Problem Set 2 **Statistical Methods In Engineering And Science**

Due Date: 10:00 PM, January 20, 2023 Prof. Alexander Giessing

Last Update: January 13, 2023 Winter Quarter, 2022

Study Group: ______________ ______________ ______________

Please upload your solution in a single pdf file on Canvas. Include all calculations, R-code, and figures (if applicable). All data sets are available on Canvas https://canvas.uw.edu/courses/1614615.

Question 1. A ball is drawn at random from an urn containing one white and one black ball. If the white ball is drawn, it is returned to the urn. If the black ball is drawn, it is put back into the urn together with two more black balls. Then a second draw is made. What is the probability that a black ball is drawn on both the first and the second draws?

Question 2. The Panama disease, also called *banana wilt*, is a devastating disease of bananas caused by the soil-inhabiting fungus species *Fusarium oxysporum forma specialis cubense*. Let A be the event that region R_1 is infested with the pathogen and B the event that region R_2 is infested. Suppose that $P(A) = 2/3$, $P(B) = 2/5$ and $P(A \cup B) = 4/5$. If the food inspection detects the pathogen in a shipment of bananas from region R_1, what is the probability that region R_2 is infested, too?

Question 3. Roughly 26% of the smartphones (in the US) run on iOS, 70% on Android, and 4% on another operating system. Suppose that 48% of the smartphones running on iOS, 57% of those running on Android, and 25% of those running on another operating system are infected by a virus.

(a) Draw the (probability) tree diagram for this problem.

(b) What is the probability that a randomly selected smartphone is infected by a virus?

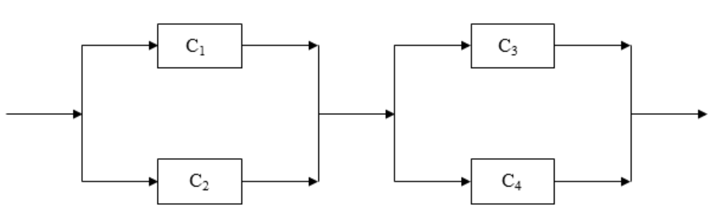
(c) Given that a randomly selected smartphone is infected by the virus, what is the probability that it runs on Android?

(d) Given that a randomly selected smartphone is not infected by the virus, what is the probability that it does not run on iOS?

Question 4. Suppose that a system consists of 4 independent components $C_1, \dots, C_4$ connected as below. The probability that each component works properly is 85%.

(a) What is the probability that the system functions properly?

(b) Given that the system works, what is the probability that component C_2 *is not* working?



Question 5. A cab was involved in a hit-and-run accident in West Seattle. 85% of the cabs in Seattle are yellow, but a witness identified the cab as *not*-yellow. The court tested the reliability of the witness under the circumstances that existed on the night of the accident and concluded that the witness had a chance of 8 out of 10 to correctly distinguish between yellow and *not*-yellow cars. What is the probability that the cab involved in the accident was indeed *not*-yellow?

Question 6. The Seattle Police Department reaches out to you to help them calibrate their breath analyzer used to test whether drivers exceed the legal limit set for blood alcohol content. Let A be the event "breath analyzer indicates that the driver's blood alcohol content exceeds the legal limit" and B the event "driver's blood alcohol content exceeds legal limit". The breath analyzer can be calibrated by choosing $0 < p < 1$ where $p = P(A \mid B)$. Due to technical constraints, $P(A \mid B) = P(A^c \mid B^c)$. On a typical Saturday night about 6% of the drivers are known to exceed the legal limit.

(a) Describe in words the meaning of $P(B^c \mid A)$.

(b) Determine $P(B^c \mid A)$ if $p = 0.85$.

(c) What value of p should the police use so that $P(B \mid A) = 0.95$?

Question 7. Consider adjusted closing prices of the SP500 index from Jan. 1, 2000 to September 8, 2016 (data set `SP500.cvs` on Canvas).

(a) What is the probability that on a randomly chosen day the SP500 index is down, i.e. the log-returns are negative?

(Hint: Let `ndays = length(rSP500)` be the number of days and then the probability that we are interested in is `sum(rSP500 < 0)/ndays`).

(b) What is the probability that the SP500 is down given it was down the previous day? Are the signs of the log-returns of two consecutive days approximately independent?

(Hint: `sum(rSP500[1:(ndays-1)]<0 & rSP500[2:ndays]<0)` is the number of two consecutive downs and `sum(rSP500[1:(ndays-1)]<0` is the number of previous days that the SP500 index is down.)

(c) What is the probability that the absolute value of the log-returns of the SP500 on a randomly selected day is at least 1.5%?

(d) If the absolute value of the log-returns of the SP500 is at least 1% on a randomly selected day, what is the probability that the absolute value of the log-return of the following day is at least 1.5%?

Remark: The first two problems are designed to show unpredictability of stock returns and the next two problems are intended to illustrate the dependence (predictability) of stock volatility.