Math 151 - Probability Theory - Homework 3

your name here

Due: Friday, February 8, 2019, in class

```r
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=3, fig.width=5,
                      fig.align = "center")
```

[7] DeGroot, section 1.10 Suppose that a school band contains 10 students from the freshman class, 20 students from the sophomore class, 30 students from the junior class, and 40 students from the senior class. If 15 students are selected at random from the band, what is the probability that at least one student will be selected from each of the four classes? *Hint:* First determine the probability that at least one of the four classes will not be represented in the selection.

[2] DeGroot, section 2.1 If A and B are disjoint events and $P(B) > 0$, what is the value of $P(A|B)$?

[6] DeGroot, section 2.1 A box contains three cards. One card is red on both sides, one card is green on both sides, and one card is red on one side and green on the other. One card is selected from the box at random, and the color on one side is observed. If this side is green, what is the probability that the other side of the card is also green?

[12] DeGroot, section 2.1 For any three events A, B, and D, such that $P(D) > 0$, prove that

$$P(A \cup B|D) = P(A|D) + P(B|D) - P(AB|D)$$

[2] DeGroot, section 2.2 Assuming that A and B are independent events, prove that the events A^c and B^c are also independent.

[10] DeGroot, section 2.2 The probability that any child in a certain family will have blue eyes is 1/4, and this feature is inherited independently by different children in the family. If there are five children in the family and it is known that at least one of these children has blue eyes, what is the probability that at least three of the children have blue eyes?

[11] DeGroot, section 2.2 Consider the family with five children described in Exercise 10.

a. If it is known that the youngest child in the family has blue eyes, what is the probability that at least three of the children have blue eyes?
b. Explain why the answer in part (a) is different from the answer in Exercise 10.

[20] DeGroot, section 2.2 Suppose that $A_1, \dots A_k$ are a sequence of k independent events. Let $B_1, \dots B_k$ be another sequence of k events such that for each value of j, $(j = 1, \dots k)$, either $B_j = A_j$ or $B_j = A_j^c$. Prove that $B_1, \dots B_k$ are also independent events. *Hint:* Use an induction argument (See the "Proof by Induction Example" in the Notes section for an example.) based on the number of events B_j for which $B_j = A_j^c$.

[R1] Assume that every time you buy a box of Wheaties, you receive one of the pictures of the n players on the LA Dodgers. Over a period of time, you buy $m \geq n$ boxes of Wheaties. [From Introduction to Probability by Grinstead and Snell]

(a) Use Theorem 1.10.2 to show that the probability that you get all n pictures is

$$1 - \binom{n}{1}\left(\frac{n-1}{n}\right)^m + \binom{n}{2}\left(\frac{n-2}{n}\right)^m - \dots + (-1)^{n-1}\binom{n}{n-1}\left(\frac{1}{n}\right)^m.$$

Hint: consider the event that you do not get the k^{th} player's picture.

(b) Write a computer program to ESTIMATE this probability. Use the program to estimate, for given n,, the smallest value of m which will give probability greater than 0.5 of getting all n pictures. Consider

$n = 10, 50, 100$ and show that $m = n\log(n) + n\log(2)$ is a good (if slightly conservative) estimate for the number of boxes needed. (For a derivation of this estimate, see Feller(1968).)

```
n.box=10
n.samp = 20
n.reps=5
mycereal = sample(1:n.box, n.samp,replace=TRUE)
mycereal                  # remove this line after you know what it does
```

```
##  [1]  3  2  2  5  3  9  3  6  9  7  1  9  8  6  3  5  7 10  6  6
```

```
unique(mycereal)          # remove this line after you know what it does
```

```
## [1]  3  2  5  9  6  7  1  8 10
```

```
length(unique(mycereal))  # remove this line after you know what it does
```

```
## [1] 9
```

```
mycerprop = c()
for (m in n.box:n.samp){
mycercount = c()
  for (j in 1:n.reps){
   mycereal = sample(1:n.box,m, replace=TRUE)
   mycercount = c(mycercount, length(unique(mycereal)))
  }
  mycerprop = c(mycerprop, sum(mycercount==n.box)/n.reps)
}

mycercount  # remove this line later changes, note it gets written over every time m changes
```

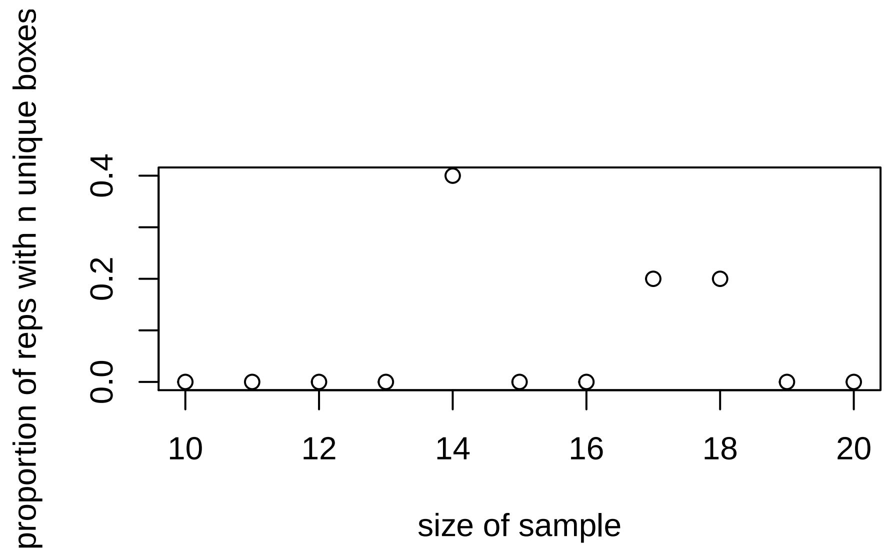
```
## [1] 8 9 8 8 8
```

```
mycerprop   # remove this line later, note it should be as long as m
```

```
##  [1] 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.2 0.2 0.0 0.0
```

After you have the simulation running, bump up the reps to get a good empirical estimate of the probability. A plot will give you an understanding of what is being estimated:

```
plot(n.box:n.samp, mycerprop, xlab="size of sample", ylab="proportion of reps with n unique boxes")
abline(h=0.5, col="red")
abline(v=n.box*log(n.box) + n.box*log(2), col="blue")
```



Write a few sentences reflecting on what you observe. Note that your sentences should come outside of the R chunk, but they should be within the R Markdown file and therefore typed (not pencil) on the pdf which you print and turn in. [n.b. This is the whole point of R Markdown: to create a way for reproducible code + analysis in one documnet.]