Homework Assignment
Section 1

Carlos M. Carvalho
Statistics

Problem 1

$X \sim N(5, 10)$ (Read X distributed Normal with mean 5 and var 10) Compute:

(i) Prob$(X > 5)$

$$Prob(X > 5) = Prob\left(Z > \frac{5-5}{\sqrt{10}}\right) = Prob(Z > 0) = 0.5 \quad \text{(or 50\%)}$$

You can do this without looking at the normal table or using your calculator... just remember that in the normal distribution, 50% of the probability is above the mean.

(ii) Prob$(X > 5 + 2 \times \sqrt{10})$

$$Prob(X > 5 + 2 \times \sqrt{10}) \approx 0.025 \quad \text{(or 2.5\%)}$$

Rule of thumb: In the normal distribution, 95% of the probability is between -2 and 2 standard deviations.

(iii) Prob $(X = 8)$

$$Prob(X = 8) = 0$$

(iv) Express Prob$(-2 \leq X \leq 6)$ in terms of Z, the standard normal random variable.

$$Prob(-2 \leq X \leq 6) = Prob\left(\frac{-2-5}{\sqrt{10}} \leq Z \leq \frac{6-5}{\sqrt{10}}\right) = 0.61$$

(v) $Prob(X > 7)$

$$Prob(X > 7) = 1 - \text{NORMDIST}(7, 5, \sqrt{10}, 1) = 0.2635$$

Problem 2

A company can purchase raw material from either of two suppliers and is concerned about the amounts of impurity the material contains. A review of the records for each supplier indicates that the percentage impurity levels in consignments of the raw material follow normal distributions with the means and standard deviations given in the table below. The company is particularly anxious that the impurity level in a consignment not exceed 5% and want to purchase from the supplier more likely to meet that specification. Which supplier should be chosen?

	Mean	Standard Deviation
Supplier A	4.4	0.4
Supplier B	4.2	0.6

Let X_A represent the percentage of impurity level in a randomly chosen consignment of raw material from Supplier A. Therefore, $X_A \sim N\left(4.4, 0.4^2\right)$. Similarly, X_B represents the percentage of impurity level in a randomly chosen consignment of raw material from Supplier B, and, $X_B \sim \left(4.2, 0.6^2\right)$.

We need to compute $Prob(X_A > 5)$ and $Prob(X_B > 5)$.

$$Prob(X_A > 5) = Prob\left(Z > \frac{5-4.4}{0.4}\right) = Prob(Z > 1.5)$$

and

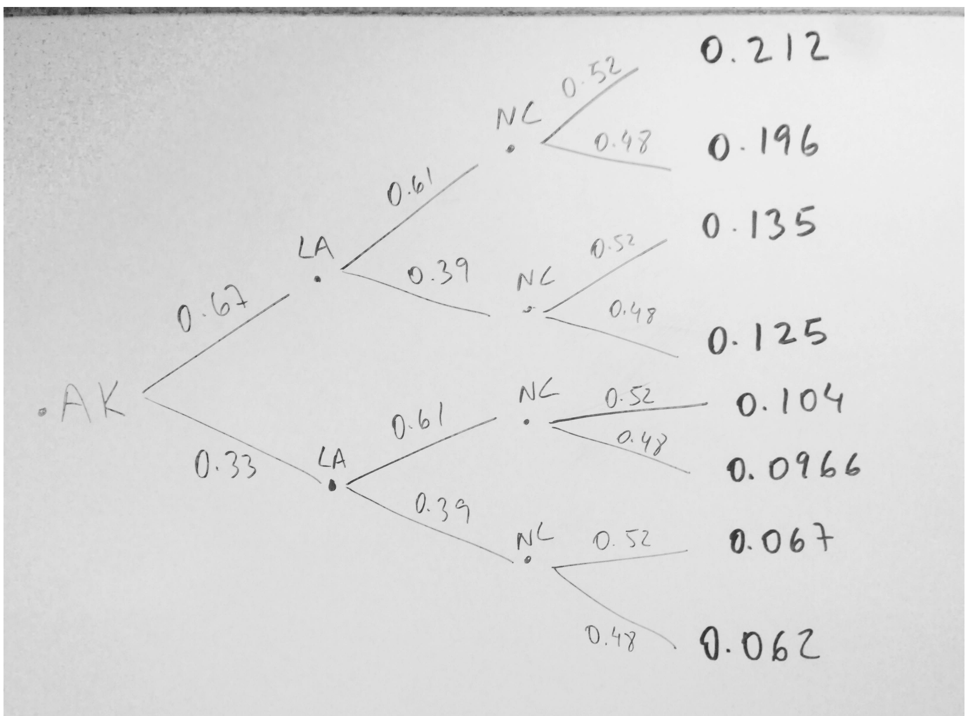
$$Prob(X_B > 5) = Prob\left(Z > \frac{5-4.2}{0.6}\right) = Prob(Z > 1.33)$$

We therefore conclude that Supplier A is better and should be chosen.

Problem 3: Senate Probabilities (Election 2014)

According to some analysis 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 probabilities 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 on the numbers above, what is the probability of the Republicans taking control of the senate?



So, republicans need to win at least 2 seats... therefore the probability they take over the senate equals 0.212 + 0.196 + 0.135 + 0.104 = 0.647

2. The betting markets are currently trading at a 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? (*Hint:* Did you have to make any assumption to answer the first questions)?

   We assumed independence across the 3 states. My guess is that those 3 events are not really independent of each other... for example, given that LA went to republicans on election night that should impact the conditional probability of NC going republican, right?

Problem 4

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 diagram depicting the marginal of D and the conditional of $T \mid D$. (you know, the one that branches as you go left to right).

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

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

(b)

```
       T0     T1
D0 0.9801 0.0099
D1 0.0003 0.0097
```

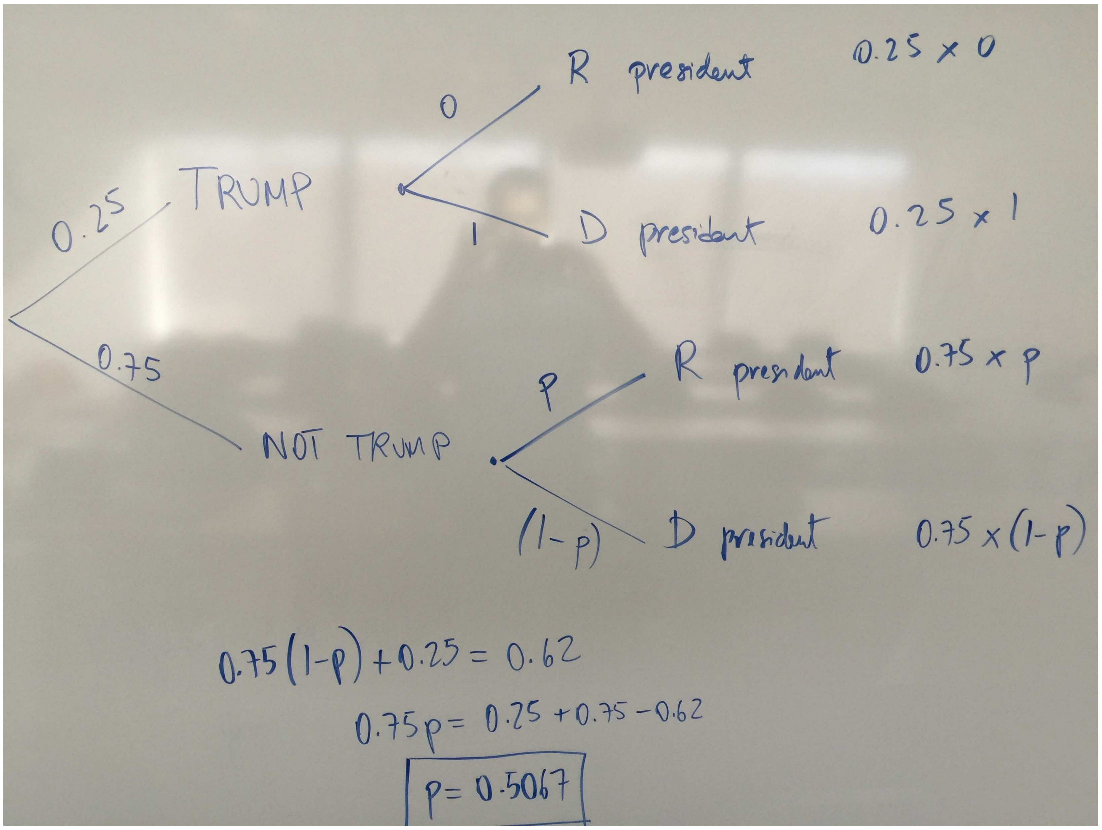
```
> .99*.99
[1] 0.9801
> .99*.01
[1] 0.0099
> .01*.03
[1] 3e-04
> .01*.97
[1] 0.0097
```

(c)
$P(D = 1 \mid T = 1) = P(D = 1, T = 1)/P(T = 1) = .0097/(.0097 + .0099) = 0.494898.$

Problem 5: Trump

Based on betting markets the probability of Donald Trump being the Republican nominee is 25%. The same markets have the probability that the next President will be a Democrat at 62%.

Assume that if Trump is the nominee he has no chance of becoming the President... so, if the nominee is someone NOT Donald Trump, what is the probability of a Republican becoming the President?



Problem 6

This problem is named after the host of the long running TV show *Let's make a deal*.
There has been a vigorous debate about what the correct answer is!!
A contestant must choose one of three closed doors.
There is a prize (say a car) behind one of the three doors.
Behind the other two doors, there is something worthless (traditionally a goat).
After the contestant chooses one of the three doors, Monty opens one of the other two, revealing a goat (never the car!!).
There are now two closed doors.
The contestant is asked whether he would like to switch from the door he intially chose, to the other closed door.

The contestant will get whatever is behind the door he has finally chosen.

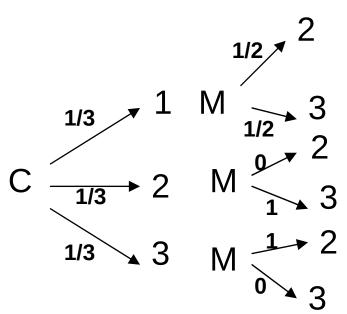
Should he switch?

Assume (I claim without loss of generality) that you initially select door 1.

The game is about to be played.

Up to the point where you must decide whether or not to switch,
there are two things about which we are uncertain:
C: the door the car is behind, C is 1, 2, or 3
M: which door Monte will open, M is 2 or 3 (given you have selected door 1).

We need the joint distribution of (C,M).
We will first write it in terms of the marginal for C and the conditional
for M given C since it is the most obvious in this form.



=>

		M 2	M 3
C	1	1/6	1/6
C	2	0	1/3
C	3	1/3	0

=> P(C=2|M=3)
= 1/3/((1/3)+(1/6))
= 2/3

Problem 7

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

We are worried about the term *"Nigerian general"* and our IT team has figured that $pr(\text{"Nigerian general"}|\text{junk mail}) = 0.20$ and $pr(\text{"Nigerian general"}|\text{NOT junk mail}) = 0.001$ In addition they figured that half of our emails is junk.

1. What is the marginal probability of seeing *"Nigerian general"* in a message? In other words, what is the $pr(\text{"Nigerian general"})$?

2. If the spam filter always classify a message containing *"Nigerian general"* as junk, how often will it make a mistake? In other words, what is the $pr(\text{NOT junk mail}|\text{"Nigerian general"})$?

Let's first build the joint distribution table:

	Junk	Not Junk
Nigerian general	0.2×0.5	0.001×0.5
Not NG	0.8×0.5	0.999×0.5

	Junk	Not Junk
Nigerian general	0.1	0.0005
Not NG	0.4	0.4995

1. Therefore, the $pr(\text{"Nigerian general"}) = 0.1 + 0.0005 = 0.1005$

2. $pr(\text{NOT junk mail}|\text{"Nigerian general"}) = \frac{pr(\text{NOT junk mail} \quad \text{and} \quad \text{"Nigerian general"})}{pr(\text{"Nigerian general"})}$

$pr(\text{NOT junk mail}|\text{"Nigerian general"}) = \frac{0.0005}{0.10005} = 0.005$

i.e., only 0.5% of the emails will be wrongly flagged as junk.

Problem 8

An oil company has purchased an option on land in Midland, TX. Preliminary geological studies have assigned the following probabilities of finding oil in the land:

$$Pr(\text{high quality oil}) = 0.5 \quad Pr(\text{medium quality oil}) = 0.2 \quad Pr(\text{NO oil}) = 0.3$$

After buying the option the company decided to perform a soil test. They found soil "type A". The probabilities of finding this particular type of soil are as follow:

$Pr(\text{soil} = \text{"type A"}|\text{high quality oil}) = 0.2$
$Pr(\text{soil= "type A"}|\text{medium quality oil}) = 0.8$
$Pr(\text{soil= "type A"}|\text{NO oil}) = 0.2$

1. Given the information from the soil test what is the probability the company will find oil in this land?
2. Before deciding to drill in the land the company has to perform a cost/benefit analysis of the project. They know it will cost $1,000,000 to drill and start operating a well. In addition, under current oil prices, they access that if oil is found (any kind) the revenue stream will be of $1,500,000. Should they exercise the option, ie, should they drill?

1. We are looking for the conditional probability of finding oil given we have already found soil type A. In other words, we want $\Pr(Oil|SoilTypeA)$.

The table below gives the join probabilities of Soil types and types of oil. We get to it by multiplying the marginal probabilities of oil with conditional probabilities of soil type given oil.

	High	Low	No Oil
Type A	0.1	0.16	0.06
Not A	0.4	0.04	0.24

So, we know type A was found therefore we only care about the first row of the table... we want to know how large in the SUM of the first two numbers in that row relative to the total...

$$\Pr(Oil|SoilTypeA) = \frac{\Pr(oil \quad \text{and} \quad typeA)}{\Pr(typeA)} = \frac{0.1 + 0.16}{0.1 + 0.16 + 0.06} = \frac{0.26}{0.32} = 0.8125$$

2. We need to calculate our expected payoff. If we drill we find oil with probability 0.81. So the expected payoff from that decision is

$$0.81 \times 1,500,00 = 1,215,000$$

therefore we should drill! The expected value of our profits is $215,000.