Bayesian Inference Name: _____

In activity #4, we introduced Bayes' Theorem: $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$

and used the Law of Total Probability to calculate the denominator: $P(A) = P(A|B)P(B) + P(A|B')P(B')$

In this assignment, you'll learn how to employ some basic Bayesian methods for inference.

Scenario: I walk into class one day and offer you a chance to play a game. In my hands, I have two dice: one 6-sided; the other 12-sided. I will keep one die in each hand (so we'll have one die in my left hand, DIE L, and one in my right hand, DIE R.

I will let you choose a hand (left or right) and I will roll the die. If the die shows a number greater than or equal to 4, you win and I give you $1. If the die shows a 1, 2, or 3, I get to keep the $1.

We will play this multiple times. I will not swap the sides the dice are on at any point. Each time the die rolls less than 4, you will lose $1 (so there is a cost associated with playing too many times).

When you think you have enough information, you will be allowed to guess which hand holds the 12-sided die. If you choose correctly, I will give you $100. Choose incorrectly and you will get nothing

1. What's the probability of getting a number ≥4 when rolling a 6-sided die? How about a 12-sided die?

$P(\text{roll} \geq 4 \mid \text{6-sided die}) =$ _____ $P(\text{roll} \geq 4 \mid \text{12-sided die}) =$ _____

2. The following table shows the possible outcomes from this game. Note that there is a cost associated with sampling too much data (playing too many rounds): you could lose money.

		Truth	
		12-sided die in left hand	12-sided die in right hand
Your decision	Pick left hand (L)	You win at least $100	You lose
	Pick right hand (R)	You lose	You win at least $100

3. You have no idea if I've chosen my right or left hand to hold the 12-sided die. So before we play this game, what are the probabilities associated with the following hypotheses?

H_1: The left hand holds the 12-sided die. $P(H_1)$ = _____

H_2: The right hand holds the 12-sided die $P(H_2)$ = _____

These are your **prior probabilities** for the two competing hypotheses. These priors represent what you believe before seeing any data.

4. I played this game with someone and got the following results. If you saw these results, what decision would you make? Is the 12-sided die in the right hand or the left hand? Briefly explain your decision.

Round	Choice (Left or Right)	Result (win or loss)
1	L	Loss
2	L	Win
3	L	Win
4	R	Loss
5	R	Loss
6	R	Win
7	R	Loss
8	L	Win
9	L	Loss
10	L	Win
11	R	Loss
12	R	Win

I believe the 12-sided die is in the _____ hand.

Brief explanation:

5. What's the probability (based on the outcome of round #1 in the above table) that the left hand holds the 12-sided die? To help get you started, notice that you want to calculate: P(left hand | lost in round #1).

We know that: **P(lose | left hand holds the 12-sided die) = 0.25** (from your answer to question #1)
P(lose | left hand holds the 6-sided die) = 0.50 (from your answer to question #1)
P(left hand holds the 12-sided die) = 0.50 (as you answered in question #3)

We can use Bayes' Theorem:

$$P\left(\text{left hand holds 12-sided die} \middle| \text{lost 1st round}\right) = \frac{P\left(\text{lost 1st round} \mid \text{left hand holds 12-sided die}\right) P(\text{left hand holds 12-sided die})}{P\left(\text{lost 1st round}\right)}$$

and the Law of Total Probability to get the denominator:

$$= \frac{P\left(\text{lost 1st round} \mid \text{left hand holds 12-sided die}\right) P(\text{left hand holds 12-sided die})}{\text{P}\left(\text{lost round} \mid \text{left hand holds 12-sided}\right)\text{P(left hand holds 12-sided)}+\text{P}\left(\text{lost round} \mid \text{left hand holds 6-sided}\right)\text{P(left hand holds 6-sided)}}$$

Go ahead and try to calculate this. Write your answer here: = _____.

6. In the previous question, you should have calculated a probability of 0.125 / 0.375 = 0.333. We call this the **posterior probability** (a probability that has been updated with results from an experiment).

   Prior probability: P(left hand holds the 12-sided die) = 0.50 (your answer to question #3)
   Posterior probability: P(left hand holds the 12-sided die | lost round 1) = 0.333 (answer to #5)

   Please make sure you can calculate this posterior probability correctly before continuing. If you need help, let me know.

   Notice that we can think of a posterior probability as: P(hypothesis | data). We now believe, based on results from round #1, that there is a 1/3 chance the left hand holds the 12-sided die.

   Let's now look at the results from round #2: <u>You chose the left hand and won the round</u>.

   Use this information to update the posterior probability. To do this, note that we now believe:

   P(left hand holds the 12-sided die) = 0.333
   P(win | left hand holds the 12-sided die) = 0.75
   P(win | left hand holds the 6-sided die) = 0.50

   We can now calculate:

$$P\left(\text{left hand holds 12-sided die} \middle| \text{won round}\right) = \frac{P\left(\text{won round} \mid \text{left hand holds 12-sided die}\right) P(\text{left hand holds 12-sided die})}{P\left(\text{won round}\right)}$$

$$= \frac{P\left(\text{won round} \mid \text{left hand holds 12-sided die}\right) P(\text{left hand holds 12-sided die})}{\text{P}\left(\text{won round} \mid \text{left hand holds 12-sided}\right)\text{P(left hand holds 12-sided)} + \text{P}\left(\text{won round} \mid \text{left hand holds 6-sided}\right)\text{P(left hand holds 6-sided)}}$$

   Show your calculations below and write your answer here: = _____

7. You should have calculated an updated posterior probability of 0.42857. You now believe there is a 42.857% chance the left hand holds 12-sided die.

   We'll do this one more time by hand. In round #3: <u>You chose the left hand and won the round</u>.

   Calculate: $P\left(\text{left hand holds 12-sided die} \middle| \text{won round}\right)$ = _____

8. Let's speed things up using a calculator. If you go to the following website, you'll find a Bayes' Theorem calculator: http://statpages.org/bayes.html

   Carefully read the "instructions and simple example." I'll admit that I didn't read the instructions at first and screwed up. Don't be like me.

   To see that this calculator gives correct answers, let's verify our answers to questions #5, #6, and #7 on this assignment.

   To verify the answer to question #5, you'll need to enter information like this:

Hypotheses:	12sided	6sided	Hyp 3	Hyp 4	Hyp 5
Prior Probabilities:	.50	.50	0	0	0
☐ WIN	.75	.50	0	0	0
☑ LOSS	.25	.50	0	0	0
☐ Outcome 3:	0	0	0	0	0
☐ Outcome 4:	0	0	0	0	0
☐ Outcome 5:	0	0	0	0	0
Revised Prob:	0	0	0	0	0

   The column names of "12sided" and "6sided" refer to our hypothesis about the hand we choose each round. In this case, we have a prior belief that there is a 50% chance the left hand holds the 12-sided die.

   Click COMPUTE and the answer to question #5 will appear at the bottom of the first column. Because we want to use these results to update our prior probabilities, go ahead and click REVISED-TO-PRIOR. This will put the posterior probabilities up in the prior probabilities column.

   Since we won round #2, you now need to click the WIN checkbox. Then click COMPUTE to see the answer to question #6.

Hypotheses:	12sided	6sided	Hyp 3	Hyp 4	Hyp 5
Prior Probabilities:	0.333	0.667	0.000	0.000	0.000
☑ WIN	.75	.50	0	0	0
☐ LOSS	.25	.50	0	0	0
☐ Outcome 3:	0	0	0	0	0
☐ Outcome 4:	0	0	0	0	0
☐ Outcome 5:	0	0	0	0	0
Revised Prob:	0.428	0.572	0.000	0.000	0.000

   Click REVISED-TO-PRIOR to update your prior probabilities once again.

   Since we also won round #4, you can simply click COMPUTE to see the answer to question #7. Click REVISED-TO-PRIOR to update your prior probabilities once again.

9. I've recorded our updated probabilities in the right column of the following table. Use the online calculator to estimate the updated probabilities for rounds 4-12 and record them in the table.

Round	Choice (Left or Right)	Result (win or loss)	P(left hand holds 12-sided die)	P(right hand holds 12-sided die)
(prior)			0.500	0.500
1	L	Loss	0.333	0.667
2	L	Win	0.428	0.572
3	L	Win	0.529	0.471
4	R	Loss		
5	R	Loss		
6	R	Win		
7	R	Loss		
8	L	Win		
9	L	Loss		
10	L	Win		
11	R	Loss		
12	R	Win		

Uh oh... I just noticed something. In rounds 1-3, you chose the LEFT hand and we used that as the basis for our calculations. In round 4, you chose the RIGHT hand and lost the round. Will we need to change our calculations?

Let's see. In choosing the right hand, we're interested in P(right hand holds 12-sided die | lost round 4).

$$P\left(\text{right hand holds 12-sided die}\middle|\text{lost round}\right)=\frac{P\left(\text{lost round}\mid\text{right hand holds 12-sided die}\right)P(\text{right hand holds 12-sided die})}{P\left(\text{lost round}\right)}$$

$$=\frac{P\left(\text{lost round}\mid\text{right hand holds 12-sided die}\right)P(\text{right hand holds 12-sided die})}{\text{P}\left(\text{lost round}\mid\text{right hand holds 12-sided}\right)\text{P(right hand holds 12-sided)}+\text{P}\left(\text{lost round}\mid\text{right hand holds 6-sided}\right)\text{P(right hand holds 6-sided)}}$$

We know: P(right hand holds 12-sided die) = 0.471 (the complement of our answer to question #7)
P(loss | right hand holds 12-sided die) = 0.25
P(loss | right hand holds 6-sided die) = 0.50

We can then calculate:

$$P\left(\text{right hand holds 12-sided die}\middle|\text{lost round}\right)=\frac{(0.25)(0.471)}{(0.25)(0.471)+(0.50)(0.529)}=\frac{0.11775}{0.2645}=0.308$$

10. Now that we know the updated probability, let's verify that our online calculator gives us that answer. Since we're choosing the RIGHT hand in round #4, we need to swap our prior probabilities in the calculator. We need to tell the calculator that we only believe there is a 47.1% chance the hand we chose (RIGHT) holds the 12-sided die.

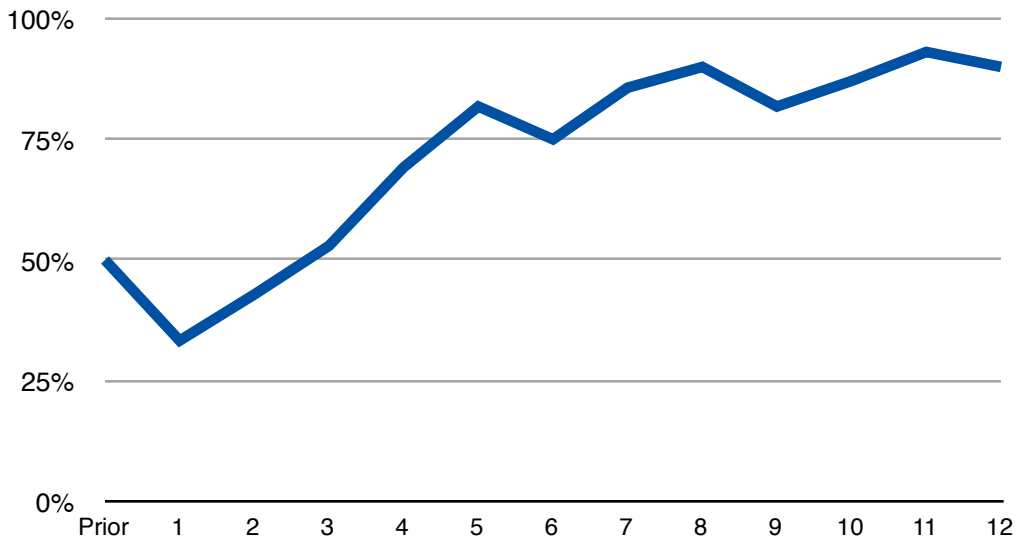
    So we need to enter the following:

Hypotheses:	12sided	6sided	Hyp 3	Hyp 4	Hyp 5
Prior Probabilities:	0.471	0.529	0.000	0.000	0.000
☐ WIN	0.75	0.5	0	0	0
☑ LOSS	0.25	0.5	0	0	0
☐ Outcome 3:	0	0	0	0	0
☐ Outcome 4:	0	0	0	0	0
☐ Outcome 5:	0	0	0	0	0
Revised Prob:	0.308	0.692	0.000	0.000	0.000

    Click REVISED-TO-PRIOR to update your prior probabilities and then complete the table on the previous page.

11. Based on the 12-rounds of this game you played, which hand do you believe holds the 12-sided die? How confident are you in this answer? (The results of this game were simulated on a program that was informed the LEFT hand held the 12-sided die).

The following graph displays our updated probabilities for P(left hand holds the 12-sided die):



Recap: Bayesian information takes advantage of prior information. No matter what we're investigating, we could take advantage of information from prior studies or physical models.

This process naturally integrates data as we collect it and updates our prior beliefs.

As we'll learn, Bayesian methods avoid counterintuitive notions like p-values (the probability of observing something as or more extreme than what we observed | the null hypothesis is true). Instead, Bayesian methods base decisions on posterior probabilities (the probability that the null hypothesis is true | our observed data). This is something we'll explore throughout the semester.

Bayesian methods depend on good prior probabilities, which can be difficult to obtain.

Scenario: The American Cancer Society estimates that about 1.7% of women have breast cancer.
(Source: http://www.cancer.org/cancer/cancerbasics/cancer-prevalence)

The Susan G. Komen For the Cure Foundation reports that mammography correctly identifies about 78% of women who truly have breast cancer.
(Source: http://ww5.komen.org/BreastCancer/AccuracyofMammograms.html)

A 2003 article suggests that up to 10% of all mammograms are false positives.
(Source: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360940)

12. A woman undergoes a routine mammogram and finds a positive result (indicating she has breast cancer). Given this positive result, what's the probability this woman has breast cancer?

Prior probability that the woman has breast cancer = _____

Updated probability = P(breast cancer | positive result) = _____

13. Using this updated probability of breast cancer, calculate the probability that she has breast cancer given a second test yields a positive result.

Updated probability = P(breast cancer | a second positive result) = _____

Scenario: Suppose a student in this class send me an email containing the word "radom."

Having never seen the word "radom" before, I guess the student could have either:

- intended to write the word "radom"
- intended to write the word "random" but, due to a typo, forgot the "n"
- intended to write the word "radon" but, due to a typo, changed the "n" to an "m"

We're going to attempt to find the likelihood of the student doing each of those three things.

Source: Gelman, A., et. al (2013). Bayesian Data Analysis, 3rd edition. ISBN: 978-1439840955

14. Let's first take a look at the probability we will compute. Using Bayes' Theorem, we have:

$$P(\text{intended word} \mid \text{"radom"}) = \frac{P(\text{"radom"} \mid \text{intended word})P(\text{intended word})}{P(\text{"radom"})}$$

Take a look at the denominator. How can we calculate the probability that the student typed "radom?" In this scenario, there are 3 different ways the student could have written that word. He could have written it (1) on purpose, (2) mistakenly instead of "random," or (3) mistakenly instead of "radon." Using the Law of Total Probability, we can expand the denominator of this formula and rewrite Bayes' formula:

$$\frac{P(\text{"radom"} \mid \text{intended word})P(\text{intended word})}{P(\text{"radom"} \mid \text{intended radom})P(\text{intended radom}) + P(\text{"radom"} \mid \text{intended random})P(\text{intended random}) + P(\text{"radom"} \mid \text{intended radon})P(\text{intended radon})}$$

We can then use this formula 3 times to estimate the likelihood of each of our 3 scenarios.

But how can we calculate the numerator of this formula? How could we possibly know the probability of a student typing radom, random, or radon in an email? Researchers at Google found the relative frequencies of these words in a large database of emails. They found the following probabilities:

P(radom) = 0.000000312
P(random) = 0.0000760
P(radon) = 0.00000605

Wait a second! How could the probability of "radom" be so close to the probabilities of the other two words? To figure this out, a quick look at Wikipedia showed me that "radom" is a medium-sized city in Poland (home to the largest air show in Poland) and the word is also an unofficial name for a semiautomatic 9mm Para pistol of Polish design. It looks like some people really do write "radom" intentionally in emails.

To use Bayes' formula, we also need to know P("radom" | intended word) for each of our 3 intended words. Thankfully, Google comes to our rescue once again. Researchers at Google provide the following conditional probabilities (based on their spelling and typing errors model):

P("radom" | radom) = 0.975
P("radom" | random) = 0.00193
P("radom" | radon) = 0.000143

To me, these values seem reasonable. There's a 97.5% that the word was typed correctly, a 0.2% chance the word "random" was misspelled, and a 0.01% chance the word "radon" was misspelled. I'll trust that these are good estimates of these conditional probabilities.

Using the given information and Bayes' formula, calculate each of the following:

$P(\text{intended "radom" | "radom"}) =$ _____

$P(\text{intended "random" | "radom"}) =$ _____

$P(\text{intended "radon" | "radom"}) =$ _____

Based on your calculations, which word did the student most likely intend to type in the email?

The student most likely intended to type the word: _____

15. If you did the calculations correctly, you should have found the likelihood of the student intending to write "radom" was more than twice as much as the likelihood that the student intended to write either "random" or "radon." Do you really believe the most likely scenario is one in which the student (a student from this class) intended to write "radom"? If not, explain what went wrong. Do you have reason to not believe in some of the given information supplied in this scenario?