Problem Set 5: Probability II

1. Suppose 80% of people like peanut butter, 89% like jelly, and 78% like both. Given that a randomly sampled person likes peanut butter, what's the probability that he also likes jelly?

2. The American Cancer Society estimates that about 1.7% of women have breast cancer. Susan G. Komen For The Cure Foundation states that mammography correctly identifies about 78% of women who truly have breast cancer. An article published in 2003 suggests that up to 10% of all mammograms result in false positives for patients who do not have cancer.

   When a patient goes through breast cancer screening there are two competing claims: patient had cancer and patient doesn't have cancer. Assuming these approximations are correct and if a mammogram yields a positive result, what is the probability that patient actually has cancer?

3. A common epidemiological model for the spread of diseases is the SIR model, where the population is partitioned into three groups: Susceptible, Infected, and Recovered. This is a reasonable model for diseases like chickenpox where a single infection usually provides immunity to subsequent infections. Sometimes these diseases can also be difficult to detect.

   Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. (Note: In this case accurate means returning a negative result for susceptible and recovered individuals and a positive result for infected individuals).

   If the individual has tested positive, what is the probability that they are actually infected?

4. A Pew Research poll asked 1,306 Americans "From what you've read and heard, is there solid evidence that the average temperature on earth has been getting warmer over the past few decades, or not?". The table below shows the distribution of responses by party and ideology, where the counts have been replaced with relative frequencies.

		Response			
		Earth is warming	Not warming	Don't Know Refuse	Total
Party and Ideology	Conservative Republican	0.11	0.20	0.02	0.33
	Mod/Lib Republican	0.06	0.06	0.01	0.13
	Mod/Cons Democrat	0.25	0.07	0.02	0.34
	Liberal Democrat	0.18	0.01	0.01	0.20
	Total	0.60	0.34	0.06	1.00

   a. Are believing that the earth is warming and being a liberal Democrat distjoint / mutually exclusive?
   b. What is the probability that a randomly chosen respondent believes the earth is warming or is a liberal Democrat?
   c. What is the probability that a randomly chosen respondent believes the earth is warming given that he is a liberal Democrat?
   d. What is the probability that a randomly chosen respondent believes the earth is warming given that he is a conservative Republican?
   e. Does it appear that whether or not a respondent believes the earth is warming is independent of their party and ideology? Explain your reasoning.
   f. What is the probability that a randomly chosen respondent is a moderate/liberal Republican given that he does not believe that the earth is warming?

5. Cost of breakfast Maria gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of 30 cents (\$0.30), the average price of a muffin is \$2.50 with a standard deviation of 15 cents, and the two prices are independent of each other.

   a. What is the mean and standard deviation of the amount she spends on breakfast daily?
   b. What is the mean and standard deviation of the amount she spends on breakfast weekly (7~days)?

6. An airline charges the following baggage fees: \$25 for the first bag and \$35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.

   a. Define a random variable that describes the baggage fee revenue for a single passenger, with the possible values that it can take along with their probabilities. The compute the average revenue per passenger, and compute the corresponding standard deviation.
   b. About how much revenue should the airline expect for a flight of 120 passengers? With what standard deviation? Note any assumptions you make and if you think they are justified.

7. Suppose we have independent observations X_1 and X_2 from a distribution with mean μ and standard deviation σ. What is the variance of the mean of the two values: $\frac{X_1+X_2}{2}$?

8. What characteristics should a random process have to be well-modeled by the Binomial distribution?

9. Let $Y \sim \text{Bern}(p = 3)$. What is $E(Y)$? $Var(Y)$?

10. Let $Z \sim \text{Bin}(n = 10, p = .9)$. Find the following properties of Z.

   a. What are the different values that Z can take?
   b. $E(Z)$
   c. $Var(Z)$
   d. $P(Z = 10)$

11. Think back to the boba tea taste test lab, either your protocol or one of the ones we described in class. Describe one component this is well-modeled by the Bernoulli destribution and a second that is well-modeled by the Binomial distribution. Include in your description why these are sensible fits. Depending on the how you set up your own protocol, yours may not be appropriate for these distributions, so feel free to draw from a separate protocol you *could* perform or one of the protocols that we discussed in class.

12. **Optional Challenge Problem**: A *chord* of a circle is a straight line segment whose endpoints both lie on the circle. For a fixed circle, what is the probability that the length of a randomly drawn chord will exceed that circle's radius?