Monday, October 26: Chapter 11: Using Simulation to Estimate Probabilities

Suppose a cereal company places one of four toys in every box of cereal. Furthermore, the company claims that each toy is produced in the same quantity so each of the toys is equally likely to show up in a randomly chosen box. Suppose I want to get all 4 toys, but it takes me 15 boxes to find each of the four. Is this evidence that the toys are NOT uniformly distributed? Estimate the probability that it takes 15 or more boxes to get all 4 toys assuming that the company is telling the truth.

There are 7 steps for a simulation: Note: These steps are slightly modified from the book.

1. Identify the component to be repeated.

2. Explain how you will model the outcomes.

3. Explain how you will simulate the trial. A trial (or run) is the sequence of events that we are pretending will take place.

4. State clearly what the response variable is.

5. Give one example and run several trials. The more trials you run, the more precise your estimate will be.

6. Analyze the results. This is usually accomplished with a table or graph of the response variable.

7. State your conclusion in the context of the problem.

If it took me 20 boxes to find all 4, would that be convincing evidence against the company?

How else could we model the outcome (step 2)?

A person who claims to have ESP says that she can identify the symbol on the back of a card (circle, triangle, or square) without seeing it. To test this claim, you shuffle the three cards, show her the back of a randomly selected card, and record if she correctly identifies the symbol. Then, you replace the card, reshuffle, and repeat this procedure 19 more times (20 total). Overall, she identified 10 correctly. Does this support her claim of ESP, or could this have occurred by chance? Use a simulation to estimate the probability of getting 10 or more correct just by guessing. **Use a table of random digits**.

HW #25: SR (257-264), page 266 (5, 7, 11, 19)

Note: if not specified, do 10 trials per simulation.

<u>Tuesday, October 27</u>: Chapter 11: Using Simulation to Estimate Probabilities (SUB)

Suppose that 150 seniors at a particular school (including 19 members of StuGo) signed up for "senior parking," which gives each student the right to a particular parking space close to the campus for the year. Also, suppose that each student's spot is determined by a lottery. The student body became suspicious, however, when StuGo members were awarded 5 of the 10 best spots. Is the suspicion warranted or could this have occurred by chance? Conduct a simulation to estimate the probability of StuGo members getting at least 5 of the 10 best spots, assuming the lottery is fair.

Suppose that a certain member of the Phoenix Suns is considered a streaky player. That is, many people believe that his shots are not independent. If he makes a shot, they believe he is more likely to make his next shot and vice-versa. As evidence, people point to a recent game where this 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 60% of his shots and that his shots are independent, design and conduct a simulation to estimate the probability of having at least one streak of 7 or more made shots in 30 attempts.

Chapter 14: Probability

The _________________________ of an event is its long-run relative frequency.

For any random phenomenon, each attempt, or _________, generates an _____________. An outcome is what happened on a trial. For example, each time we waited for a bus was a _______ and the possible outcomes were 1 minute, 2 minutes, etc…

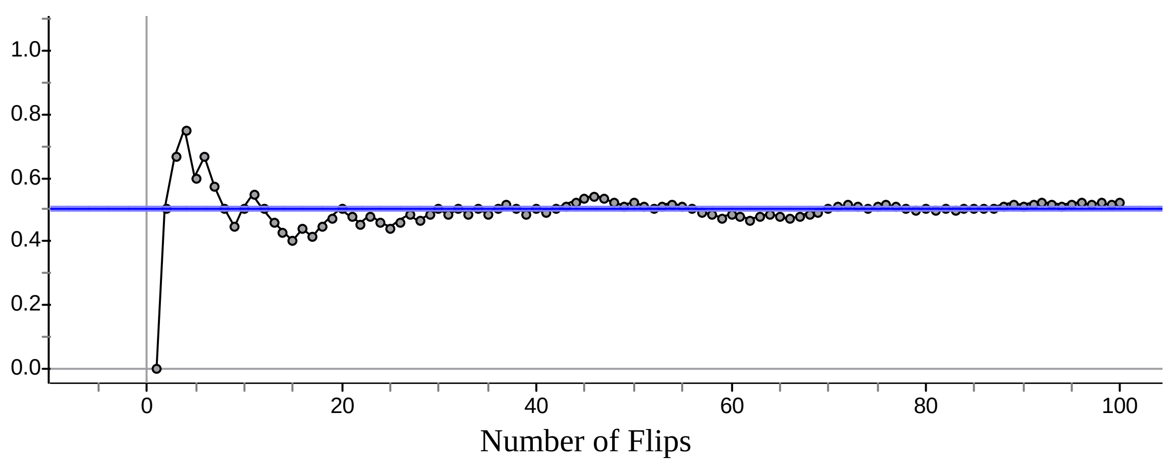
Instead of individual outcomes, sometimes we consider _________, which are combinations of outcomes. For example, if a trial consisted of rolling a die, the outcomes would be 1, 2, 3, 4, 5, 6. An event might be:

If we want to know about combinations of outcomes, it helps if the individual trials are independent. If trials are _____________________, then the outcome of one trial doesn't influence or change the likelihood of another trial. For example, when we were waiting for the bus, the trials are independent, since knowing how long I waited in one trial does not affect how long I waited in the next trial.

However, if we are drawing cards without replacement from a deck, the trials are not independent. If we draw a club on our first trial, then we know the next card is less likely to be a club and more likely to be a heart, spade, or diamond. This is why casinos use 6 decks for Blackjack instead of 1.

The _____________________________ says that the long-run relative frequency of repeated independent events get closer and closer to the true relative frequency (probability) as the number of trials increases.

The picture below shows the relative frequency of "heads" after each of 100 flips of a coin. Notice how when there are a small number of flips, the relative frequency of heads can vary quite a bit from 0.5, but when the number of flips increases the relative frequency gets closer to 0.5.



Many people misunderstand this Law and apply it to situations with a small number of trials. For example, if we flip a coin 5 times and it lands on heads 5 times, some people will believe that the 6^{th} trial is more likely to be tails, since "tails is due." This is sometimes called the "Law of Averages." Don't believe it.

However, since the trials are independent, the probability that the coin lands on heads never changes, no matter what has happened before. The Law of Large Numbers only guarantees that the relative frequency of an event will approach the true probability after a ____________ number of trials. The coin doesn't purposefully compensate for its run of heads by landing on tails, but the results of many trials will eventually overwhelm any departures from what's expected.

For example, we may start with 5 heads in a row (100% heads), but after 100 more flips, we should be close to 55 heads and 50 tails (52.4% heads), and after 900 more flips, we should be close to 505 heads and 500 tails (50.2% heads).

Note: This is why casinos have boards which post the past 15 or 20 outcomes on a roulette wheel. The casino knows that gamblers misunderstand the law of large numbers and will be encouraged to bet more since they think they know what numbers are more likely to come up.

HW #26: SR (326-330), page 266 (6, 8, 13, 31), page 339 (1-7 odd)
Note: if not specified, do 10 trials per simulation.

<u>Thursday, October 29</u>: Chapter 14: Basic Probability Rules

1. A probability is any number between 0 and 1. That is, for any event A, $0 \le P(A) \le 1$.
   - Remember that a probability is a relative frequency. We cannot have more successes than the number of trials (so P(A) ≤ 1) and we cannot have a negative number of successes (so P(A) ≥ 0).
   - If an event has probability = 0, then the event will never occur.
   - If an event has probability = 1, then the event must occur.

2. Something has to happen. That is, when we consider a single trial of a random phenomenon, one of possible the outcomes must occur. For example, if we are flipping a coin, we must either get heads or tails.
   - All of the possible outcomes of a random phenomenon are called the ____________________ (S), thus P(S) = 1
   - If we add the probabilities of all the possible outcomes and it doesn't equal 1, then the assignment of probabilities is not legitimate (or there may be rounding error).

3. The probability of an event occurring is 1 minus the probability it doesn't occur. That is, P(A) = 1 – $P\left(A^c\right)$. For example, if the probability of rain today is .1, then the probability that it will not rain today is 1 - .1 = .9.
   - The event "not A" is called the complement of A and is denoted A^c.
   - Other books may use different notation, including: A', ~A, etc.

4. For two disjoint events A and B, the probability that one <u>or</u> the other event occurs is the sum of the probabilities of the two events. That is, P(A or B) = $P\left(A \cup B\right)$ = P(A) + P(B) if A and B are disjoint.
   - Two events are ______________ if they do not share any outcomes. For example, suppose you randomly selected a registered voter and asked him which political party he belonged to. Since you can only register as a member of one party, being a democrat and being a republican are disjoint events.
   - Disjoint events are also called ______________________________ events.
   - So, if the probability of being a Republican is .4 and the probability of being a Democrat is .45, the probability of being a Republican or a Democrat =
   - This rule also works for more than 2 disjoint events. For example, if the probability of being a Libertarian is 5%, then P(R or D or L) =
   - However, you cannot use this addition rule when the events are not disjoint. For example, for students in statistics, P(senior) = .75 and P(own a TI calculator) = .9, but P(senior or own a TI) $\neq$ .75 + .9 = 1.65. The events are not disjoint and we cannot have probabilities over 1!

5. For two *independent* events A and B, the probability that both A ***and*** B occur is the product of the probabilities of the two events. That is, P(A and B) = $P(A \cap B)$ = P(A) · P(B) if A and B are independent.

- Recall from yesterday that two events are ______________________ if knowing the outcome of one event does not change the probability of the other event.
- For example, if we flip a coin and roll a die, the events "heads" and "6" are independent, since knowing the outcome of the coin flip does not change the probability that the die lands on 6.
- Thus, P(heads and 6) =

- This rule also works for more than 2 independent events. For example, the probability of rolling 3 sixes in a row is: P(6 and 6 and 6) =

- However, you cannot use this multiplication rule when the events are not independent. For example, if a class of 30 students has 16 females, the probability of selecting (without replacement) 2 females is not (16/30)(16/30), since these events are not independent. Knowing that the first person selected was female changes the probability that the second person selected is female (from 16/30 to 15/29).

Summary: When you want P(A or B), you add the probabilities (assuming the events are disjoint) and when you want P(A and B), you multiply the probabilities (assuming the events are independent).

If the probability of rain in Oro Valley is 50% and the probability of rain in Marana is 50%, what is the probability that it will rain in both places?

If the probability of rain in New York City is also 50%, what is the probability that is will rain in both Oro Valley and NYC?

Suppose that 50% of cars sold in this country are made in the USA, 30% in Japan, and 10% in Germany.

1. Is this a valid assignment of probabilities? How do you know?

2. Assume that the remaining 10% of cars are made in other countries.
   a. State the probability distribution for "make of car". Note: a probability distribution lists all the outcomes of a variable and their probabilities.

b. What is the probability that a car selected at random is not made in the USA?

c. What is the probability that a car selected at random is made in Germany or Japan?

d. This rule used in (c) only works for disjoint events. Explain why these events are disjoint.

e. What is the probability that two cars selected at random are both from Japan?

- Note: In cases where the population is very large compared to the sample (population size greater than 10 times the sample size), it is OK to treat trials as independent even when sampling without replacement since the change in probabilities will be very small.

f. What is the probability that none of the 3 cars came from Germany?

g. What is the probability that at least one of the three cars is made in the Germany?

h. What is the probability that the first time you get a Japanese car is on your fourth selection?

i. What is the probability that two cars selected at random are both from the USA or both from Japan?

Simulation Question: Suppose we did a gift exchange in our class. To decide who each person will give a present to, everybody's name is placed in a hat and each person draws a name from the hat. What is the probability that no one will draw his/her own name from the hat?

HW #27: SR (331-337), page 268 (33 – do 10 trials), page 339 (2-8 even, 9-21 odd)

Monday, November 2: Chapter 15: Probability Rules!

Review: For any random phenomenon, each ________ generated an ________________. An ________ is any set or collection of outcomes. The collection of all possible outcomes is called the _______________________ (S).

For example, if a family has 2 children, the sample space is {BB, BG, GB, GG}. In this case all four outcomes are equally likely.

If you are dealt 5 cards from a deck and record the number of aces you get, what is the sample space?

Are these outcomes equally likely?

When all of the outcomes in the sample space are equally likely, the probability of an event A is:

$$P(A) = \frac{\text{count of outcomes in A}}{\text{count of all possible outcomes}}$$

For example, the probability that a family with two children has two boys is 1/4. But, the probability of getting 4 aces in 5 cards is not 1/5 since the outcomes are not equally likely.

What is the probability of drawing a single card from a deck and getting a red card or an ace?

To answer this problem, we need to use the GENERAL ADDITION RULE, which works when events are disjoint and when events are not disjoint.

Note: When we use the word "or" we mean it in the inclusive sense: A or B means A occurs or B occurs or both A and B occur.

Suppose that 75% of students in statistics are seniors, 90% have a TI-83, and 70% are seniors with TI-83's. What is the probability that a randomly selected statistics student is a senior or has a TI-83?

Chapter 15: Conditional Probability

Sometimes the knowledge that one event has occurred changes the probability that another event will occur. In this case, the events are NOT independent.

- Your estimate of the probability of being in a car accident increases if you know that it is raining outside
- Suppose that the pass rate on the AP Statistics exam is 80%. That is, for a randomly selected student, P(pass) = .80. However, if you know that the student got a B in the class, then the probability increases to 95%. That is, for a randomly selected student, P(pass given that you have a B) = .95.
- Notation: P(pass | B) = .95

Probabilities in the form P(A | B) are called _________________ probabilities and pronounced "the probability of A occurring given that B has already occurred"

The following data is about the 2201 passengers on the Titanic. 367/1731 males survived and overall 711 survived. Express this in two-way table with gender and survival as the variables.

Review: A _____________________ is a way to summarize the relationship between two categorical variables. It lists the outcomes of one variable down the left side, the outcomes of the other variable across the top, and the frequencies (or relative frequencies) in each cell.

Review: The overall distribution for gender is called the __________________distribution of gender (ex: 79% are males and 21% are females). The distribution for a specific gender category is called a _______________ distribution (ex: 73% of the females survived and 27% of the females didn't survive).

Suppose you randomly selected a name from the Titanic's passenger list.

a. P(survived) = ?

b. P(male) = ?

c. P(female $\cap$ survived) = ?

d. P(female $\cup$ survived) = ?

e. P(survived | female) = ?

f. P(male | survived) = ?

g. P(died | female) = ?

h. P(female | died) = ?

i. P(female | survived) = ?

Def: Let E and F be two events. The conditional probability of event E given that event F has occurred is:

HW #28: SR (344-350), page 268 (25 – do 10 trials), page 362 (1-13 odd)

<u>Tuesday, November 3</u>: Chapter 15: The General Multiplication Rule and Independence

In the last chapter, we learned a rule for finding P(A and B) when A and B are independent:

P(A and B) = P(A) P(B)

However, now that we understand conditional probability, we can use the general multiplication rule which works whether or not the events are independent:

P(A and B) = P(A) P(B|A)

This rule comes directly from the conditional probability formula we learned yesterday

What is the probability of drawing 2 aces in a row from a deck (without replacement)?

Suppose that 23% of CDO students are seniors and that 65% of seniors are taking a math class. What is the probability that a randomly selected CDO student is a senior and is taking a math class?

Suppose that there are 12 girls on a softball team. Six of them are freshmen, 4 are sophomores, and 2 are juniors.

a. If two players are selected at random, what is the probability that they are both juniors?

b. Neither is a junior?

c. At least 1 is a junior?

d. They are of same grades?

Reminder: When we are sampling without replacement, successive selections are NOT independent. When the population size is known, you should adjust the probabilities accordingly.

Also, we can now formally define independence. Events A and B are INDEPENDENT whenever $P(B \mid A) = P(B \mid A^c) = P(B)$. That is, knowing the outcome of event A does not change the probability of event B. Whenever you are asked about independence, this is the definition you should use to justify your answer.

Also, we can see that the original multiplication formula is just a special case of the general formula, since $P(A)\,P(B|A)$ simplifies to $P(A)\,P(B)$ when A and B are independent.

Using the examples from yesterday:

- Were being male and surviving independent on the Titanic? Disjoint?

- Were being a senior and having a drivers license independent at CDO? Disjoint?

Note: If events E and F are independent and the data we have is from a random sample, then $P(E \mid F) \approx P(E)$. In other words, the probabilities may not be exactly the same because of sampling variability. How close do they need to be? Wait until chapter 26…

Suppose that in a certain company, 5% of the employees use drugs. The company decides to test all of its employees for drugs with a test that is 95% successful (it correctly identifies 95% of users as users and 95% of non-users as non-users). Let D = event that the employee uses drugs and P = event that the person tests positive.

a. Express the information given in terms of P and D.

b. Express the information in a tree diagram.

c. What is the probability that a person uses drugs and tests positive?

d. What is the probability that a person tests positive?

e. What is the probability that a person uses drugs or tests positive?

f. If a person tests positive, what is the probability that they use drugs? Does this result surprise you?

Note: the false positive rate of a test is P(positive | no drugs) and the false negative rate is P(negative | drugs)

- Which of these errors is worse? For the employees? For the company?
- How can we change these probabilities?

HW #29: SR (350-361), page 268 (27 – estimate average using 10 trials), page 364 (15-21 odd, 29, 35-41 odd)

<u>Thursday, November 5</u>: Chapter 15: Review Conditional Probability (SUB)

In a certain city, 45% of registered voters are Republicans, 40% are Democrats and the rest are Independents. Also, 40% of the Republicans are women, 55% of the Democrats are women and 60% of the Independents are women.

Make a tree diagram to display this information.

Suppose you randomly selected one registered voter.

a. What is the probability that you choose a person who is a Republican and a woman?

b. What is the probability that you choose a person who is a woman?

c. What is the probability that you choose a person who is a male or an Independent?

d. What is the probability that a randomly chosen male is a Republican?

Are being a woman and being a Democrat independent? Justify.

It is possible to use a table to solve this problem, but it takes more work to construct than a tree diagram:

In general, if you are given counts, then a table (or venn diagram) is easier to use. But, if you are given conditional probabilities, then a tree diagram is usually better.

In a survey of 1200 college students, 380 said they had they tried smoking, 800 had tried drinking, and 315 had tried both. Make a two-way table to display this information.

a. What is the probability that a randomly chosen respondent has tried neither?

b. What is the probability that a randomly chosen drinker has also tried smoking?

c. What is the probability that a randomly chosen smoker hasn't had a drink?

d. What is the probability that a randomly chosen respondent has tried at least one of them?

Is smoking independent of drinking? Justify.

Are smoking and drinking disjoint events?

AP Question:

HW #30: page 268 (34 – do 10 trials), page 341 (23-35 odd), page 362 (23, 25, 31, 43, 45)

<u>Tuesday, November 10</u>: Review chapters 11, 14, 15

Here is data from a random sample of 100 CDO students. Are being a male and playing a sport independent for students at CDO?

	Male	Female	Total
Sport	16	14	30
No Sport	32	38	70
total	48	52	100

Assuming the variables are independent in the CDO population, how likely is it to get a difference this large due to sampling variability. Use a simulation to estimate this probability. Is it possible that the variables really are independent?

AP Question:

<u>Thursday, November 12</u>: Chapter 11, 14, 15 Test