ECS 132 Fall 2021: Assignment 3

Gabriel Yap

October 16, 2021

Instructions

1. You may in no circumstances upload your homework to private tutoring websites such as CourseHero or Chegg. Remember all material related to this course is a property of the University of California and posting them is a violation of the copyright laws.
2. If you refer to a source (either a book or the internet), you must cite it.
3. You are highly urged to work on these problems on your own. See the homework grading policy. Not getting the answer correct has very low penalty. However, trying it and then figuring out where you went wrong will really help you understand the material better and you will be much better prepared for the exams. If you do discuss with others, you must list their names.
4. Write your answers in Latex using the template provided, generate a PDF, and upload in Canvas.

1 Problem

Consider the following game in a game arcade. The probability of winning a game is p and hence loosing the game is $1 - p$. In order to win a prize you are given three choices:

- A: win at least once in 6 games;
- B: win at least twice in 12 games; and
- C: win at least 3 times in 18 games.

1. If $p = \frac{1}{6}$ which of A, B, or C should you choose to maximize your probability of winning the prize? You are required to compute the probabilities for each option to justify your answer.
   Let x be the number of times I win the game.
   A: P(x$\geq$1) = 1 - P(x = 0) = 1 - [$\binom{6}{0} * \frac{1}{6}^0 * (1 - \frac{1}{6})^6] = 1 - 0.3349 = 0.6651$
   B: P(x$\geq$2) = 1 - P(x = 0) - P(x = 1) = 1 - [$\binom{12}{0} * \frac{1}{6}^0 * (1 - \frac{1}{6})^{12}$] - [$\binom{12}{1} * \frac{1}{6}^1 * (1 - \frac{1}{6})^{11}] = 1 - 0.38133 =$ 0.6187
   C: P(x$\geq$3) = 1 - P(x = 0) - P(x = 1) - P(x = 2) = 1 - [$\binom{18}{0} * \frac{1}{6}^0 * (1 - \frac{1}{6})^{18}$] - [$\binom{18}{1} * \frac{1}{6}^1 * (1 - \frac{1}{6})^{17}$ - $\binom{18}{2} * \frac{1}{6}^2 * (1 - \frac{1}{6})^{16}$] = 1 - 0.4027 = 0.5973
   Assuming each game is independent of each other, I would choose option A to maximize my probability of winning the prize since it has the highest probability of a prize-winning trial.

2. Using the formulas derived in part (1), write a R-code to compute (not simulate) and plot the probability of winning the prize for options A, B, and C for different values of p varying from 0.1 to 0.9 in increments of 0.1. Summarize your observation.

   Regarding the plot: i) All three curves should be drawn in the same plot, ii) Make sure that the three curves are distinguishable (using different line styles, colors, and/or markers), and iii) make sure to add a legend that identifies the curves for the three different options (A, B, and C). Once you use the `plot` function, you can use `lines` to add lines to the plot. Also, use the `legend` function to add the legend.

```
    vecA <- c()    #creating vectors for each event A, B, and C
vecB <- c()
vecC <- c()
for (i in 1:9) { #storing data for event A
    probWinA = 1 - ((1 - i * 0.1)^6)
    vecA[i] = probWinA
}
for (i in 1:9) { #storing data for event B
    probWinB = 1 - ((1 - i * 0.1)^12) - (12 * (i * 0.1) * (1 - i * 0.1)^11)
    vecB[i] = probWinB
}
for (i in 1:9) { #storing data for event C
    probWinC = 1 - ((1 - i * 0.1)^18) - (18 * (i * 0.1) * (1 - i * 0.1)^17) - (153
    vecC[i] = probWinC
}
print(length(vecA))
print(length(vecB))
plot(vecA, type = "o", col = "red", xlab = "Probability to win * 0.1", ylab = "Prob
lines(vecB, type = "o", col = "blue")
lines(vecC, type = "o", col = "green")
```

Event A has the highest chance to win a prize before p $\approx$ 0.24, then afterwards it has the lowest chance to win a prize. Event C has the lowest chance to win a prize before p $\approx$ 0.24, but has the highest chance to win a prize afterwards. Event B was consistently the in between event A and C before and after p $\approx$ 0.24.

2 Problem

Suppose the university has designed a e-mail spam filter that attempts to identify by looking for commonly occurring phrases in spam. E-mail analysis has shown that 80% of email is spam. Suppose that 10% of the spam email contain the phrase “Large inheritance”, whereas this phrase is only used in 1% of non-spam emails. Suppose a new email is received with the phrase ”Large inheritance”, what is the probability that it is spam?
Let S be the event that the email is spam and let L be the event that an email has the phrase ”Large Inheritance”. Also, let S^c be the event than email is not spam and L^c be the event that an email does not have the phrase ”Large Inheritance”.
P(S) = 0.8 P(S^c) = 0.2 P(L|S) = 0.1 P(L|S^c) = 0.01
$P(S \mid L) = \frac{L|S*P(S)}{P(L)} = \frac{0.1*0.8}{P(L|S)*P(S)+P(L|S^c)*P(S^c)} = \frac{0.1*0.8}{0.1*0.8+0.01*0.2} = 0.9756$

3 Problem

A binary communication channel carries data as one of two types of signals denoted by 0 and 1. Owing to noise, a transmitted 0 is sometimes received as a 1 and a transmitted 1 is sometimes received as a 0. For a given channel, assume a probability of 0.94 that a transmitted 0 is correctly received as a 0 and a probability 0.91 that a transmitted 1 is received as a 1. Further assume a probability of 0.45 of transmitting a 0. If a signal is sent, determine
Let event T be the event that a 0 is transmitted and event R be the event that a 0 is received. T^c is the event a 1 is transmitted and R^c is the event a 1 is received.

1. Probability that a 1 is received
   $P(R^c) = 1 - P(R) = 1 - 0.4725 = 0.5275$

2. Probability that a 0 is received
   $P(R) = P(R \mid T) * P(T) + P(R \mid T^c) * P(T^c) = 0.94 * 0.45 + 0.09 * 0.55 = 0.4725$

3. Probability that a 1 was transmitted given that a 1 was received
   $P(T^c \mid R^c) = \frac{P(R^c|T^c)*P(T^c)}{P(R^c)} = \frac{0.91*0.55}{0.5275} = 0.9488$

4. Probability that a 0 was transmitted given that a 0 was received
   $P(T \mid R) = \frac{P(R|T)*P(T)}{P(R)} = \frac{0.94*0.45}{0.4725} = 0.8952$

5. Probability of an error
   $P(''error'') = P(R^c \mid T) * P(T) + P(R \mid T^c) * P(T^c) = 0.06 * 0.45 + 0.09 * 0.55 = 0.0765$

4 Problem

Suppose that a die is rolled twice. What are the possible values that the following random variables can take

1. the maximum value to appear in the two rolls;
   Let X be the maximum value to appear in the two rolls: X = max{roll1, roll2} = {1, 2, 3, 4, 5, 6}

2. the value of the first roll minus the value of the second roll?
   Let Y be the value of the first roll minus the value of the second roll: Y = roll1 - roll2 = {-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5}

3. Calculate the probabilities associated with the above two random variables?
   There are 36 total combinations and each combination has a probability of $\frac{1}{36}$
   $P(X = 1) = (1, 1) = \frac{1}{36}$
   $P(X = 2) = (1, 2), (2, 1), (2, 2) = \frac{3}{36}$
   $P(X = 3) = (1, 3), (2, 3), (3, 3), (3, 2), (3, 1) = \frac{5}{36}$
   $P(X = 4) = (1, 4), (2, 4), (3, 4), (4, 4), (4, 3), (4, 2), (4, 1) = \frac{7}{36}$
   $P(X = 5) = (1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (5, 4), (5, 3), (5, 2), (5, 1) = \frac{9}{36}$
   $P(X = 6) = (1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 5), (6, 4), (6, 3), (6, 2), (6, 1) = \frac{11}{36}$
   $P(Y = -5) = (1, 6) = \frac{1}{36}$
   $P(Y = -4) = (1, 5), (2, 6) = \frac{2}{36}$
   $P(Y = -3) = (1, 4), (2, 5), (3, 6) = \frac{3}{36}$
   $P(Y = -2) = (1, 3), (2, 4), (3, 5), (4, 6) = \frac{4}{36}$
   $P(Y = -1) = (1, 2), (2, 3), (3, 4), (4, 5), (5, 6) = \frac{5}{36}$
   $P(Y = 0) = (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6) = \frac{6}{36}$

$P(Y=1) = (6,5),(5,4),(4,3),(3,2),(2,1) = \frac{5}{36}$
$P(Y=2) = (6,4),(5,3),(4,2),(3,1) = \frac{4}{36}$
$P(Y=3) = (6,3),(5,2),(4,1) = \frac{3}{36}$
$P(Y=4) = (6,2),(5,1) = \frac{2}{36}$
$P(Y=5) = (6,1) = \frac{1}{36}$

5 Problem

The probability that a patient recovers from a rare blood disease is 0.4 and 10 people are known to have contracted this disease. Let X denote the random variable which denotes the number of patient who survive from the disease.

1. Plot the probability mass function (pmf) of X.

```
      n <- 10
p <- 0.4
x <- 0:n
plot(x, dbinom(x, size = n, prob = p), main = "PMF of X", xlab = "# of patients
```

2. Plot the cumulative distribution function (cdf) of X.

```
      n <- 10     # for #5.2
p <- 0.4
x <- 0:n
plot(x, pbinom(x, size = n, prob = p), type = "s", main = "CDF of X", xlab = "
```

3. What is the probability that at least 8 survive, i.e., $P\{X \geq 8\}$?
   The probability of at least 8 people surviving will the the sum of P(X=8), P(X=9), and P(X=10). So we can calculate it as:
   $\sum_{i=8}^{10} \binom{10}{n} * 0.4^n * 0.6^{10-n} \approx 0.0123$

4. What is the probability that 3 to 8 survive, i.e., $P\{3 \leq X \leq 8\}$?
   THe probability of 3 to 8 patients surviving is the sum of P(X = 3), P(X = 4), P(X = 5), P(X = 6), P(X = 7), P(X = 8). We can calculate it as:
   $\sum_{i=3}^{8} \binom{10}{n} * 0.4^n * 0.6^{10-n} \approx 0.8310$

6 Problem

This problem is to show that determining if two events are independent is not always obvious.

1. Consider a family of 3 children. Consider the following two events. A is the event that the family has children of both sexes and B is the event that there is at most one girl. Are events A and B independent?
   Yes, A and B are independent. The definition of independence for two events is $P(A \cap B) = P(A) * P(B)$. Event A can happen 6 out of the 8 possible combinations of children (BGG), (GBG), (GGB), (BBG), (BGB), (GBB). Event B can happen 4 out of the 8 possible combinations, (BBB),(GBB),(BGB),(BBG). Events A and B jointly occur 3 out of the 8 possible outcomes (GBB), (BGB), (BBG). This is the same as $\frac{6}{8} * \frac{4}{8} = \frac{3}{8}$.

2. What is the answer in a family with 4 children?
   When there are 4 children, A and B are not independent. Event A occurs $1 - \{P(BBBB) + P(GGGG)\} = 1 - \frac{2}{16} = \frac{14}{16}$. Event B occurs $\{$(BBBB), (GBBB), (BGBB), (BBGB), (BBBG)$\}$. The joint event of A and B occurs $\frac{4}{16}\{GBBB, BGBB, BBGB, BBBG\}$. However, when computing the joint probability via multiplication of P(A) and P(B), we get $P(A) * P(B) = \frac{14}{16} * \frac{4}{16} = 0.2188 \neq \frac{4}{16}$.