Statistics Problem Set

(Send corrections to cbruni@uwaterloo.ca)

(i) Compute some of the following probabilities

1. Probability of drawing a green ball from a bag with 4 brown, 5 green, 6 blue and 7 yellow balls.
2. Probability of not drawing a green ball from previous bag.
3. The probability of drawing a brown ball from the previous bag.
4. The probability of drawing a brown ball given that you drew a ball with five letters in its name.
5. Assuming that having children is random, what is the probability that a person who has two children has two girls? What is the probability that this does not occur?
6. At a bar you get talking with a person and find out they have two children. From asking questions, you find out that the person has a girl in ballet. Determine the probability that the other person is a boy.

Solution:

1. $5/(4+5+6+7) = 5/22$.
2. $1 - 5/22 = 17/22$
3. $4/22$
4. $4/(5+4) = 4/9$ (Note: This is $(4/22)/(9/22)$ by Bayes Theorem)
5. $1/4$ and $1 - 1/4 = 3/4$
6. $2/3$. There are four total possibilities: BB, BG, GB, GG. We remove one of them leaving BG, GB, and GG. Thus the other child is twice as likely to be a boy.

(ii) Birthday Paradox. Throughout for simplicity, assume there are 365 days in every year.

1. Suppose the probability of an event A occurring is p. What is the probably of A^c, the event not A, occurring?
2. Given two people, what is the probability that they do not share a common birthday? What is the probability that they share a common birthday?
3. Given three people, what is the probability that no two share a common birthday?
4. How many people are needed to have a 50% chance that two people share a common birthday?
5. How many people are needed to have a 100% chance that they share a common birthday?
6. Suppose you have a game collection consisting of 500 games. Each night you randomly choose one to play. You do this for 50 nights. What is the probability that you played a different game each night?

Solution:

1. $1 - p$
2. Do not share a birthday: 364/365. Do share a common birthday: 1/365.
3. No pair shares a common birthday: $364 \cdot 363/(365)^2$. Do share a common birthday $1 - 364 \cdot 363/(365)^2$.
4. Want to find a value of n such that $364 \cdot 363 \cdot ... \cdot (364 - n + 1)/(365)^{n-1} \approx 0.5$. This is approximately $n = 23$ which is surprising giving the next result.
5. We require 366 people in order to guarantee a collision.
6. $499/500 \cdot 498/500 \cdot ... \cdot 450/500 \approx 7.13\%$

(iii) Another Disease Example. A new population is suffering from a disease epidemic. Empirical studies have shown that 2% of the population suffer from the disease. There is a test that can determine whether a person has the disease that has a 2.5% false positive rate and a 2.5% false negative rate. (So in some sense, this test is also 95% accurate like in the lesson).

1. Write down an expectation table (like in the lesson) for a population of 10000 people.
2. What is the probability that a person does not have the disease?
3. What is the probability that a person tests positive given that they have the disease?
4. What is the probability that a person has the disease given that they test positive?
5. What is the probability that a person has the disease given that they test negative?
6. Solve the above using Bayes Theorem and conditional probability notation.
7. Compare this with the example from class. Which test would you use? Why?

Solution:

1.

	Has Disease	Does Not Have Disease
Positive	195	245
Negative	5	9555

2. There is a 98% chance that the person doesn't have the disease.
3. The probability is $195/(195 + 5) = 195/200 = 0.975$.
4. There is a $195/(245+195) = 195/440 = 39/88 \approx 44\%$ chance of having the disease
5. There is a $5/(9555+5) = 5/9560 \approx 0.05\%$ chance of having the disease given that they tested negative.

6. Let A be the event of having the disease and let B be the event of testing positive. Then $P(A) = 0.02$, $P(B^c \mid A) = 0.025$ and $P(B \mid A^c) = 0.025$. Now, the probability that a person does not have the disease is $P(A^c) = 1 - 0.02 = 0.98$. The probability that a person tests positive given that they have the disease is by definition:

$$\begin{aligned}
P(B \mid A) &= P(A \text{ and } B)/P(A) \\
&= (P(A) - P(A \text{ and } B^c))/P(A) \\
&= 1 - P(B^c \mid A) \\
&= 1 - 0.025 \\
&= 0.975.
\end{aligned}$$

Now, the probability that a person has the disease given that they test positive for the disease is $P(A \mid B)$ which is by Bayes Theorem $P(B \mid A)P(A)/P(B)$. First we find $P(B)$:

$$\begin{aligned}
P(B) &= P(B \text{ and } A) + P(B \text{ and } A^c) \\
&= P(B \mid A)P(A) + P(B \mid A^c)P(A^c) \\
&= (0.975)(0.02) + (0.025)(0.98) \\
&= 0.044
\end{aligned}$$

Thus,

$$\begin{aligned}
P(A \mid B) &= P(B \mid A)P(A)/P(B) \\
&= (0.975)(0.02)/(0.0538) \\
&= 39/88 \approx 0.443...
\end{aligned}$$

Lastly, $P(A \mid B^c)$ can be computed via Bayes Theorem:

$$\begin{aligned}
P(A \mid B^c) &= P(B^c \mid A)P(A)/P(B^c) \\
&= P(B^c \mid A)P(A)/(1 - P(B)) \\
&= (0.025)(0.02)/(1 - 0.044) \\
&= 5/9560 \approx 0.0005.
\end{aligned}$$

7. Answers may vary. I would use the test from class because of the lack of a false negative.

(iv) Consider two Chess Champions Bobby and Gary. Suppose they competed in two consecutive years (with a varying number of games in each year per player) and they have the following winning percentages: Who is the better player? If you can determine

	Year 1	Year 2
Bobby	95%	89%
Gary	91%	88%

this, justify. If you can't determine, give a situation where Bobby is the better player and give a situation where Gary is the better player.

Solution: Notice that you cannot determine the better player. As an example, suppose that Bobby won 19/20 games in year 1 and 89/100 in year 2 whereas Gary won 910/1000 and 88/100 over the two years. Then, Bobby has an overall average of 90% and Gary has an overall average of 90.7%.