ECS 132 Fall 2020: Assignment 3 - 15 Points

October 24, 2020

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 R Markdown and submit the knitted pdf on Gradescope; for due date and other details see the Homework Policy and Schedule.

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.
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.

1.1 Answer

First, a note about this problem. In 1693 Samuel Pepys and Isaac Newton corresponded over a very similar problem posed by Pepys in relation to a wager he planned to make. Newton, of course solved the problem!

Also interesting fun fact! In 1665, “social distancing” orders emptied campuses throughout England, as the bubonic plague raged, killing 100,000 people (roughly one-quarter of London’s population), in just 18 months. Isaac Newton then a 24-year-old student from Trinity College, Cambridge was among those forced to leave campus and return indefinitely to his childhood home. During this time away he discovered differential and integral calculus, formulated the laws of gravity, and explored optics, experimenting with prisms and investigating light.

1. It is easy to find the probability that you loose all 6 games. Our required event is the complement of that.

$$\begin{aligned} P(A) &= 1 - (5/6)^6 \\ &= 0.665 \end{aligned}$$

2. In this case we find the probability of loosing all 12 games and the probability of winning only 1 game and then subtract the sum of those from 1

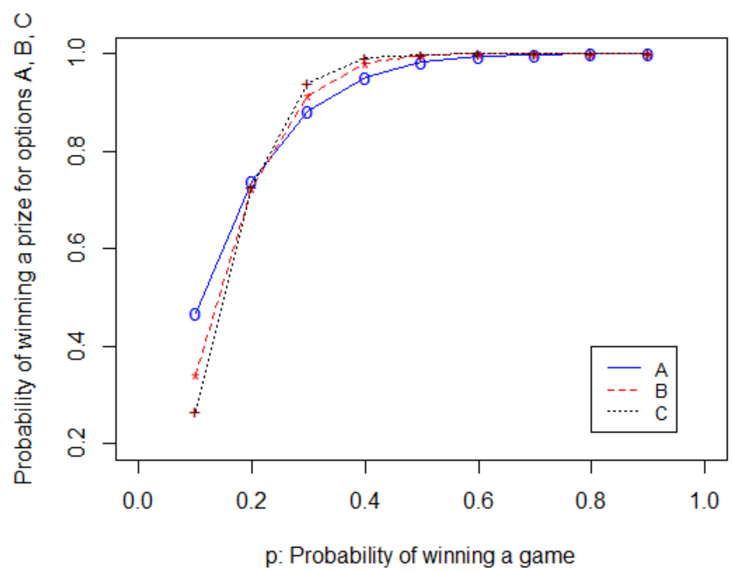
$$\begin{aligned} P(B) &= 1 - P(\text{0 wins in 12 games}) - P(\text{1 win in 12 games}) \\ &= 1 - (5/6)^{12} - 12 \times (5/6)^{11}(1/6)^1 \\ &= 0.619 \end{aligned}$$

3. If we define P(i) to be the probability that there are at least three wins in 18 games, then

$$\begin{aligned} P(C) &= 1 - \sum_{0}^{2} \binom{18}{i} (1/6)^i (5/6)^{18-i} \\ &= 0.597 \end{aligned}$$

```
pa = numeric(9)
pb = numeric(9)
pc = numeric(9)

p = c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

for (i in 1:9){

    pa[i] = 1 - (1 - p[i])^6

    pb[i] = 1 - (1 - p[i])^12 - choose(12,1)*p[i]*(1 - p[i])^11

    pc[i] = 1 - (1 - p[i])^18 - choose(18,1)*p[i]*(1 - p[i])^17 - choose(18,2)*(p

}
plot(p, pa, type="o", col="blue", pch="o", lty=1,  ylim= c(0.2,1), xlim=c(0,1), xl
points(p, pb, col="red", pch="*")
lines(p, pb, col="red",lty=2)
```

```
points(p, pc, col="dark red", pch="+")
lines(p, pc, col="black", lty=3)
legend(x = 0.8,y=0.4, legend=c("A", "B", "C"),
      col=c("blue", "red", "black"), lty=1:3, cex=0.8)
```



1.2 Points - 3 pts

All of the points below apply for each of scenario A, B, and C. So, you can divide the points below into thirds for each scenario – that is to say if you get A and B correct, but C incorrect, award 2/3 pts

- 1 point for properly setting up and identifying the probabilities (from part 1)
- 1 point for properly setting up the R code
- 1 point for properly setting up and generating the plots

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?

2.1 Answer

We define the following 2 events

- S: event that an e-mail is spam
- F: event that the email has the phrase ”Large inheritance”

Using Bayes’ Rule we find

$$\begin{aligned} P(S|F) &= \frac{P(F|S)P(S)}{P(F)} \\ &= \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|S^c)P(S^c)} \\ &= \frac{0.1 \times 0.8}{0.1 \times 0.8 + 0.01 \times 0.2} \\ &\approx 0.9756 \end{aligned}$$

2.2 Points - 1pt

- 1/2 pt for properly identifying the events labeled S and F in the answer (it’s fine if you called them something different)
- 1/2 pt for properly setting up Bayes’ Theorem. Please note, you must use Bayes’ Theorem for this problem.

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

1. Probability that a 1 is received
2. Probability that a 0 is received
3. Probability that a 1 was transmitted given that a 1 was received
4. Probability that a 0 was transmitted given that a 0 was received
5. Probability of an error

3.1 Answers

Define the following events:

- T_0 : A 0 is transmitted
- R_0 : A 0 is received
- T_1 : $\overline{T_0}$
- R_1 : $\overline{R_0}$

Then events in parts 1 through 4 are $R_1, R_0, T_1|R_1, T_0|R_0$ and $P(\text{Error}) = P(T_1 \cap R_0) \cup P(T_0 \cap R_1)$
We can easily obtain the following:

- $P(R_1|T_0) = P(\overline{R_0}|T_0) = 1 - P(R_0|T_0) = 0.06$
- $P(R_0|T_1) = P(\overline{R_1}|T_1) = 1 - P(R_1|T_1) = 0.09$
- $P(T_1) = P(\overline{T_0}) = 1 - P(T_0) = 0.55$

Using the theorem of total probability and Bayes theorem we get:

1. $P(R_1) = P(R_1|T_0)P(T_0) + P(R_1|T_1)P(T_1) = (0.06)(0.45) + (0.91)(0.55) = 0.5275$
   OR
   $P(R_1) = P(\overline{R_0}) = 1 - P(R_0) = 0.5275$
2. $P(R_0) = P(R_0|T_0)P(T_0) + P(R_0|T_1)P(T_1) = (0.94)(0.45) + (0.09)(0.55) = 0.4725$
   OR
   $P(R_0) = P(\overline{R_1}) = 1 - P(R_1) = 0.4725$
3. $P(T_1|R_1) = \frac{P(R_1|T_1)P(T_1)}{P(R_1)} = \frac{(0.91)(0.55)}{(0.5275)}$
4. $P(T_0|R_0) = \frac{P(R_0|T_0)P(T_0)}{P(R_0)} = \frac{(0.94)(0.45)}{(0.4725)}$
5. $$\begin{aligned} P(\text{ Error}) &= P(T_1 \cap R_0) + P(T_0 \cap R_1) \\ &= P(T_1|R_0)P(R_0) + P(T_0|R_1)P(R_1) \\ &= (1 - P(T_0|R_0))P(R_0) + (1 - P(T_1|R_1))P(R_1) \\ &= (0.1048)(0.4725) + (0.0512)(0.5275) \end{aligned}$$

3.2 Points - 5 pts

Correctly computing all of the following probabilities. Symbols are consistent with the answer section – it does not matter what you named these events, any correct calculations are acceptable.

- 1 pt: $p(R_1)$
- 1 pt: $p(R_0)$
- 1 pt: $p(T_1|R_1)$
- 1 pt: $p(T_0|R_0)$
- 1 pt: $p(R \neq T)$

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;
2. the value of the first roll minus the value of the second roll?
3. Calculate the probabilities associated with the above two random variables?

4.1 Answer

Let X denote the random variable which is the maximum value to appear in two rolls of a die. X can take values $\{1, 2, 3, 4, 5, 6\}$.

Let Y denote the random variable which is the value of the first roll minus the second roll. Y can take values $\{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$.

$$
\begin{aligned}
toremovenumbering(beforeeachequation) P\{X = 1\} &= P\{(1, 1)\} \\
&= \frac{1}{36} \\
P\{X = 2\} &= P\{(2, 1), (1, 2), (2, 2)\} \\
&= \frac{3}{36} \\
P\{X = 3\} &= \frac{5}{36} \\
P\{X = 4\} &= \frac{7}{36} \\
P\{X = 5\} &= \frac{9}{36} \\
P\{X = 6\} &= \frac{11}{36}
\end{aligned}
$$

$$
\begin{aligned}
P\{Y = -5\} &= P\{(1, 6)\} \\
&= \frac{1}{36} \\
P\{Y = -4\} &= P\{(2, 6), (1, 5)\} \\
&= \frac{2}{36} \\
P\{Y = -3\} &= \frac{3}{36} \\
P\{Y = -2\} &= \frac{4}{36} \\
P\{Y = -1\} &= \frac{5}{36} \\
P\{Y = 0\} &= \frac{6}{36}
\end{aligned}
$$

The probabilities of the other values of the random variable Y can written by noting the symmetry.

4.2 Points - 3 pts

- 1/2 pt: Properly define the sample space for max(2 rolled dice)
- 1/2 pt: Properly define sample space for the difference between two dice (die1-die2)
- 2 pts: Properly compute the probabilities from above sample spaces (1 pt each)

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.
2. Plot the cumulative distribution function (cdf) of X.
3. What is the probability that at least 8 survive, i.e., $P\{X \geq 8\}$?
4. What is the probability that 3 to 8 survive, i.e., $P\{3 \leq X \leq 8\}$?

5.1 Answers

1. Plot the probability mass function (pmf) of X. We can use R to get all the values and plot them.

```
n <- 10
k <- seq(0,n)
p <- 0.4
pr <- dbinom(k, n, p)
pr
[1] 0.0060466176 0.0403107840 0.1209323520 0.2149908480 0.2508226560
[6] 0.2006581248 0.1114767360 0.0424673280 0.0106168320 0.0015728640
[11] 0.0001048576
barplot(pr, names.arg=k, ylab = "p(i)", xlab = "i")
```

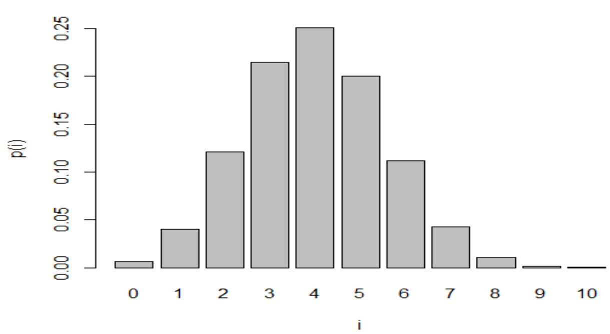


Figure 1: The probability mass function of Binomial random variable with parameter $n = 10$ and $p = 0.4$.

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

```
pr <- pbinom(k, n, p)
[1] 0.006046618 0.046357402 0.167289754 0.382280602 0.633103258 0.833761382
[7] 0.945238118 0.987705446 0.998322278 0.999895142 1.000000000
barplot(pr, names.arg=k, ylab = "P(X <= i)", xlab = "i")
```

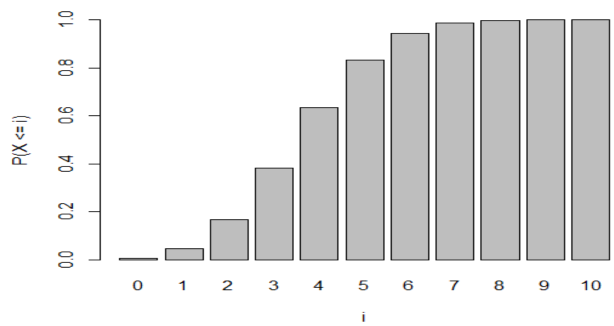


Figure 2: The cumulative distribution function of Binomial random variable with parameter $n = 10$ and $p = 0.4$.

3. What is the probability that at least 8 survive, i.e., $P\{X \geq 8\}$? $P\{X \geq 8\} = 1 - P\{X \leq 7\}$. Using R, we can do the following

```
pr = pbinom(7, n, 0)
ans <- 1 - pr
ans
0.01229455
```

4. What is the probability that 3 to 8 survive, i.e., $P\{3 \leq X \leq 8\}$? The required probability is

$$\begin{aligned} P\{3 \leq X \leq 8\} &= P\{X = 3\} + P\{3 < X \leq 8\} \\ &= P\{2 < X \leq 8\} \\ &= pbinom(8, n, p) - pbinom(2, n, p) \end{aligned}$$

Again, we can use R

```
pbinom(8, n, p) - pbinom(2, n, p)
```

5.2 Points - 3 pts

- 1 pt: Plot of the PMF of X
- 1 pt: Plot of the CDF of X
- 1/2 pt: $p(X \leq 8)$
- 1/2 pt: $p(3 \leq X \leq 8)$