Homework 1

Jared S. Murray
Statistics
McCombs School of Business

Instructions

Submit your assignment at the **beginning** of class on **Monday Feb 4**. Please make sure that the pages of your homework assignment are in the correct order and securely stapled together. You may (but are not required to) work in groups of up to four students. Each group should hand in one copy of the homework.

Grading

A small number of problems will be graded for correctness, and the remainder graded for completeness. A complete response answers the question posed and also shows your work. This means showing the steps of a mathematical calculation, or including R output that justifies your conclusions. For questions that are not just calculations (e.g., more than computing an expected value from a table) you should answer in complete, concise sentences. Detailed solutions will be available – you should always check your work against these solutions.

Problem 1: Senate Probabilities (Election 2014)

According to some analyses, the control of the Senate in the upcoming elections will be determined by the races in 3 states: Arkansas, Louisiana and North Carolina, where 3 Democratic incumbents face very competitive opponents. Based on predictions by experts at the NY Times, the Republicans have the following chances of wining each of these races: Arkansas 67%, Louisiana 61% and North Carolina 52%.

1. To win control of the senate, Republicans need to win at least two of these races. Based only on the numbers above, give an estimate the probability of the Republicans taking control of the senate. State any assumptions you need to make in order to compute your estimate using only the information given above.

2. The betting markets are currently trading at a 0.80 probability for the Republicans to control the senate. How does your answer from the question above compare to this number? Can you explain why you are seeing a difference?

Problem 2

Imagine that you've graduated with your BBA and taken a spectacular job that involves significant travel. You'll be traveling routinely to Boston, Orlando, Philadelphia and San Diego... You're trying to decide whether you want to fly US Airways or Delta, so you can start accumulating miles and status, and you also want to minimize the time you spend sitting in airports or missing connecting flights due to delays. Assume that each airline has an identical frequent flyer program. After a quick look online you find in the U.S. Bureau of Transportation Statistics the following table describing the delays of these two airlines:

	Delta	US Airways
Delayed	20%	22%
On Time	80%	78%

Is this enough information for you to make a decision? (Assume that this data only includes flights to the four cities of interest.) If not, can you describe a possible scenario in which choosing Delta **doesn't** make sense?

Problem 3

Suppose a person is randomly drawn from a large population and then tested for a disease.
Let $D = 1$ if the person has the disease and 0 otherwise.
Let $T = 1$ if the person tests positive and 0 otherwise.

Suppose
$P(D = 0) = .99$.
$P(T = 1 \mid D = 0) = .01$.
$P(T = 1 \mid D = 1) = .97$.

(a) Draw the tree diagram depicting the marginal of D and the conditional of $T \mid D$.

(b) Give the joint distribution of D and T in the two way table format.

(c) What is $P(D = 1 \mid T = 1)$?

Problem 4

Here's a simplified look at a spam filter algorithm:

We are getting a tremendous volume of mail referencing a *"Nigerian prince"* and our IT team has figured out that 20% of spam emails contain the term "Nigerian prince", while only 0.1% of legitimate emails contain the term "Nigerian prince" In addition, they estimate that half of our emails are spam.

1. What is the marginal probability of seeing *"Nigerian prince"* in a message? In other words, what is Pr("Nigerian prince")?

2. If the spam filter always classifies a message containing *"Nigerian prince"* as junk, how often will it make a mistake?
   In other words, what is the Pr(NOT junk mail|"Nigerian prince") ?

Problem 5

1. Compute the probability distribution and expected value of the number of heads in 3 tosses of a fair coin.

2. In class we used the following R code to estimate the distribution of the number of heads in 2 tosses of a fair coin, using 10,000 simulations. Modify the code to estimate the probability distribution of the number of heads in **3 tosses** of a fair coin using **15,000 simulations**. *Hint: you will need to modify the call to the* `sample` *function; type* `?sample` *into the R console to get help regarding these arguments if you need it.*

```
library(mosaic)
num.sim = 10000
num.heads.sample = do(num.sim) * {
  coinflips.result = sample(x = c(0, 1),
                            size = 2, replace = TRUE)
  sum(coinflips.result)
}
table(num.heads.sample)/num.sim
```

TO HAND IN: The results of the simulation (your estimated probabilities).