Q4. Pr(X<6)?	91.51%
4	0.350083	0.675042	Q5. Pr(3< X < 8)?	67.47%

Now suppose that the number of successes in the bin, r, changes to 12 and that the number of draws, n, changes to 7. Make this change by entering 12 into cell C4 and 7 into cell C5, and the probabilities in the pdf and cdf change automatically, as do the answers to the questions. Following this change, the top part of your worksheet will look like the following.

Hypergeometric PDF and CDF

N	20
r	12
n	7

x	p(x)	F(x)		
0	0.000103	0.000103	Q1. Pr(X=5)?	28.61%
1	0.004334	0.004438	Q2. Pr(X=3 or X=4)?	55.62%
2	0.047678	0.052116	Q3. Pr(X>3)?	74.92%
3	0.198658	0.250774	Q4. Pr(X<6)?	89.44%
4	0.357585	0.608359	Q5. Pr(3< X < 8)?	74.92%

34. Save this file as YourName_Lab3.xlsx. You need to turn this file in, as is, so be careful not to change it.

Exercises

For the following problems, create separate Excel files (or worksheets) that allow you to calculate the indicated probabilities. Report all probabilities in percentages with two decimal places. You do not need to turn in your Excel file for the exercise portion of the lab (though you will turn in the file you created in the tutorial section).

1. Create the pdf and cdf for $X \sim$ Bin(0.25, 240). Do this in such a way that allows you to change the probability, π, at will. Have Excel produce the following probabilities: Pr(X > 130), Pr(X < 118), Pr(X ≤ 107 or X ≥ 132). Fill in the answer sheet for the following different probabilities of success: π = 0.46, 0.50, 0.56.

2. Create the pdf and cdf for $X \sim$ Pois(94). Do this in such a way that allows you to change the expected number of arrivals, λ, at will. Have Excel produce the following probabilities: Pr(X > 115), Pr(X < 128), Pr(94 ≤ X ≤ 122). Fill in the answer sheet for the following different expected number of arrivals: λ = 94, 100, and 115.

3. Create the pdf and cdf for $X \sim$ Hyper(3000, 1800, 280). Do this in such a way that allows you to change the number of draws, n = 280, at will. Have Excel produce the following probabilities: Pr(X < 180), Pr(X > 185), Pr(160 ≤ X ≤ 190). Fill in the answer sheet for following the different number of draws: n = 280, 300, 325.

Turning in your work

Email YourName_Lab3.xlsx to your professor as a file attachment to an email with the subject heading Excel Lab 3: Your Name. Also print and turn in your filled-in answer sheet.

Answer Sheet for Lab #3: Discrete Random Variables

Name: _____

Report all probabilities in percentages with two decimal places.

1. Fill in the following table using the pdf and cdf for X ~ Bin(π, 240).

X ~ Bin(π, 240)			
	π = 0.46	π = 0.50	π = 0.56
P(X > 130)			
Pr(X < 118)			
Pr(X ≤ 107 or X ≥ 132)			

2. Fill in the following table using the pdf and cdf for X ~ Pois(λ).

X ~ Pois(λ)			
	λ = 94	λ = 100	λ = 115
P(X > 115)			
Pr(X < 128)			
Pr(94 ≤ X ≤ 122)			

3. Fill in the following table using the pdf and cdf for X ~ Hyper(3000, 1800, n).

X ~ Hyper(3000, 1800, n)			
	n = 280	n = 300	n = 325
P(X < 180)			
Pr(X > 185)			
Pr(160 ≤ X ≤ 190)			