301056620

*Ebrahim Adeeb*

*2/9/2017*

## Stat 460 Assignment #2

**#1) Exercise 4.6 in text**

$$P(Increased\ Attendance) = P(A) = 0.95$$

$$P(Increased\ Attendance\ and\ Increased\ Sales) = P(A \cap B) = 0.97$$

$$P(Increased\ Sales\ Given\ Increased\ Attendance) = P(B \mid A) = P(A \cap B)/P(A) = 0.97/0.95 = 1.02$$

For those two event probabilities to be true the conditional probability (increased sales given incresed attendance) is greater than 1! Which is impossible because kolmogorovs axiom is not satisfied. Thus, Abe (The Statistical Stud) is incoherent but thats okay because he cant always be perfect.

**#2) Exercise 4.7 in text**

$$P(New\ Boyfriend\ by\ April) = 0.4$$

We cannot think about it in the symmetry POV. The probability of occurence in every month is not symmetric. Some months are better for dating than others and we dont know if Fiona is even looking for a boyfriend in specific months. The problem of interest does not involve a finite number of equally likely outcomes. Thinking about it in a frequential way may work by saying under the (theoretical) repitition of a year 40/100 times the experiment yields a single boyfriend by April. This is not very applicable or realistic because the reproducibility of the experiment is impossible because time is always moving forward.

The subjective probability makes more sense because if you know Fiona and have and understanding of her dating behaviour historically and know her current emotional state ( ie. Looking for a bf/not looking for a bf ). As long as your probility is coherent, which in this case it is, the subjective probability definition makes the most sense to use.

**#3)**

**CR7:**

$$X \equiv num\ of\ goals\ scored\ by\ C.Ronaldo\ in\ the\ 16/17\ La\ Liga\ season\ based\ on\ n_x\ shot$$

$$X = 12$$

$$n_x = 69$$

**LM10:**

$$Y \equiv num\ of\ goals\ scored\ by\ L.Messi\ in\ the\ 16/17\ La\ Liga\ season\ based\ on\ n_y\ shot$$

$$Y = 15$$

$$n_y = 61$$

**Distribution:**

$$X \sim Bin(n_x, p_x)$$

$$Y \sim Bin(n_y, p_y)$$

$$p_x \sim Beta(a_x, b_x)$$

$$p_y \sim Beta(a_y, b_y)$$

**a)**

This model assumes independence between shots and games which is untrue. A player could get his own rebound and shoot, thus the last shot affects the current one. Also after a bad game where many shots were missed a coach may ask the player to take less or more shots. In those examples our independent trials assumption is broken. Our number of trials is also not fixed because we dont know how many total shots a player will take. The probability of success is also not fixed per trial because the quality of the shot and location of the shot affects the probability of success.

I would suggest using a poisson binomial which is defined as discrete probability distribution of a sum of independent Bernoulli trials that are not necessarily identically distributed. The concept is named after Siméon Denis Poisson.

In other words, it is the probability distribution of the number of successes in a sequence of n independent yes/no experiments with success probabilities $p_1, p_2, \dots, p_n p_1, p_2, \dots, p_n$. The ordinary binomial distribution is a special case of the Poisson binomial distribution, when all success probabilities are the same, that is $p_1 = p_2 = \cdots = p_n p_1 = p_2 = \cdots = p_n$.

**b)**

I collected data from the past 4 seasons, I ommited the data from the current season because our prior cannot include X's/data points used in our likelihood function.

```r
C.Ronaldo=as.data.frame(matrix(c("12/13",46,235,"13/14"
,31,216,"14/15",48,225,"15/16",35,227), ncol=4,nrow=3))
rownames(C.Ronaldo) <- c('Season', 'Goals', "Shots")
colnames(C.Ronaldo) <- c('','','','')
C.Ronaldo
```

```
##
## Season 12/13 13/14 14/15 15/16
## Goals     46    31    48    35
## Shots    235   216   225   227
```

Mean of ratio (goals/shots)= 0.1766954

Standard Deviation of ratio (goals/shots)= 0.03322968

```r
L.Messi=as.data.frame(matrix(c("12/13",34,163,"13/14"
,28,160,"14/15",43,187,"15/16",26,158), ncol=4,nrow=3))
rownames(L.Messi) <- c('Season', 'Goals', "Shots")
colnames(L.Messi) <- c('','','','')
L.Messi
```

```
##
## Season 12/13 13/14 14/15 15/16
## Goals     34    28    43    26
## Shots    163   160   187   158
```

Mean of ratio (goals/shots)= 0.1945231

Standard Deviation of ratio (goals/shots)= 0.03017608

Mean of Beta Distribution $= \frac{\alpha}{\alpha+\beta}$

Variance of Beta Distribution $= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

- If we have a beta prior to a binomial parameter we can choose $\alpha =$ Success and $\beta =$ Failure as our Beta distribution parameters

$$p_x \sim Beta(160, 743)$$
$$p_y \sim Beta(131, 537)$$

**c)**

## Please see attached sheet for calculations

I was unable to use R to integrate out one variable, the code below integrates out both variables. If I was able to find the marginal posterior of "w" I would plot the distribution and the peak of the distribution would be my mode!

```r
library(cubature) # load the package "cubature"
f <- function(x) { (x[1]-x[2])^171*(1-x[1]+x[2])^799*
    (x[1]^145)*(1-x[1])^582 } # "x" is vector
adaptIntegrate(f, lowerLimit = c(0, 0), upperLimit = c(1, 1))
```

```
## $integral
## [1] -2.448374e+39
##
## $error
## [1] 2.407897e+34
```

```
##
## $functionEvaluations
## [1] 2873
##
## $returnCode
## [1] 0
```