CHAPTER 2

Visualizing Random Processes

CONCEPTS

- Stochastic Process, Event, Sample Space, Random Variable, Variation, Probability of an Event, Empirical Probability, Probability Distribution, Centrality, Dispersion, Skewness, Parameter, Histogram

OBJECTIVES

- Recognize that the outcomes of a stochastic process may exhibit regularity even though the process is random

- Learn through experimentation how changing the parameters can affect the outcomes of an experiment

- Learn how a histogram can summarize the results of an experiment and suggest the shape of a probability distribution

- Visualize random data-generating processes that give rise to common probability distributions

- Understand how relative frequencies can be used to estimate the probability of an event if the sample is large

Overview of Concepts

A **stochastic process** is an experiment of chance whose outcomes cannot be predicted (e.g., flipping a coin). The **sample space** is the set of all possible outcomes (called elementary outcomes). For example, if you flip three coins the sample space {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT} contains eight elementary outcomes. An **event** is a collection of elementary outcomes. For example, the set {HHT, HTH, THH} consists of all elementary outcomes containing two heads. A **random variable** is a number that is assigned to each event in the domain of X. For example, if X is the number of heads in the three-coin experiment, its possible values are {0, 1, 2, 3}.

A **probability distribution** assigns a probability P(X) to every value of X, such that $0 \leq P(X) \leq 1$. If the event cannot occur then P(X) = 0, and if the event is certain to occur then P(X) = 1. The **probability of an event** can sometimes be known *a priori* by thinking about the process. For example, in the three-coin experiment P(2) is 3/8 because three of the eight events in the sample space have two heads. An **empirical probability** is found by doing the experiment many times and finding the relative frequency of occurrence of each event. If we can flip three coins repeatedly, the relative frequency of 2's will be approximately 0.375 or 3/8. An empirical probability becomes better as the number of repetitions grows. Yet in a real-life stochastic experiment there is **variation** due to sampling, so unexpected things can happen in a given sample (e.g., you can get 0 heads in 3 flips).

We can describe a probability distribution in terms of its **centrality** ("middle" or "typical" value), its **dispersion** (the "spread" around the center), and its **skewness** (degree to which it lacks symmetry). The mode (most frequent value) is a simple measure of centrality. The range (from lowest to highest value) is a simple measure of dispersion. The visual length of the right and left tails are a simple measure of skewness. A **histogram** shows the sample frequencies.

Sometimes we can derive a mathematical equation for a probability if we assume certain conditions. For example, a binary experiment with independent trials and constant probability of "success" follows a known model called a *binomial distribution*. Its two **parameters** are the number of coins and the probability of "heads." In many binary processes the probability of "success" is not 0.50. For example, if motorists are stopped at random to check for expired licenses, the probability that a driver will have an expired license (the probability of "success") might be 0.03. An unknown parameter can be estimated empirically if a large sample is taken (e.g, using police records).

A second type of stochastic processe involves counting events over an interval of time or space. For example, how many phone calls will arrive at an L. L. Bean call center between 10:00 A.M. and 10:01 A.M. on the first Thursday in October? The lowest possible value is 0 but there is no obvious upper limit. If phone calls are independent and randomly distributed over time, such a probability distribution follows a known mathematical model called the *Poisson distribution* with one parameter (the mean arrival rate). This parameter is generally unknown *a priori* but can be estimated empirically (e.g., from L. L. Bean's database).

A third type of stochastic process is found in games of chance where all probabilities are the same (e.g., rolling a die), giving rise to a *uniform distribution*. Combining events gives rise to a more complex experiment. For example, if we roll two dice and add their outcomes we get a *triangular distribution*. These are examples of known distributions.

Illustration of Concepts

In a certain assembly facility, cars are produced with an average of 5 defects per car. The number of defects in a particular car is a **random variable**. Any particular **event** (such as finding 4 defects in a particular car) is the result of a **stochastic process**. The **histogram** in Figure 1 shows a count of the number of defects in each of 50 cars inspected at random. The range is from 1 to 10. It is symmetric and has two modes (4 and 6). Some events (such as 0 and 8) did not occur. There is **variation** around the presumed process mean of 5.
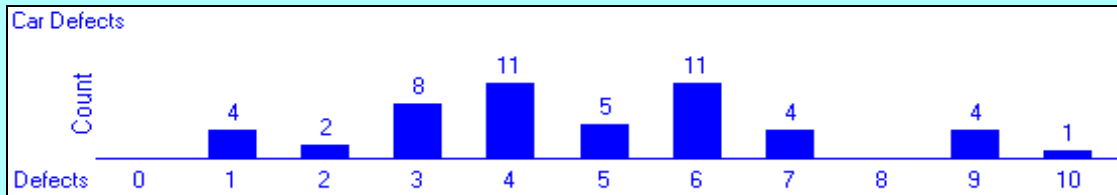

Figure 1: Histogram of Number of Defects in 1st Sample of 50 Cars

Figure 2 shows the results of inspecting 50 more cars. This sample histogram is quite different. Its range is from 2 to 10. It is unimodal (4) and has no "gaps" but is skewed (longer right tail). Logically the **sample space** {0, 1, 2, 3, ...} is best regarded as having no upper limit, but these two samples suggest that more than 10 defects is unlikely. Overall, these two samples reveal considerable **dispersion**, not very much **centrality**, and some skewness.
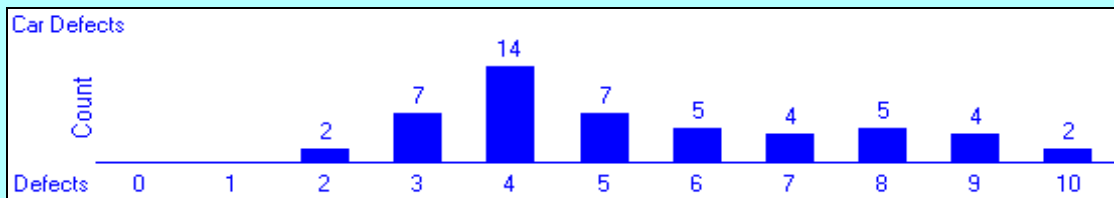

Figure 2: Histogram of Number of Defects in 2nd Sample of 50 Cars

To estimate the **probability of an event** empirically, we would prefer many repetitions of the experiment. Figure 3 shows the results of 1000 car inspections. The **empirical probability** of 4 defects is estimated to be 176/1000 or 0.176. For such a large sample, this histogram should provide a pretty good idea of the underlying **probability distribution**. Its range is from 0 to 14 and the mean number of defects is 4.942, which is very close to 5 (the true **parameter** of this Poisson distribution).
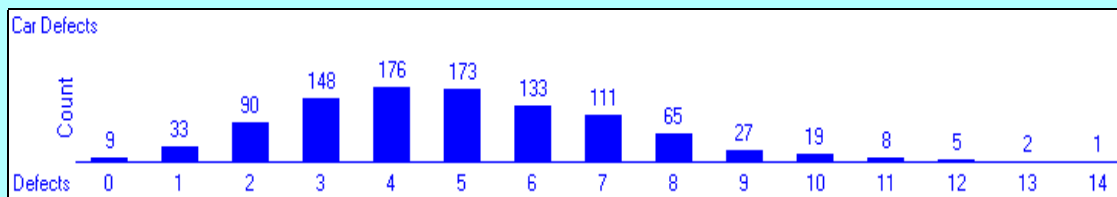

Figure 3: Histogram of Number of Defects in a Sample of 1000 Cars
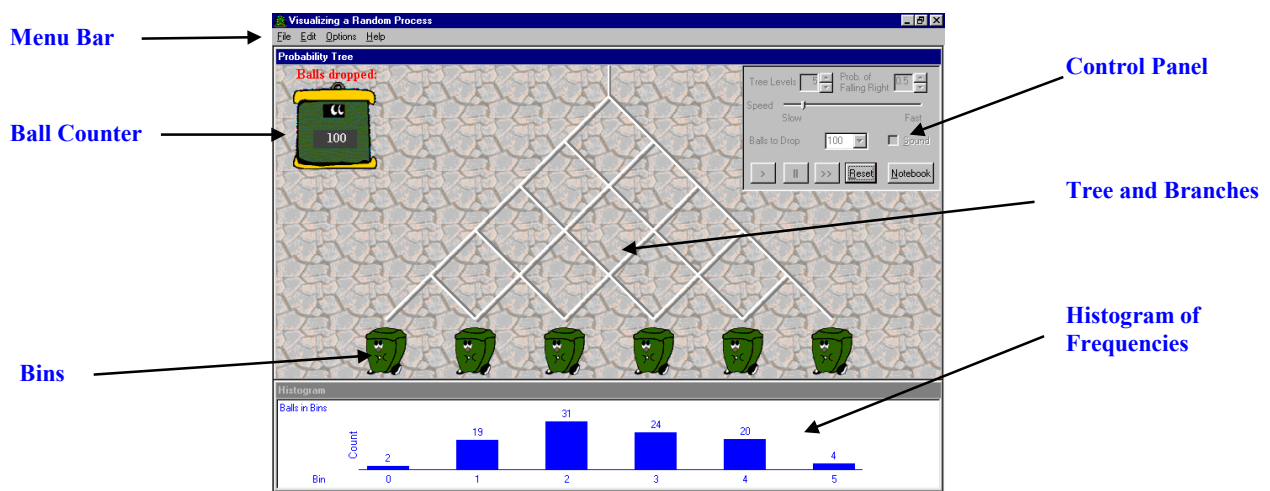
As the number of repetitions in the experiment increases, we may also begin to see the shape of the unknown probability distribution. The histogram in Figure 3 (1000 cars) is somewhat more regular in its shape than the histograms in Figures 1 and 2 (50 cars). We would expect that another sample of 1000 would produce a histogram much like Figure 3. In smaller samples we would be reluctant to generalize about the distribution's shape.

Orientation to Features

This module is a simulation that illustrates random processes. You can choose four different scenarios from the Notebook.
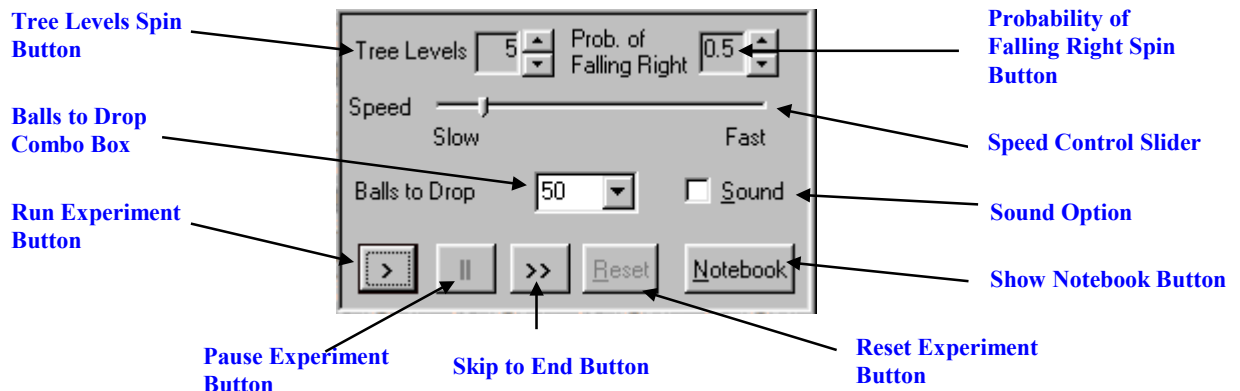
1. **Opening Screen**
    Start the module by clicking on the module's icon number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. The Introduction and Concepts sections describe what will be covered in this module. Click on the Scenarios tab, select Probability Trees, and press OK. The upper part of the screen shows a tree with branch points, bins, a counter, and a Control Panel in the upper right. On the bottom is a histogram that tabulates the frequency of balls that are dropped.
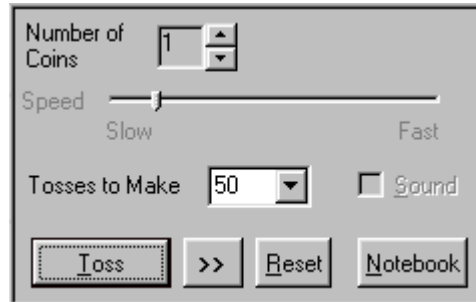

2. **Control Panel**
    a. In the Probability Trees scenario, at each branch point, a ball falls either left or right. The Tree Levels spin button sets the number of branch points (1 to 10). The Probability of Falling Right spin button goes from 0.1 to 0.9 (initially 0.5). The Speed slider controls the rate at which balls are dropped. The Balls to Drop combo box allows 10, 20, 50, 100, 200, 500, or 1000 balls. The Sound check box adds sound effects. The Reset Experiment button resets the counter to zero. The Run Experiment button initiates ball-dropping and the Pause Experiment button lets you pause to look at the histogram. If you get tired of watching each ball, the Skip to End of Experiment button finishes the simulation.

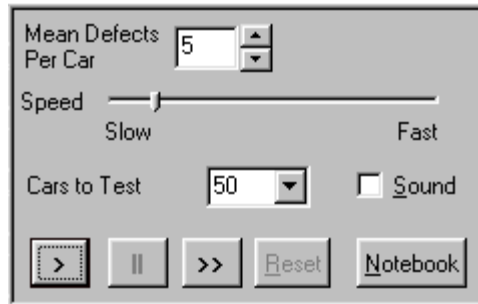b. Click the Notebook button. Select Tossing Coins, then click OK. In this scenario, several coins are flipped and the random variable is the total number of heads. Use the Number of Coins spin button to set the number of coins (1 to 8). These are fair coins, so the probabilities of heads and tails are the same at 0.5. The Speed slider is disabled. The Sound check box adds sound effects. With the Tosses to Make combo box you can choose 10, 20, 50, 100, 200, 500, or 1000 tosses. The Reset Experiment button resets the counter to zero. The Toss button flips the coins once. If you want to see the end result, click the Skip to End button to finish the random simulation.

c. Click the Notebook button. Select Throwing Dice, then click OK. In this scenario, you can toss several dice, then see the sum of the dots that you get. Use the Number of Dice spin button to set the number of dice (1 to 4). Each die is fair, so the probabilities of 1, 2, 3, 4, 5, 6 are identical at 1/6. The Speed slider is disabled. The Sound check box adds sound effects. With the Throws to Make combo box you can choose 10, 20, 50, 100, 200, 500, or 1000 throws. The Reset Experiment button resets the counter to zero. The Throw button throws the dice once. If you want to see the end result, click the Skip to End button to finish the random simulation.

d. Click the Notebook button. Click Scenarios, select Inspecting Cars, then click OK. In this scenario, you can set the Mean Defects Per Car (1 to 10). The Speed slider controls the rate at which cars move along the assembly line. Using the Cars to Test combo box you can choose 10, 20, 50, 100, 200, 500, or 1000 cars. The Sound check box adds sound effects. The Reset Experiment button resets the car counter to zero. The Run Experiment button initiates the process, and the Pause Experiment button lets you pause to look at the histogram. If you get tired of watching the cars go by, the Skip to End button finishes the random simulation.

3. **Options**
   You can have each histogram frequency labeled as a relative frequency (the number of occurrences of each event divided by the number of times the experiment is repeated). This option makes it easier to estimate empirical probabilities.

4. **Copying Graphs**
   To copy the histogram, highlight the histogram, then select Copy from the Edit menu bar at the top of the screen (or type Ctrl-C).   It can then be pasted into another application, such as a spreadsheet or document.

5. **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 *Visual Statistics*.

6. **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.

Basic Learning Exercises   Name _____

Tossing Coins

In the notebook, click Scenarios, click Tossing Coins to select the scenario, then click OK.

1. Set the Number of Coins spin button to 1 and the Tosses to Make spin button to 50. Click the Toss button ten times. How many times did you get tails (0)? Heads (1)? How many would you expect? Is your sample result much different from what you would expect?

2. Click the Skip to End of Experiment button. Now you have 50 tosses. How many times did you get tails (0) and heads (1)? How many would you have expected? Is your sample result very different from what you would expect? Would you expect the small sample (10 tosses) to differ more from its expected value than the large sample (50 tosses)? Explain.

3. Repeat Exercises 1 and 2. Did your results in this second set of experiments behave like the first series? How does this experiment support the idea that outcomes of a stochastic process exhibit regularity even though the underlying process is random?

Probability Trees

4. Click the Notebook button. Click Scenarios, click Probability Trees to select the scenario, then click OK. Set the Tree Levels to 1, Probability of Falling Right to 0.5, and Balls to Drop to 100. Click the Run Experiment Button. Drag the Speed slider to see its effect. If you wish, you can click the Skip to End of Experiment button. Are the sample histogram frequencies close to what you would expect? Is this experiment like the one-coin experiment? Explain.

5. Decrease the Prob of Falling Right to 0.3. How many times did you get "success"? Was this what you expected? Explain why this is like a baseball player whose batting average is 0.300 going to bat 100 times.

6. Set the Prob of Falling Right to 0.5. Increase the Tree Levels to 4. Conduct the experiment. What is the most frequent outcome? Explain why this is like flipping four coins.

7. If a professional basketball player is shooting free throws, what would be the probability of "success"? Set the Prob of Falling Right to that value. Set the number of Tree Levels to 10, to simulate shooting 10 free throws. Conduct the experiment. Was the most frequent outcome what you would expect? Why is this unlike flipping 10 coins?

8. Set the Prob of Falling Right to 0.20 and Tree Levels to 5. On the menu bar, select Options and click Display Relative Frequencies. Run the experiment and record the relative frequencies in the table below. Increase Balls to Drop to 500 and repeat. Repeat for 1000 balls. The theoretical (binomial) probabilities for 5 trials and a 0.2 probability of success are $P(0) = 0.3277$, $P(1) = 0.4096$, $P(2) = 0.2048$, $P(3) = 0.0512$, $P(4) = 0.0064$, and $P(X=5) = 0.0003$. Which of the three experiments (100, 500, 1000 balls) should give the best empirical estimate of the true probabilities? Explain your reasoning.

Balls	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)
100						
500						
1000						

9. Why might it be advantageous to use relative frequencies in summarizing a random experiment? Why might it be disadvantageous?

Intermediate Learning Exercises   Name _____

Throwing Dice

In the notebook, click Scenarios, click Throwing Dice to select the scenario, then click OK. Select Options and deselect Display Relative Frequencies so that raw frequencies will be displayed.

10. Set the Number of Dice to 1 and the Number of Throws to 100. Click the Throw button 6 times. How many times did you get each outcome? How many would you expect? Then click the Throw button 6 more times (12 throws altogether). How many times did you get each outcome? How many would you expect? Would any of your sample results so far make you doubt that the die is fair?

11. Click the Skip to End of Experiment button. Describe the distribution. How many of each outcome would you expect?   How different are the actual frequencies from your expectation? Do you think the die is fair? Are 100 throws enough to give you confidence in your answer?

12. Increase the Number of Throws to 1000. How many of each outcome would you expect? How different are the actual frequencies from your expectation? Does a large sample guarantee frequencies that are what we expect?

13. Set the Number of Dice to 2 and the Number of Throws to 100. Click the Skip to End of Experiment button. How many of each outcome would you expect? Describe the resulting sample histogram (e.g., its domain, mode, and general shape). Is this what you expected? Increase the Number of Throws to 1000 and repeat the experiment. Was 100 throws enough to give you confidence that you know the shape of the true distribution? Was 1000 enough?

Inspecting Cars
In the notebook, click Scenarios, click Inspecting Cars to select the scenario, then click OK. Select Options and select Display Relative Frequencies.

14. Set the Mean Defects per Car to 1 and the Number of Cars to 100. Click the Run Experiment button. Observe the process of data accumulation for a while and then click the Skip to End of Experiment button. Describe the resulting histogram (domain, mode, skewness, tails).

15. Increase the Number of Cars to 1000 and click the Skip to End of Experiment button. Record the relative frequencies below. Describe the resulting histogram (domain, mode, skewness, tails). Repeat this experiment two more times. The theoretical (Poisson) probabilities are P(0) = 0.3679, P(1) = 0.3679, P(2) = 0.1839, P(3) = 0.0613, P(4) = 0.0153, P(X=5) = 0.0031, P(X=6) = 0.0005, and P(X=7)= 0.0001. How close did you come?

Sample	P(X=0)	P(X=1)	P(X=2)	P(X=3)	P(X=4)	P(X=5)	P(X≥6)
1							
2							
3							

16. Keep the Number of Cars at 1000 but set the Mean Defects per Car to 5. What is the domain? Is there a clear mode? Is the distribution more or less skewed than in the previous exercise?

17. Keep the Number of Cars at 1000 but set the Mean Defects per Car to 20. Describe the domain, centrality, dispersion, and skewness of the distribution.   Can you generalize about the effect of increasing the mean on the skewness of the distribution? On the mode?

Advanced Learning Exercises   Name _____

Distribution Shape
In the notebook, click Scenarios, click Tossing Coins to select the scenario, then click OK. Select Options and choose Show Relative Frequencies.

18. Set the Number of Coins to 2 and set the Tosses to Make to 1000. What is the expected relative frequency of 0, 1, and 2 heads? Explain why (i.e., describe the sample space). Click the Skip to End of Experiment button. Is 1000 a large enough number of repetitions to feel confident that the empirical probabilities are reliable?

19. Set the Number of Coins to 4. List a few elements of the sample space and tell how it differs from the previous exercise. What is the expected histogram shape for the number of heads? What mode would you expect? Click the Skip to End of Experiment button. Describe the resulting histogram. Were your *a priori* predictions correct? What is the empirical probability of P(1)? Do you have confidence in this empirical probability (repeat the sample a few times if you are not sure, to see how much variation there is)?

20. Set the Number of Coins to 8. What is the expected histogram shape for the number of heads? What mode would you expect? Click the Skip to End of Experiment button. Describe the resulting histogram. Were your *a priori* predictions correct? Would you say this histogram resembles a "normal" or "bell-shaped" curve?

Modes

In the notebook, click Scenarios, click Throwing Dice to select the scenario, then click OK. Select Options and choose Show Relative Frequencies.

21. Set the Number of Dice to 3 and the Number of Throws to 1000. Click the Skip to End of Experiment button. Describe the resulting sample histogram (e.g., its domain, mode, and general shape). Record the mode(s) for your sample. Repeat the sampling exercise four more times and record the mode(s) below. Is there a clear-cut mode?

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Mode(s)1					

22. Set the Number of Dice to 4 and the Number of Throws to 1000. Click the Skip to End of Experiment button. Describe the resulting sample histogram (e.g., its domain, mode, and general shape). What "ought" the mode to be? Would you consider this sample mode a reliable estimate of the theoretical mode? Generalize about what happens to centrality and dispersion as the number of dice increases.

Mean and Variation

In the notebook, click Scenarios, click Inspecting Cars to select the scenario, then click OK. Select Options and select Display Relative Frequencies.

23. Set the Number of Cars at 100 and set the Mean Defects per Car to 4. Click the Skip to End of Experiment button. Record the minimum, maximum, and calculate the range. Repeat for Mean Defects per Car of 8, 12, 16, and 20 defects. Describe what is happening to the range. Explain why the range changes as you increase the mean number of defects.

Sample	Mean = 4	Mean = 8	Mean = 12	Mean = 16	Mean = 20
Maximum					
Minimum					
Range					

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 either the coin tossing or dice scenario and the relative frequency option. Choose 1 coin and 10 tosses or 1 die and 10 throws. Do the sampling experiment 5 times, saving each sample histogram for your report. What was the maximum departure from the expected proportions? Discuss how the histograms illustrate the concept of sample variation. Repeat the experiment using 1000 throws. What was the maximum difference from the expected proportion? Discuss how these five histograms illustrate regularity. Using one histogram, deselect the relative frequency option. Explain how the count histogram contains the same information as the relative frequency histogram.

2. Select the probability tree scenario and deselect the relative frequency option (vertical axis on histogram should say Count). Choose 5 tree levels, 10 balls, and a 0.3 probability of falling right. Do the sampling experiment 5 times, saving each sample histogram for your report. Compare the histograms and explain how they illustrate the concept of sample variation. Repeat the experiment once using 1000 balls. Compare the histograms and explain how they illustrate the concept of regularity. Set the probability of falling right to 0.7 and conduct the experiment once with 1000 balls. For 1000 balls, describe the similarities and differences between the 0.3 and 0.7 probability of falling right.

3. Select the inspecting cars scenario and the relative frequency option. Choose 10 cars and set the mean number of defects to 1. Do the sampling experiment 5 times, saving each sample histogram for your report. In each experiment, what was the empirical probability of getting exactly the mean that you set? How does this illustrate sample variation? Repeat using 1000 cars. How does this demonstrate regularity? Are the 1000 car histograms symmetric? What is the mode in each case? Explain why the mean and mode are generally different in a non-symmetric histogram.

Team Learning Projects

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

1. This project is for a team of 3. Select the probability tree scenario and the relative frequency option. Choose 1000 balls. One team member should choose 4 tree branches, another 6 branches, and the last 8 branches. Each team members should vary the probability of falling right from 0.1 to 0.9 in increments of 0.1, producing a histogram for each experiment. The objective is to demonstrate variation and regularity and to illustrate how the shape of a histogram varies as you change the parameters of the process. What type of distribution is this, and what are its parameters?

2. This is a project for a team of 2. Select the dice scenario and raw frequency option. Choose 100 throws. Each team member should choose a different number of dice (1 or 2). Calculate the *a priori* probability and expected frequency for each outcome. Do each sampling experiment 5 times, saving each sample histogram. Discuss how the histograms illustrate the concept of sample variation. Compare the actual and *a priori* frequencies by calculating their differences and their ratios. What was the maximum departure from the expected frequency? Repeat each experiment 5 times using 1000 throws. Compare the actual and *a priori* frequencies by calculating their differences and their ratios. Was the maximum departure from the expected frequency less than with 100 throws? The objective is to show if the empirical probabilities improve as you increase the number of repetitions.

3. This is a project for a team of 3 or 4. Select the inspecting cars scenario and the relative frequency option. Choose 10 cars. One team member should choose 4 as the mean number of defects, another 8, the third 12, and the fourth (if any) 16. Do the sampling experiment 5 times, producing a histogram for each experiment. In each experiment, what was the empirical probability of getting exactly the mean that you set? Repeat using 50 cars. Repeat using 200 cars. The objectives are to demonstrate variation and regularity, to see how the probability of getting exactly the process mean varies as the mean changes, and to see if the empirical probabilities stabilize as the number of repetitions increases.

Self-Evaluation Quiz

1. What is a stochastic process?
   a. One whose outcomes are known in advance.
   b. One that produces the same outcome every time.
   c. One whose outcomes are not known in advance.
   d. One involving certainty.
   e. Both c. and d. are correct.

2. The sample space contains
   a. elementary outcomes of an experiment.
   b. random variables.
   c. probabilities of events.
   d. parameters of a process.
   e. both b. and c.

3. A random variable is
   a. an elementary outcome.
   b. an event in the sample space.
   c. a number assigned to an event.
   d. a probability of an event.
   e. a parameter of a process.

4. The histogram is used to assess
   a. centrality.
   b. variation.
   c. dispersion.
   d. symmetry.
   e. all of the above.

5. The probability P(A) of an event A has which characteristics?
   a. P(A) must not exceed 1.
   b. P(A) must be non-negative.
   c. P(A) may be determined empirically using relative frequencies.
   d. P(A) may be known *a priori* in some cases.
   e. P(A) has all of the above characteristics.

6. The parameter(s) of a distribution
   a. are needed to characterize the probability distribution.
   b. are impossible to estimate from a large sample.
   c. are usually known *a priori*.
   d. have no effect on centrality or variation.
   e. must be specified by design engineers.

7. As we increase the number of random trials we get
    a. a better idea of the probability of events.
    b. a smaller degree of dispersion.
    c. a narrower histogram scale.
    d. increasing parameter values.
    e. all of the above.

8. Which are the *parameters* of the distribution for the probability tree experiment?
    a. The number of balls dropped.
    b. The number of tree levels.
    c. The speed at which balls are dropped.
    d. The probability of falling right.
    e. Both b. and d. are correct.

9. If you roll 3 dice instead of 2, which will *not* occur?
    a. The center of the histogram will shift right.
    b. The mode of the histogram will increase.
    c. The range of the histogram will increase.
    d. The histogram will remain symmetric.
    e. The probability of rolling 7 will stay the same.

10. If you flip 4 coins 100 times
    a. you would see regularity in the relative frequencies of heads.
    b. you would expect a range from 0 to 4.
    c. you would expect to get 2 heads about 37.5 percent of the time.
    d. you would be observing a stochastic process.
    e. you would find all of these things.

11. The binomial model is used to describe a process
    a. with a constant probability of success for each trial.
    b. with a specified number of independent trials.
    c. that has only two outcomes.
    d. whose outcomes are discrete (e.g., 0, 1, 2, 3).
    e. with all of the above characteristics.

12. Which is *not* a characteristic of the Poisson model?
    a. It involves counting events over time or space.
    b. It has a definite upper limit on X.
    c. It describes a stochastic process with a known distribution.
    d. It has one parameter.
    e. It might be used to describe arrivals of telephone calls per hour.

Glossary of Terms

A priori probability A probability that is known prior to looking at actual sample results. The probability is calculated by reasoning alone, using the known sample space of the experiment. See **Empirical probability**.

Binomial distribution Two-parameter distribution describing discrete data generated by a binary (success/failure) experiment with n independent trials and constant probability of success.

Centrality General characterization of the location of the middle or "typical" values in a distribution (e.g., mean, median, mode).

Discrete random variable Random variable with a countable number of values.

Dispersion General characterization of the degree of variation about the center of a distribution (e.g., range).

Domain The values that can be assumed by a random variable (e.g., X = 0, 1, 2, 3).

Empirical probability A probability that is estimated by looking at relative frequencies of actual outcomes in a sampling experiment. It is based on evidence rather than reasoning. An empirical probability becomes closer to the *a priori* probability as the number of repetitions of the experiment increases. See **A priori probability**.

Event A collection of elementary outcomes of a stochastic experiment. For example, when you flip a coin once, "heads" or "tails" are the two events that could occur.

Frequency The number of occurrences of an event in a particular stochastic experiment.

Histogram Bar chart showing on the horizontal axis the values of a random variable and on the vertical axis the frequency of occurrence of each value.

Independent Reference to events that do not influence one another.

Mode The data value that occurs most frequently in a sample. It is not necessarily unique. If there are two modes, the data are called bimodal. The mode is most useful for discrete data with a small range. In a probability distribution, the mode is the X value that has the highest probability of occurrence.

Normal distribution The "bell-shaped" or Gaussian distribution. It is sometimes used as a point of reference with which to compare a sample histogram.

Parameter One or more numerical characteristics of a distribution that determine its probability distribution (centrality, dispersion, skewness, and other characteristics).

Probability distribution Value assigned to every value X of a random variable. These probabilities must sum to unity.

Probability of an event A number between 0 and 1 assigned to every value of a random variable.

Random variable An event of a stochastic experiment that is assigned a numerical value. For example, the number of heads in 15 flips of a fair coin is a random variable.

Regularity Tendency for stable patterns to emerge in relative frequencies when a random experiment is repeated many times.

Relative frequency The number of times an event occurs divided by the number of repetitions of the random experiment. For example, if you throw a pair of dice 216 times and get 7 exactly 35 times, the relative frequency of 7 is 35/216 or 0.162. See **Empirical probability**.

Sample space The set of all possible outcomes of an experiment of chance. Each element in the sample space is called an elementary outcome.

Skewness Degree of asymmetry of a probability distribution or a histogram. *Positive* skewness indicates a long right tail and *negative* skewness indicates a long left tail. See **Symmetry**.

Stochastic process Data-generating situation that produces an outcome which cannot be predicted in advance.

Symmetry Shape of a probability distribution or histogram whose tails are of equal length.

Triangular distribution Three-parameter distribution whose shape has a single sharp peak on either side of which the probabilities decline toward zero in a linear fashion.

Uniform distribution Two-parameter distribution which assigns the same probability to every value of X.

Variation Natural tendency for outcomes of a stochastic process to vary. See **Dispersion**.

Solutions to Self-Evaluation Quiz

1. c   Do Exercises 1–5. Consult the Glossary. Read the Overview of Concepts.
2. a   Do Exercises 1–7. Consult the Glossary. Read the Overview of Concepts.
3. c   Consult the Glossary. Read the Overview of Concepts.
4. e   Do Exercises 6–9. Consult the Glossary. Read the Overview of Concepts.
5. e   Do Exercises 7–8. Consult the Glossary. Read the Overview of Concepts.
6. a   Do Exercises 4–8. Read the Overview of Concepts. Do Team Projects 1 or 2.
7. a   Do Exercises 8–9. Consult the Glossary. Read the Overview of Concepts.
8. e   Do Exercises 6-9. Consult the Glossary.
9. e   Do Exercises 10–13. Do Individual Project 3 or Team Project 2.
10. e   Do Exercises 1–5 and 18–20. Do Individual Project 2.
11. e   Read the Overview of Concepts. Consult the Glossary.
12. b   Do Exercise 23. Read the Overview of Concepts. Consult the Glossary.