CHAPTER 4

Visualizing Discrete Distributions

CONCEPTS

- Discrete Variable, Binomial Distribution, Hypergeometric Distribution, Poisson Distribution, Uniform Distribution, Probability Distribution, Cumulative Distribution, Approximations

OBJECTIVES

- Recognize common discrete distributions and their cumulatives
- Identify the parameters of common discrete distributions and how they affect a distribution
- Interpret descriptive statistics for common discrete distributions
- Understand when to apply approximations and learn to assess their accuracy

Overview of Concepts

A **discrete variable** is a random variable that can take on only a countable number of alternative values. In this module, discrete variables assume only integer values. The **probability distribution** tells the chance that each value will occur. The binomial, Poisson, uniform, and hypergeometric distributions all describe the probability distribution of discrete variables under a variety of different situations or experiments.

A Bernoulli process refers to a two-valued or binary experiment (0, 1; no, yes; failure, success) whose probability of a success is fixed throughout and in which one trial has no effect on another trial. The **binomial distribution** tells the statistician the probability of obtaining r successes in an experiment of n trials (sample size n) when the binary variable is generated by a Bernoulli process. Generally, this means that the n observations are being selected from an infinite or very large population. It can, however, also mean that sampling is done with replacement from a smaller population. The binomial distribution can be used to find the probability of winning in many games of chance.

If the population size is finite and sampling is done without replacement, then the probability of a success changes as each sample is drawn. In this case the **hypergeometric distribution** tells the statistician the probability of obtaining r successes in an experiment of n trials (or sample size n) when sampling without replacement from a population of size N containing A successes. The hypergeometric distribution can be used to find the probability of an event in many card games.

When the statistician is interested in finding the probability of the number of occurrences (successes) over a fixed time or space, the **Poisson distribution** is used. It is named after the French mathematician Siméon Poisson (1781–1840). This distribution requires that the probability of the event (success) be proportional to the size of the interval being measured, that two events have zero probability of occurring at the same point in time or space, and that each event be independent of every other event.

The **uniform distribution** describes the situation where there are a finite number of events that each occur with equal probability. If there are k events in a uniform experiment, each event has a probability of 1/k of occurring.

Associated with every probability distribution is a **cumulative distribution**. This distribution tells the statistician the probability of selecting a value less than or equal to a specific value of the random variable. It sums the probability distribution from left to right or, in other words, accumulates the left tail of the probability distribution. Sometimes statisticians also like to accumulate the right tail of the distribution, which tells the probability of selecting a value greater than or equal to a specific value of the random variable.

The binomial, hypergeometric and Poisson distributions can each be **approximated** by other distributions. These approximations are useful because they often simplify the process of calculating probabilities. The accuracy of every approximation can be assessed using certain criteria. If the criteria are not met, the accuracy of the approximation must be questioned. The criteria are rough guidelines only. At times the criteria can suggest that an approximation is appropriate when it displays more error than the statistician deems appropriate.

Illustration of Concepts

Discrete distributions have useful applications in the everyday world. Consider the following example. Assume you operate a large produce distribution center in a major urban area. Every day numerous large trucks arrive with fresh fruits and vegetables, which are purchased from you by local retailers. Your trucks then deliver the fresh produce to retailers throughout the urban area for resale to consumers. Let us consider two problems that you would face.

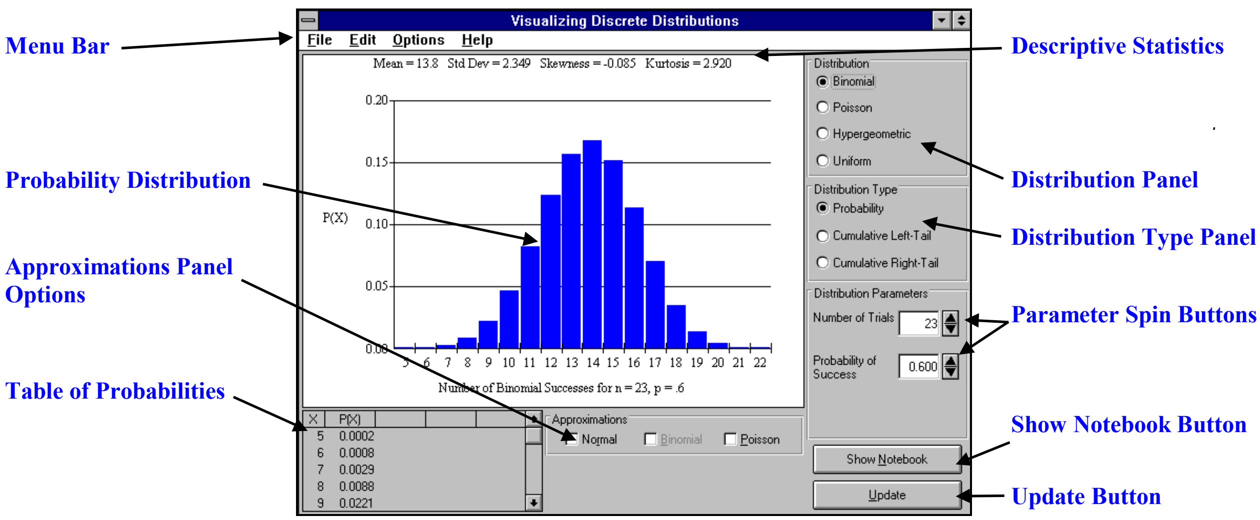
First, in order to operate you need a loading dock to accommodate both the large trucks delivering the produce and your own delivery trucks. If you have too small a dock, trucks will have to wait to load or unload. If they are your own delivery trucks, you are not efficiently using your trucks or drivers. If they are growers' trucks delivering produce to you, you will alienate your suppliers who must pay their drivers to wait at the dock. However, building too large a dock is a waste of your own resources (both land and money to build the dock). In this scenario the number of trucks arriving in one hour is a **discrete variable** having a **Poisson distribution**. This can be used to help decide the size of loading dock that is needed. For example, if you know that a truck can be loaded or unloaded in one hour and that on average 6 trucks arrive every hour, a loading dock that holds 10 trucks would be insufficient 8.39% of the time (obtained from the **cumulative distribution**) while one that holds 11 trucks would be insufficient only 4.26% of the time (we have assumed that the trucks arrive independently and do not prefer certain hours of the day). Therefore, if you want less than a 5% risk of having a truck wait for docking space, an 11-berth dock would be appropriate for current needs.

Second, in order to operate a firm that delivers quality produce you must decide when to reject (or receive at a discount) a shipment from a grower as being of inferior quality. This means that you will need to sample produce from each delivery to determine if it meets your quality standard. For example, assume you have a contract with a grower that states if more than 5% of a shipment is damaged, you get a 10% reduction in price and if more than 10% of a shipment is damaged you get a 25% reduction in price. Experience tells you that it is equally likely that produce boxes throughout the truck contain damaged produce and that all produce within a box has an equal chance of being damaged. Because each produce box is equally likely to hold damaged produce, the boxes follow a **uniform distribution**. Knowing this **probability distribution** means that using random sampling will provide you with a representative sample of produce boxes. If 20 boxes are selected and each produce box holds 200 pieces of produce, you would have 4,000 pieces of produce to examine. If you sampled 200 pieces, the distribution of selecting damaged produce would follow a **hypergeometric distribution**. However, because 200/4,000 ≤ 0.05 (a criterion for using the binomial distribution to approximate the hypergeometric distribution) we can **approximate** this distribution with a **binomial distribution**. For example, assume 22 damaged pieces of produce were discovered out of the 200 samples. If only 5% of the produce is damaged, the probability of selecting 22 or more damaged pieces would be only be 0.0005 (very unlikely). If 10% of the produce is damaged, the probability of selecting 22 or more damaged pieces would be 0.3517 (a likely event). Therefore, in this example the distributor should get a 10% price reduction because more than 5% of the produce is damaged, but there is not enough evidence to suggest that more than 10% of the produce is damaged.

Orientation to Basic Features

This module illustrates the binomial, Poisson, hypergeometric and uniform discrete probability distributions and their cumulative distributions. Appropriate approximations can also be displayed.

1. **Opening Screen**
   Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. Read the introduction to see the types of questions this module will enable you to answer. Click on the Concepts tabs to see the concepts that this module covers. Click on the Scenarios tab. Select the Binomial Distribution from the table of choices. Select a scenario and press OK. The upper left of the screen depicts the binomial distribution. The module's Control Panel is on the right. On the bottom left is the table of the probabilities that are graphed. On the bottom right below the distribution are approximation options you may select. Other features are controlled from the menu bar at the top of the screen. A flashing Update button will indicate when you have changed one or more control settings.



2. **Changing Parameter Values**
   Use the spin buttons to change the parameters of the binomial distribution or to type in parameter values. Click on Update to view the new distribution.

3. **Type of Display**
   Select either Cumulative Right-Tail or Cumulative Left-Tail on the Distribution Type panel. Click on Update. Notice that both the probability and cumulative probability values are displayed in the table below the graph.

4. **Selecting a Distribution**
   Return to the Notebook by clicking on the Show Notebook button. Click on Return to the Scenarios contents page. Select either Poisson distribution, Hypergeometric distribution, or Uniform distribution. Select a scenario from the choices given. If you do not wish to deal with scenarios, you can simply select Poisson, Hypergeometric, or Uniform from the Distribution panel of the Control Panel and then click on Update.

5. **Options**

Three options are available from the Options menu on the menu bar at the top of the screen.

a. You can choose to not display the descriptive statistics for the distribution. To remove the display, deselect Display Statistics on the Options menu.

b. You can also display your discrete distributions using either 2D or 3D histograms. Select 3D Graph on the Options menu in order to get 3D histograms. Select the kind of graph you prefer. Since 2D histograms more accurately display probabilities, they are the default.

c. Select Auto Update on the Options menu to automatically update the screen as the parameter values are changed.

6. **Copying Graph**

Select Copy Graph or Copy Table from the Edit menu (on the menu bar at the top of the screen) to copy the graph or table onto the clipboard. It can then be pasted into other applications, such as Microsoft Word or WordPerfect.

7. **Help**

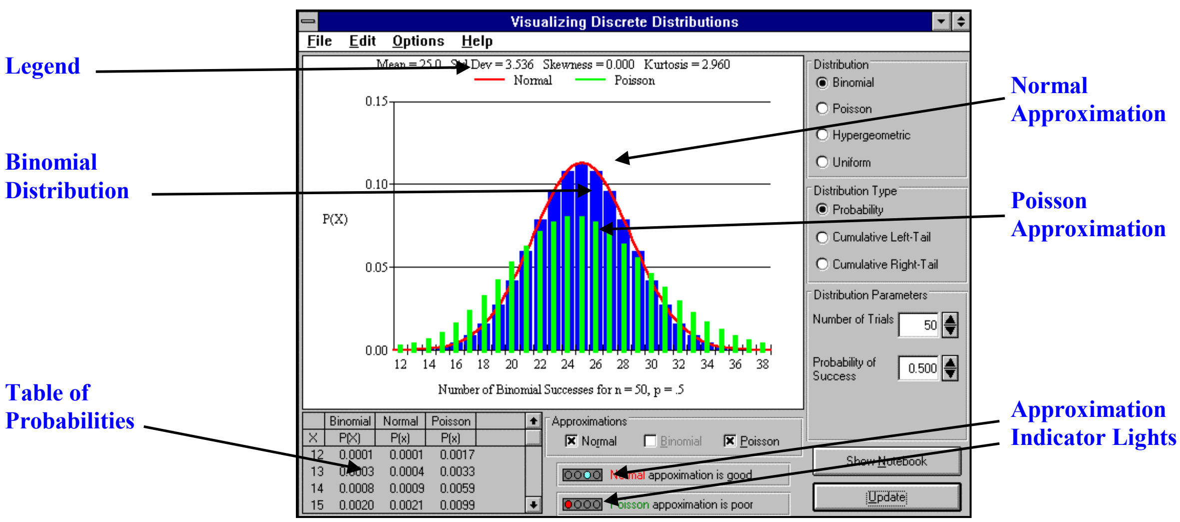
Click on Help on the menu bar at the top of the screen. Search for Help lets you search an index for this module, Contents shows a table of contents for this module, Using Help gives instructions on how to use Help, and About gives licensing and copyright information about this module. Click on Contents. Click on any topic of interest (shown as green "hot text"). If there is green "hot text" on any screen, click on it to jump to a related help screen. Close Help by selecting Exit from the File menu on the Help screen.

8. **Exit**

Close the module by selecting Exit in the File menu (or click ☒ in the upper right-hand corner of the window). You will be returned to the *Visual Statistics* main menu.

Orientation to Additional Features

1. **Approximations**
   Depending on the distribution you have selected either one or two standard approximations to the distribution may be superimposed on the display. For the binomial, both the normal and Poisson approximations are available; for the Poisson, the normal approximation is available; for the hypergeometric, both the normal and binomial approximations are available. No approximations are available for the uniform distribution. These options are shown in the Approximations panel below the distribution (if an option is not available, the approximation will be grayed out). The quality of the approximation is shown in the Approximation Indicator light at the bottom of the screen (red = poor, yellow = adequate, light blue = good, green = excellent). Notice that the table below the graph contains the probabilities using the approximations that have been selected. The figure below illustrates the display for a binomial distribution with n = 50 and p = 0.50.



Basic Learning Exercises

Name ________________

Binomial Distribution

1. Press the Show Notebook button. Select the Scenarios tab and click on Binomial Distribution. Select the Green Lights scenario and read it. Press the OK button. In this example, what are the values of its descriptive statistics? What parameters are associated with the binomial distribution and what are their values? How would you describe this distribution?

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

2. Change the Probability of Success from 0.60 to 0.50. What are the new values of the descriptive statistics? How would you describe the new distribution? What would this tell you about how the traffic light is timed?

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

3. Change the Probability of Success from 0.50 to 0.40. What are the new values of the descriptive statistics? How would you describe the new distribution? What would this mean about how the traffic light is timed?

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

4. Change the Number of Trials to 70 from 23. a) What are the new values of the descriptive statistics? b) What happened to the shape of the distribution? b) Change the Probability of Success to 0.70. c) What are the new values of the descriptive statistics? d) What do you think the descriptive statistics would be if the Probability of Success were changed to 0.30? **Hint:** Reread exercises 1 and 3 together with your answers. Check your answer.

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______
   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______
   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

Poisson Distribution

5. Press the Show Notebook button. Select the Scenarios tab and click on Poisson Distribution. Select and read the Billboards on I-75 scenario. Record the descriptive statistics for this distribution. What parameter defines the Poisson distribution?

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

6. a) How would you describe the distribution? b) What is the meaning of λ in this scenario? c) What is the relationship between Lambda (λ) and the distribution's mean and variance?

7. If we changed the scenario to be the average number of billboards per 10 miles, what would the new λ be? What are the new descriptive statistics? What is the meaning of λ in this modified scenario? **Hint:** λ is proportional to the size of your interval.

   λ = _______
   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

Uniform Distribution

8. Press the Show Notebook button, select the Scenarios tab, click on Uniform Distribution, select the Gas Pump scenario and read it. What are the values of the descriptive statistics? What parameters are associated with the uniform distribution and what are their values? How would you describe the uniform distribution?

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

Hypergeometric Distribution

9. Deselect all approximations. Press the Show Notebook button. Select the Scenarios tab and click on Hypergeometric Distribution. Select and read the Refrigerator Configuration scenario. a) What are the values of its descriptive statistics? b) What parameters are associated with the hypergeometric distribution? c) How would you describe the distribution? d) What is the probability of a success (selecting a side-by-side) before the first refrigerator is selected?

   Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _______

10. Increase Sample Size to 6. What are the new values of the descriptive statistics? Change Sample Size to 4. Change Successes in Population to 6. What are the new values of the descriptive statistics? Notice that in both cases the distribution is symmetric. **Hint:** *A hypergeometric distribution is symmetric only if the sample size or the number of successes in the population are half the population size.*

    Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _________
    Mean ________ Standard Deviation ________ Skewness _________ Kurtosis _________

Intermediate Learning Exercises Name ________________

Normal Approximation to the Binomial Distribution

11. Press the Show Notebook button, select the Scenarios tab, click on Binomial Distribution, select the Green Lights scenario and read it. Press the OK button. In the Approximations panel select Normal. Set the Probability of Success (p) to 0.5 and the Number of Trials (n) to 5 and press Update. a) What is the mean and variance of the approximating normal distribution? b) *In an "excellent" approximation, the normal curve goes through the middle of the top of each bar in the binomial distribution.* Does the normal curve go through the middle of each bar in this distribution? c) What does the Approximation Indicator light (at the bottom of the screen) say about the quality of the normal approximation? d) Increase n to 20 and answer questions (a)–(c).

12. Set n to 20 and p to 0.10. a) Does the normal curve intersect the center of each bar? b) Using the vertical distance between the middle of each bar and the normal distribution as a measure of the quality of the approximation, which bar has the largest absolute difference? c) From the table of probabilities in the lower left of the screen, record each binomial and normal probability, and find the absolute differences. Which absolute difference is largest? Does your answer agree with b)? d) What does the Approximation Indicator light say?

X	Binomial	Normal	Absolute Difference
0	_____	_____	_____
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
4	_____	_____	_____
5	_____	_____	_____
6	_____	_____	_____
7	_____	_____	_____
8	_____	_____	_____
9	_____	_____	_____

13. Select Auto Update under Options on the Menu Bar. Keep p = 0.10. Increase the Number of Trials spin button, at what value of n does the Approximation Indicator light change to "adequate?" To "good?" To "excellent?" Does the graph reflect these assessments?

14. The Approximation Indicator light is based on the maximum absolute difference between the two distributions. Less than 0.005 is *excellent*, less than 0.01 is *good*, less than 0.02 is *adequate*, and all others are *poor*. However, most textbooks state that if np ≥ 5 and n (1-p) ≥ 5 the normal distribution adequately approximates the binomial distribution. This is true when p is between 0.20 and 0.80, but the approximation deteriorates when p gets either too large or too small. Here are several situations for which np ≥ 5 and n (1-p) ≥ 5. What does the Approximation Indicator light say about the accuracy of each normal approximation? If some are less than good, does the graph show the problem? Explain briefly.

n	p	np	Quality Indicator
10	0.50	5	_________
15	0.35	5.25	_________
50	0.10	5	_________
100	0.05	5	_________
125	0.04	5	_________
167	0.03	5.01	_________

Poisson Approximation to Binomial Distribution

15. Most textbooks state that if n ≥ 20 and p ≤ 0.05 the Poisson distribution adequately approximates the binomial distribution. To evaluate this statement in the Approximations panel, deselect Normal and select Poisson. *For an "excellent" approximation the green and blue bar should be the same height.* Set p = 0.05 and n = 10. a) What is the value of λ in the approximating Poisson distribution? b) At what value of n does the approximation become "good", based on the Approximation Indicator light? c) Become "excellent"? d) Set p = 0.04. At what value of n does the approximation become "adequate" or "poor"? e) What does this tell you about when to use the Poisson approximation to the binomial distribution?

Normal Approximation to Poisson Distribution

16. Most textbooks state that if λ > 5 the normal distribution adequately approximates the Poisson distribution. Press the Show Notebook button, select the Scenarios tab, click on Poisson Distribution, select and read the Billboards on I-75 scenario. Press OK. Change λ to 5.1. To see the normal distribution select Normal in the Approximations panel. a) What is the mean and standard deviation of the approximating normal distribution? b)Using the Approximation Indicator light, how good is this approximation? c) At what value of λ does the approximation become "adequate?" d) Given this result, why is λ > 5 used? e) At what value of λ does the approximation become "good?"

Advanced Learning Exercises Name ________________

Investigating the Hypergeometric Distribution

17. Press the Show Notebook button. Select the Scenarios tab and click on Hypergeometric Distribution. Select and read the Refrigerator Configuration scenario. Deselect all approximations. Set Population Size to 50 and Successes in Population to 35. Select Options on the menu bar and deselect Auto Update. a) Set Sample Size to 5 and click Update. Change Sample Size to 45 and click Update. How did the shape of the distribution change? b) Try other pairs of sample sizes equal to n and N-n. What general rule can you deduce if the number of successes in the population does not change? c) Set Population Size to 50 and Sample Size to 10. Change Successes in Population to 45 and click Update. Change Successes in Population to 5 and click Update. How did the shape of the distribution change? d) Try other pairs of successes equal to A and N-A. What general rule can you deduce if the sample size is unchanged?

18. Set Population Size to 50, Sample Size to 40, and Successes in Population to 25. a) Why is this distribution symmetric? b) How does the hypergeometric distribution change if Sample Size is changed to 10? Explain. c) How does the hypergeometric distribution change if Population Size is set to 30 and Successes in Population to 15? Explain.

19. Set Population Size to 50, Sample Size to 25, and Successes in Population to 40. a) Why is this distribution symmetric? b) How does the hypergeometric distribution change if Successes in Population is changed to 10? Explain. c) How does the hypergeometric distribution change if Population Size is set to 30 and Sample Size to 15? Explain.

20. Set Population Size to 50, Successes in Population to 25, and Sample Size to 10. Select Auto Update. Use the spin button to reduce Successes in Population to one. What happens to the number of intervals? Why does this happen?

21. a) What will happen to the distribution as you increase Successes in Population from 1 to 49? Try it. Was your supposition correct? If not describe what did happen. b) Explain why this occurs as the number of successes changes.

22. Many learners are surprised by the results of exercises 20 and 21. What is unique about the hypergeometric distribution that causes these unexpected results?

Binomial Approximation to the Hypergeometric Distribution

23. Select Binomial in the Approximations panel. Textbooks usually say the binomial approximation to the hypergeometric distribution is acceptable if $n/N < 0.05$. What is the value of n and p in the approximating binomial distribution? Evaluate the $n/N < 0.05$ criterion with different values of Population Size, Sample Size and Successes in Population.

Normal Approximation to the Hypergeometric Distribution

24. Select Normal in the Approximations panel. Set Population Size to 100, Successes in Population to 50, and Sample Size to 50. a) What is the value of the mean and variance in the approximating normal distribution? b) Decrease sample size, observing the approximation quality indicator. Why does the approximation deteriorate? c) Set Sample Size to 25. Decrease Successes in Population. Why does the approximation deteriorate?

Uniform Distribution

25. Press the Show Notebook Button. Select the Scenarios tab and click on Uniform Distribution. Select Gas Pump, read it, and click OK. Why is there no option to show an approximation?

Individual Learning Projects

Write a report on one of the three topics listed below. Use the cut-and-paste facilities of the module to place the appropriate graphs in your report.

1. Select two distributions to approximate. For both distributions investigate the accuracy of approximating the probability distribution function and the cumulative distribution function. First, you must select a criterion to evaluate the accuracy of each approximation. You can use a visual criterion (e.g., the vertical distance between the approximating distribution and the original distribution) or a criterion using a statistic (e.g., how much the skewness or peakedness statistics differ from one another). Second, for each distribution you are approximating, evaluate your approximation for a variety of parameter values. Third, decide whether the cumulative or probability distribution is easier to approximate and explain why.

2. Investigate the skewness of the binomial, Poisson, hypergeometric, and uniform distribution. For each distribution, explain when (or if) the distribution is symmetric, under what conditions the distribution is positively and negatively skewed, and how each of the distribution's parameters affects skewness. **Hint:** The hypergeometric distribution is complicated because there are three parameters, all of which can affect your result.

3. The binomial distribution can be used to approximate the hypergeometric distribution. The Poisson distribution can be used to approximate the binomial distribution. Using the Approximation Light indicator, investigate under what conditions the Poisson distribution can be used to approximate the hypergeometric distribution. Develop a rule for when the Poisson distribution can be used to approximate the hypergeometric distribution.

Team Learning Projects

Select one of the three projects listed below. Produce a team project that is suitable for an oral presentation. Use presentation software or large poster board(s) to display your results. Graphs and tables should be large enough for your audience to see. Each team member should be responsible for producing some of the exhibits. Ask your instructor if a written report is also expected.

1. This project is for a team of two. The team is to investigate the normal and Poisson approximation to the binomial distribution. One team member should investigate the normal approximation and the other the Poisson approximation. Start with a value of p = 0.5 and select n such that the *adequate* criterion is just met. Systematically decrease p and adjust n so that the criterion is just met. Each time evaluate the approximation. When is the approximation no longer *adequate* in your opinion? Repeat this process for the *good* and *excellent* criteria. Create a display that illustrates the accuracy of the approximation. Your displays should clearly illustrate how the approximation changes as the parameter values change.
2. This project is for a team of three or four. The team is to investigate the shape of the binomial distribution. Each team member should select a different value of n (make sure the values selected cover the range from 10 to 100). Investigate p = 0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50. Each team member should create the binomial distribution using each of the values of p. Create a display that shows how the parameter values affect the shape of the distribution. What happens to the distribution's shape as n increases with constant p? Does the value of p affect this answer? What happens to the distribution's shape as p increases with constant n? Does the value of n affect this answer? Why is it unnecessary to investigate values of p above 0.50?
3. This project is for a team of three. The team is to investigate the normal approximation to the hypergeometric distribution. The team should select a population size N (fairly large). Each team member should select one of the three approximation levels (*adequate*, *good*, *excellent*) to investigate by selecting A so that A/N = 0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50 (or as nearly as possible for the chosen N). Why is it unnecessary to investigate values of A/N above 0.50? Each team member should create a hypergeometric distribution (with the normal approximation superimposed) using each of the values of A/N. For each of these distributions, each team member should select a value of n that is consistent with the approximation level that he or she is investigating (or as closely as possible). Each team member should discuss when the approximation began to deteriorate. The team should arrive at a general conclusion.

Self-Evaluation Quiz

1. Which distribution is always asymmetric?
   a. Binomial.
   b. Poisson.
   c. Hypergeometric.
   d. Uniform.
   e. None of the distributions are always asymmetric.

2. Which distribution is most strongly right-skewed?
   a. Binomial with n = 50, p = 0.25
   b. Binomial with n = 5, p = 0.9
   c. Binomial with n = 50, p = 0.5
   d. Poisson with $\lambda = 1$
   e. Poisson with $\lambda = 15$

3. Which distribution is left-skewed?
   a. Binomial with n = 10, p = 0.3
   b. Poisson with $\lambda = 2.7$
   c. Hypergeometric with N = 10, n = 5, A = 5
   d. Uniform with a = 10, b = 20
   e. None of the above.

4. The Poisson distribution
   a. was named after French mathematician Siméon Poisson.
   b. was first applied in the French fishing industry.
   c. is always less peaked (flatter) than a normal distribution.
   d. is always less skewed than a normal distribution.
   e. has none of the above characteristics.

5. Which *most* resembles a Poisson random variable?
   a. the number of heads in 200 flips of a fair coin.
   b. the number of power failures in one year at a computer center.
   c. the number of face cards in a bridge hand.
   d. the number of bad floppy diskettes that are on a desk full of floppy diskettes.
   e. the number of dots showing when a pair of dice are rolled.

6. Which distribution has a mean of 5?
   a. Poisson with $\lambda = 25$.
   b. Binomial with n = 200, p = .05
   c. Uniform with endpoints 2 and 8
   d. Hypergeometric with N = 100, n = 10, A = 5
   e. More than one of the above.

7. Which binomial distribution would a normal distribution best approximate?
   a. n = 200, p = 0.05
   b. n = 100, p = 0.10
   c. n = 50, p = 0.20
   d. n = 25, p = 0.40
   e. All of the above.

8. Which Poisson distribution would a normal distribution best approximate?
   a. $\lambda = 25$
   b. $\lambda = 10$
   c. $\lambda = 3$
   d. $\lambda = 1.5$
   e. More than one of the above.

9. Which binomial distribution would a Poisson distribution best approximate?
   a. n = 35, p = 0.025
   b. n = 50, p = 0.03
   c. n = 20, p = 0.01
   d. n = 200, p = 0.05
   e. None of the above.

10. A binomial distribution with n = 75 and p = .02 is best approximated by which distribution?
    a. Poisson with $\lambda = 25$
    b. Poisson with $\lambda = 7.5$
    c. Poisson with $\lambda = 1.5$
    d. Poisson with $\lambda = 15$
    e. None of the above.

11. In a hypergeometric distribution with N = 20, n = 5, A = 4 the largest possible value of X (the number of successes) is
    a. 20
    b. 10
    c. 5
    d. 4
    e. none of the above.

12. Which hypergeometric distribution would a normal distribution best approximate?
    a. N = 100, n = 10, A = 5
    b. N = 150, n = 5, A = 75
    c. N = 20, n = 10, A = 5
    d. N = 200, n =25, A = 100
    e. N = 50, n = 20, A = 10

Glossary of Terms

Approximations In this module, a reference to using one distribution to approximate another distribution.

Binomial distribution The discrete distribution that describes the number of "successes" in n independent trials with constant probability of success p.

Cumulative distribution A function that maps each value of a random variable to the probability of being less than or equal to that value. The function begins at 0 and rises to 1 as you move to the right (or, less commonly from 1 to 0 as you move to the left). See **Probability distribution**.

Descriptive statistics Statistics provided in this module are the distribution's mean, standard deviation, skewness, and kurtosis.

Discrete distribution A probability distribution whose random variable is defined over a discrete domain.

Discrete variable A random variable that takes on a countable number of different values.

Hypergeometric distribution The discrete distribution that describes the number of "successes" in a sample of size n from a finite population of size N containing a fixed number of "successes" A when sampling without replacement.

Kurtosis A measure of relative peakedness of a distribution. K = 3 indicates a normal bell-shaped distribution (mesokurtic). K < 3 indicates a platykurtic distribution (flatter than a normal distribution with shorter tails). K > 3 indicates a leptokurtic distribution (more peaked than a normal distribution with longer tails).

Mean The expected value of a random variable. It may be interpreted as the fulcrum (balancing point) of the distribution along the X-axis.

Median The point along the X-axis that defines the upper and lower 50 percent of the distribution. In a symmetric distribution, it is equal to the mean.

Mode The X value that defines the highest point of the probability function.

Normal distribution The standard bell-shaped or Gaussian distribution. It has two parameters called the mean and variance.

Parameter A numerical characteristic of a population. Each theoretical distribution is characterized by one or more parameters.

Poisson distribution The discrete distribution that describes the number of independent events occurring within an interval of time or space when the expected number of events within that interval is known.

Probability distribution A function that maps each value of a random variable to a probability. Each probability must be less than 1 and they must sum to 1. See **Cumulative distribution**.

Skewness A measure of relative symmetry. Zero indicates symmetry. The larger its absolute value, the more asymmetric the distribution. Positive values indicate a long right tail, and negative values indicate a long left tail.

Standard deviation The square root of the variance.

Uniform distribution The discrete distribution that assigns the same probability to each of n values of X in the domain. It is platykurtic (kurtosis = 1.8) and symmetric (skewness = 0).

Variance A measure of dispersion equal to the expected value of $(X - \mu)^2$. The larger the variance, the greater the dispersion or "spread" around the mean. See **Standard deviation**.

Solutions to Self-Evaluation Quiz

1. b Do Exercises 1–10.
2. d Do Exercises 1–7.
3. e Do Exercises 1–10.
4. a Read the Overview of Concepts.
5. b Do Exercises 5–7. Read the Overview of Concepts.
6. c Do Exercises 1–10.
7. d Do Exercises 11–14.
8. a Do Exercise 16.
9. c Do Exercise 15.
10. c Do Exercise 15.
11. d Do Exercises 17–21.
12. d Do Exercises 17, 24.