

Activity #4: Probability Rules, Discrete Random Variables, Conditional Probability, and Bayes' Theorem

1. In a class of 25 students, 10 are sophomores and 15 are juniors. Furthermore, 11 of the students are engineering majors, while 3 sophomores are engineering majors.

Let's use simple probability rules to calculate basic probabilities. If we randomly choose one of these 25 students:

$$P(\text{sophomore}) = \underline{\hspace{2cm}} \quad \underline{\hspace{2cm}} \leq P(\text{math major}) \leq \underline{\hspace{2cm}}^*$$

$$P(\text{freshman}) = \underline{\hspace{2cm}} \quad P(\text{sophomore or junior}) = \underline{\hspace{2cm}}^{**}$$

$$P(\text{not an engineering major}) = P(\text{engineering major})' = \underline{\hspace{2cm}}^{***}$$

From this simple example, we can generalize basic facts about probabilities. For any event A in sample space S:

- $0 \leq P(A) \leq 1$ Every probability is a number between 0 and 1
- $P(S) = 1$ The probability of the entire sample space is 1.0 (or the sum of all probabilities 1.0)
- $P(A)' = 1 - P(A)$ Complement Rule (the complement of A is defined to be "event A does not occur")

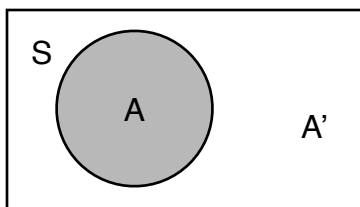
Try these:

$$P(\text{engineering major or sophomore}) = P(\text{engineering} \cup \text{sophomore}) = \underline{\hspace{2cm}}$$

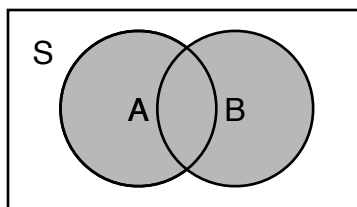
$$P(\text{junior and not an engineering major}) = P(\text{junior} \cap \text{engineering}') = \underline{\hspace{2cm}}$$

2. Those probabilities were a bit more difficult to calculate with the way the information was given. We can simplify our calculations if we organize information in one of two ways: Venn Diagrams or contingency tables.

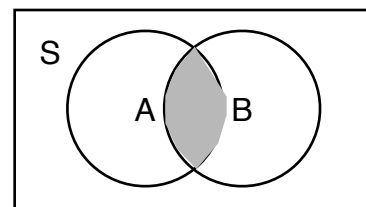
Let's start with a look at Venn Diagrams. In a Venn Diagram, the box represents the entire sample space. Circles represent variables (or categories of items). We shade areas to represent probabilities of interest.



$P(A)$
Probability of A



$P(A \text{ or } B)$
 $P(A \cup B)$
A union B
Either A or B occur



$P(A \text{ and } B)$
 $P(A \cap B)$
A intersect B
Both A and B occur

Create a Venn Diagram with the given information: In a class of 25 students, 10 are sophomores and 15 are juniors. Furthermore, 11 of the students are engineering majors, while 3 sophomores are engineering majors.

Using your Venn Diagram, calculate the following:

$$P(\text{engineering major or sophomore}) = P(\text{engineering} \cup \text{sophomore}) = \underline{\hspace{2cm}}$$

$$P(\text{junior and not an engineering major}) = P(\text{junior} \cap \text{engineering}') = \underline{\hspace{2cm}}$$

This time, create a contingency table. With two variables in our scenario, we'll create a 2-dimensional table. The rows should represent possible values of the first variable and the columns should represent possible values of the second variable.

		Total
Total		

Verify the contingency table gives you the same probabilities as the Venn Diagram. Then, answer the following:

$$P(\text{sophomore or junior}) = \underline{\hspace{2cm}}$$

$$P(\text{junior and engineering major}) = \underline{\hspace{2cm}}$$

Advice on calculating probabilities:

- In this class, we'll estimate probabilities via simulation, counting rules, and direct calculation.
- We will encounter about a dozen probability rules. Most of them will be common sense.
- While these rules can come in handy, don't rely on them. Understanding the rules > memorizing formulas.
- I think constructing contingency tables is the easiest way to get probabilities (without any formulas)
- The General Addition Rule (displayed below) is important and should be understood.

Addition Rules: ("or" probabilities)

- General Addition Rule: $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Draw a Venn Diagram to display that last rule.

3. The Three Stooges starred in 190 movies. The following table displays the movies according to the director and who played the third stooge:



3rd Stooge

		Director				Total
		Del Lord	Jules White	Ed Bernds	Other	
Curly		38	33	5	21	97
Shemp		1	55	20	1	77
Joe		0	16	0	0	16
Total		39	104	25	22	190

Source: <http://static.neatorama.com/images/2008-10/three-stooges.jpg>

If we select a movie at random...

- $P(\text{Curly is the 3rd Stooge}) = \underline{\hspace{2cm}}$
- $P(\text{Jules White directs}) = \underline{\hspace{2cm}}$
- $P(\text{Curly \& Jules White directs}) = \underline{\hspace{2cm}}$
- $P(\text{Curly or Jules White directs}) = \underline{\hspace{2cm}}$ Use formula:
- $P(\text{Curly \& Shemp}) = \underline{\hspace{2cm}}$
- $P(\text{Curly or Shemp or Other director}) = \underline{\hspace{2cm}}$ Use formula:
- $P(\text{Not Joe}) = \underline{\hspace{2cm}}$
- $P(\text{Del Lord was the director}) = \underline{\hspace{2cm}}$
- Given you know Joe is NOT in the movie, what's the probability that Del Lord was the director? $\underline{\hspace{2cm}}$
- Given you know Joe IS in the movie, what's the probability that Del Lord was the director? $\underline{\hspace{2cm}}$

Hmm... it looks like $P(\text{Del Lord is the director})$ changes when we're given more information. We'll get back to this concept (**conditional probability**) later in this activity.

Hopefully, reading probabilities from a contingency table seems easy. Can we construct contingency tables for scenarios with more than 2 variables? If so, how do we create a 3-, 4-, or higher-dimensional table? Let's see...

4. Of 565 St. Ambrose freshmen in 2013, 129 did not return to St. Ambrose for their second year. We can create a table to display the number of returning and non-returning students by gender, race, and athletic status. Note that this table will display 4-dimensions: race, gender, athletic status, and whether or not a student returned.

		Did not return			Did return		
		Male	Female		Male	Female	
Not athlete	Race unknown	7	6	13	11	24	35
	White	20	34	54	60	159	219
	Minority	5	13	18	13	17	30
		32	53	85	84	200	284
Student athlete	Race unknown	0	1	1	14	4	18
	White	19	7	26	62	52	114
	Minority	15	2	17	16	4	20
		34	10	44	92	60	152

As you can see, I constructed the rows to represent 2 dimensions: athletes and race. The columns show both retention and gender. In the red font, I included some subtotals.

From this 4-way contingency table, calculate the following:

- $P(\text{white, male student athlete who returns to St. Ambrose}) = P(\text{white \& male \& athlete \& return}) =$ _____
- $P(\text{athlete or minority student}) =$ _____
- $P(\text{non-white male student}) = P(\text{white}' \& \text{male}) =$ _____
- $P(\text{returned to St. Ambrose}) =$ _____
- Given you know a student is an athlete, what's the probability he/she returned to St. Ambrose? _____
- Given you know a student is NOT an athlete, what's the probability he/she returned to St. Ambrose? _____

5. Suppose 80% of students applying to St. Ambrose are accepted.

If we randomly select a student who applies to SAU, what's our best estimate of the probability that the student is accepted?

Suppose I tell you the student earned straight As in high school and scored a perfect 36 on the ACT. Would that change your estimated probability?

6. The table to the right lists the outcomes from rolling two 6-sided dice.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

a) What's the probability the first die shows a 6?

b) Suppose you know the sum is 7. What's the probability the first die shows a 6?

c) Suppose you know the sum is 5. What's the probability the first die shows a 6? Why does this answer differ?

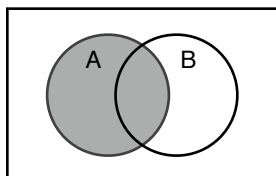
d) Suppose you know the sum is 5. What's the probability the first die shows a 2?

The **conditional probability** of an event is the revised probability we estimate when we have additional information

If we are given (or we assume) information about event B, then the revised probability of event A is written as:

$$P(A|B) = \text{the probability of A given B}$$

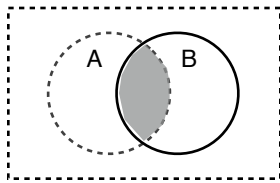
Suppose we're interested in the probability of event A occurring. In a Venn Diagram, we would display this as:



An estimate of this probability, $P(A) = \frac{\text{\# of outcomes in A}}{\text{\# of possible outcomes in the sample space}}$

Now suppose we know event B has happened. Given event B, what's the conditional probability of event A?

Since we already know event B has happened, that limits our sample space (the total number of possible outcomes).



An estimate of this probability, $P(A | B) = \frac{\text{\# of outcomes in A and B}}{\text{\# of possible outcomes in event B}}$

The **conditional probability** of event A, given event B has occurred, is estimated by: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Note that we still do not have a formula to calculate the numerator of this formula

7. In Europe, 88% of all households have a television. 51% of all households have a television and a VCR. What's the probability that a household with a television has a VCR? Use the table to verify your calculations.

	TV	No TV	Total
VCR			
No VCR			
Total			100

8. According to flightstats.com, 86.55% of flights out of Salt Lake City International Airport departed on-time in 2011. 72.68% of flights departed on-time and arrived on-time. Given a flight departs on-time, what's the probability it will arrive on-time?

	Arrive on	Arrive off	Total
Depart on			
Depart off			
Total			100

The **conditional probability** of event A, given event B has occurred, is estimated by: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

We can manipulate this equation to get the **General Multiplication Rule**: $P(A \cap B) = P(A|B)P(B)$

We can also note that: $P(A \cap B) = P(A|B)P(B) = P(B \cap A) = P(B|A)P(A)$

9. A new test is developed to detect a disease that inflicts 10% of the population. The test has a 97% chance of correctly identifying the disease in an infected individual. Suppose you take this test and you do not know if you are infected. What's the probability that you are infected and the test will correctly detect your infection?

		Total
Total		100

10. One hundred adults were surveyed to determine their opinions regarding loan debt burdens of college students. The proportion of responses, separated by adults with and without children in college, are displayed below:

	Too much loan debt	About right	Too little loan debt	Total
Has child in college	0.2	0.09	0.01	0.3
Does not have child in college	0.41	0.21	0.08	0.7
Total	0.61	0.3	0.09	1

Estimate the following probabilities:

a) $P(\text{does not have child in college} \mid \text{too little loan debt}) =$ _____

b) $P(\text{too little loan debt} \mid \text{does not have child in college}) =$ _____

c) $P(\text{does not have child in college AND too little loan debt}) =$ _____

d) $P(\text{does not have child in college OR too little loan debt}) =$ _____

11. From a Life Table, one finds that 89.835% of females can expect to live to age 60, while 57.062% can expect to live to age 80. Given that a woman has reached the age of 60, what's the probability that she will live to age 80? (Hint: Start by pretending you have a population of 100,000 women and determine how many are expected to live to each age).

		Total
Total		100