5.1A

Introduction, The Idea of Probability, Myths about Randomness

- **The Idea of Probability**

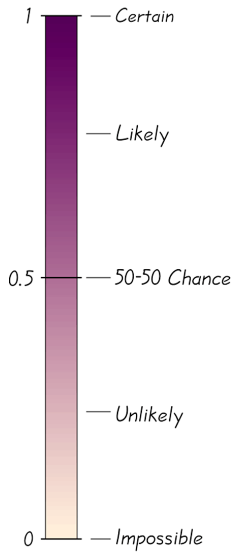
Chance behavior is unpredictable in the short run, but has a regular and predictable pattern in the long run.

The law of large numbers says that if we observe more and more repetitions of any chance process, the proportion of times that a specific outcome occurs approaches a single value.

The probability of any outcome of a chance process is a number between 0 (never occurs) and 1(always occurs) that describes the proportion of times the outcome would occur in a very long series of repetitions.

$$0 \leq P(A) \leq 1$$

Impossible to occur (0) — Certain to occur (1)



Can you think of an event that cannot occur?

Can you think of an event that will definitely occur?

Example:

The probability of getting a sum of 7 when rolling two dice is 1/6. Interpret this value.

If we roll two dice a large number of times, we can expect a sum of 7 to occur 1/6 of the time.

Example

Athletes are often tested for use of performance-enhancing drugs. Drug tests aren't perfect – they sometimes say that an athlete took a banned substance when that isn't the case (a "false positive"). Other times, the test concludes that the athlete is "clean" when he or she actually took a banned substance (a "false negative"). For one commonly used drug test, the probability of a false negative is 0.03.

a. Interpret this probability.

If a drug test is conducted many times, we can expect a false negative 3% of the time.

b. Which is a more serious error in this case: a false positive or a false negative? Justify your answer.

Myths About Randomness:

1. Short Term Regularity

Myth: Random phenomena is predictable in the short run

Truth: Patterns are very common in the short run

Example

Roll a die 12 times and record the result of each roll. Which of the following outcomes is more probable?

123456654321 154524336126

These outcomes are both equally (un)likely, even though the first set of rolls has a more noticeable pattern.

Myths About Randomness:

2. Law of Averages

Myth: If I flip Tails 6 times in a row, I am more likely to get a Heads the the 7th time.

Truth: If I flip Tails 6 times in a row, my chance of getting a Heads the 7th time is still 50%.

Example

In casinos, there is often a large display next to every roulette table showing the outcomes of the last several spins of the wheel. Since the results of previous spins reveal nothing about the results of future spins, why do the casinos pay for these displays?

Because many players use the previous results to determine what bets to make, even though it won't help them win. And as long as the players keep making bets, the casino keeps making money.

5.1B

Simulations

- **Simulation**

The imitation of chance behavior, based on a model that accurately reflects the situation, is called a simulation.

It is used to estimate probabilities when they are difficult to calculate theoretically.

Simulation Steps:

1. What is the question of interest about some chance process?
2. Describe how to use a chance device to imitate one repetition of the process. Explain clearly how to identify the outcomes of the chance process and what variable to measure.
3. Perform many repetitions of the simulation.
4. Use the results of your simulation to answer the question of interest.

Example

On her drive to work every day, Llana passes through an intersection with a traffic light. The light has probability 1/3 of being green when she gets to the intersection. Explain how you would use each chance device to simulate whether the light is green or not green on a given day.

a. A six-sided die

Assign 2 numbers (1,2) to represent a green light and 4 numbers (3,4,5,6) to represent a red light. Roll the die and record the result.

b. Table D of random digits

Let 0,1,2 represent a green light and 3,4,5,6,7,8 represent a red light. Look up a number in the table. If it is 9, ignore it.

c. A standard deck of playing cards

Let any red card represent a red light and a spade represent a green light. Shuffle the deck and pick a card. Ignore clubs.

Example

Suppose that a basketball announcer suggests that a certain player is streaky. That is, the announcer believes that if the player makes a shot, then he is more likely to make his next shot. As evidence, he points to a recent game where the player took 30 shots and had a streak of 7 made shots in a row. Is this evidence of streakiness or could it have occurred simply by chance? Assuming this player makes 48% of his shots and the results of a shot don't depend on previous shots, how likely is it for the player to have a streak of 7 or more made shots in a row?

Using randint(1,100) in your calculator, let 1-48 represent a made free throw and 49-100 represent a missed free throw. Simulate an attempted free throw 30 times and make a note of the longest streak of made shots.

Example

Suppose I want to choose a simple random sample of size 6 from a group of 60 seniors and 30 juniors. To do this, I write each person's name on an equally sized piece of paper and mix them up in a large grocery bag. Just as I am about to select the first name, a thoughtful student suggests that I should stratify by class. I agree, and we decide it would be appropriate to select 4 seniors and 2 juniors. However, since I already mixed up the names, I don't want to have separate them all again. Instead, I will select names one at a time from the bag until I get 4 seniors and 2 juniors. This means, however, that I may need to select more than 6 names (e.g. I may get more than 2 juniors before I get the 4 seniors). Design and carry out a simulation to estimate the probability that you must draw 8 or more names to get 4 seniors and 2 juniors.

Using pairs of digits from Table D, we'll label the 60 seniors 01-60 and the 30 juniors 61-90. Numbers 00 and 91-99 will be skipped. Moving left to right across a row, we'll look at pairs of digits until we have 4 *different* labels from 01-60 and 2 *different* labels from 61-90. Then, we will count how many *different* labels from 01-90 we looked at.

Here is an example of one repetition, using line 101 from Table D:

19 (senior) 22 (senior) 39 (senior) 57 (senior) 34 (senior)

05 (senior) 75 (junior) 62 (junior)

In this example, it took exactly 8 selections to get at least 4 seniors and at least 2 juniors.