**STT 810**

**ICA 2**

1. More practice with the sample function.
   1. Timing: create a sample function, starting with the following line:

x <- sample(c(0, 1), 1000, replace = TRUE)

It should be instantaneous. Increase the number of samples by a factor of 10. Keep on doing this until the generation is not instantaneous. Try to estimate the time it takes to generate (you might want to run it multiple times). Then increase by 3x, and measure the time. And finally, increase it by 10x (10 times the number you first timed). How does the time it took relate to the number of samples? You might want to do this in pairs (keep all sampling on a single computer, since results will vary).

# Question 1 a

start\_time <- Sys.time()

x <- sample(c(0, 1), 1000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 0.0009639263 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 10000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 0.0009918213 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 100000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 0.006230116 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 1000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 0.0632689 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 3000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 0.1984899 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 10000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 0.5360501 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 30000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 1.340603 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 100000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 4.713556 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 300000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 19.06151 secs

start\_time <- Sys.time()

x <- sample(c(0, 1), 1000000000, replace = TRUE)

end\_time <- Sys.time()

end\_time - start\_time

# Time difference of 1.021896 mins

The time difference varies but not in exact proportion to the number of simulations.

* 1. Use the sample function to put the numbers 1-10 in a random order.

x <- sample(seq(1:10),10, replace = FALSE)

x

7 4 9 2 1 3 10 6 8 5

1. Bayes’ Rule
   1. Take the spam filter example, with the given detection success rates. Suppose 60% of all emails are spam. What is the probability that a given message is spam, given that is marked as spam?

Ans – P(A|B) = P(B|A)\*P(A)/P(B)

P(A) = .60

P(B) = 0.8\*0.6 + 0.4\*0.1

P(B|A) = 0.8

Ans = 0.923 or 12/13 or 92.3%

* 1. What percent of all emails have to be spam in order for 50% of emails which are not marked to be spam to be actually spam?

Ans –

P (A|B) = P (B|A)\*P (A)/ P (B) = P (B|A)\*P (A) / (P (B|A)\*P (A) + P (B|Ã)\*P (Ã))

Where Ã is that A is false.

P (B|A) = 0.2

P (Ã) = 1 − P (A)

P(B|~A) = 0.9

Let P (A) = x   
0.5 = 0.2x / (0.2x + 0.9(1 − x))  
x = 0.818 or 81.8%.