MATH 241 Fall 2015 Homework #7

Professor Adam Kapelner

Due 4PM in my mail slot, Friday, October 30, 2015

(this document last updated Thursday 29th October, 2015 at 8:40am)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read Chapter 4 on the geometric, negative binomial r.v.'s. in Ross (chapter references are from the 7th edition).

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 15 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name (thanks to Anastassiya and Jasmine of Section A for figuring this out). If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records. Write your name and section below (A, B or C). 10 points off if your section is blank or wrong.

NAME: _____ SECTION (A, B or C): _____

Stopping time r.v.'s   We will be looking at the negative binomial and its special case of the geometric.

Problem 1

We will be investigating r.v.'s by imagining a trip the grocery store to buy ingredients for guacamole.

(a) [easy] You buy *one* avocado at the grocery store which is mushy. Thus, it may have brown inside because it's partially rotten. Call this probability of rottenness p. Model the number of *good* avocados you have using a random variable. Call this r.v. X. Hint: the number of good avocadoes is either zero or one since you buy one and if it's good, you have one; if it's bad you have zero. Please remember the probability of rottenness is p and I'm asking for the probability of the avocado being good *i.e.* not rotten. All you need to write is $X \sim$ something. You do not need to write the PMF, draw the PMF, draw the CDF, nor contemplate the meaning of life in the next centimeter of white space.

Here is my answer

(b) [easy] You buy 10 avocadoes. Assume the draws of avocadoes are independent. Model the number of *good* avocados you have using a random variable. Please remember the probability of rottenness is p and I'm asking for good avocados *i.e.* not rotten. Call this r.v. X and write $X \sim$ something below.

(c) [easy] Comment on why the r.v. you created in (b) is the sum of many $\overset{iid}{\sim}$ r.v.'s you modeled in (a).

(d) [easy] Write the PMF for the r.v. you created in (b).

(e) [easy] Write the support for the r.v. you created in (b).

(f) [easy] Use the sigma notation for summing (e.g. $\sum_{i=1}^{5}$) to calculate the probability that you get 3, 4, 5 or 6 good avocados. Since you don't know p you cannot actually compute a numerical value for this probability. Leave it in sigma notation.

(g) [easy] If you did this activity (bought 10 avocadoes) many times, what would be the average number of rotten avocadoes you get?

(h) [easy] If you did this activity once, what is the expcted number of rotten avocadoes you get?

(i) [easy] Now you do another activity. You take one avocado, cut it open and see if it's rotten. You keep doing this until you see a rotten avocado. Model the number of avocados you cut open using a r.v. Call this r.v. X.

(j) [easy] Write the PMF for the r.v. you created in (i).

(k) [easy] Write the support for the r.v. you created in (i).

(l) [easy] What is the probability you stop when looking at the third avocado?

(m) [easy] Use the sigma notation for summing (e.g. $\sum_{i=1}^{5}$) to calculate the probability that you stop between 4 and 37 avocados (including 4 and including 37). Since you don't know p you cannot actually compute a numerical value for this probability. Leave it in sigma notation.

(n) [harder] Let's say you learned how to detect rotten avocadoes and you used this learning to select new avocados. What assumption is violated?

(o) [harder] Let's say there were two baskets of avocados at the grocery store. The first basket comes from California-grown avocados and the second basket comes from Mexican-grown avocados. At some point in your picking of avocados you move from one basket to the other. What assumption is violated now?

Problem 2

We will rederive the negative binomial PMF as we did in class. The probability of success if p and the number of successes we wish to find is r.

(a) [easy] If we are waiting x trials to finally see exactly r successes, what does the outcome result of the last trial *need* to be?

(b) [easy] How many trials do we witness in order to witness $r - 1$ successes not counting the last trial?

(c) [easy] Can these $r - 1$ successes happen anywhere within these $x - 1$ trials?

(d) [easy] If you get $r - 1$ successes in $x - 1$ trials, how many failures do you get?

(e) [easy] How many ways is there to get $r - 1$ successes among $x - 1$ trials?

(f) [easy] What is the probability of getting $r - 1$ successes and $x - r$ failures *in that order* if successes and failures are independent?

(g) [easy] Use the answers in (e) and (f) to find the probability of getting $r - 1$ successes in $x - 1$ trials.

(h) [easy] Use the answers in (g) and the probability of a final success to finally derive the full PMF of the Negative Binomial distribution.

(i) [easy] Let $X \sim \text{NegBin}(r, p)$. What is the support of X?

(j) [harder] What is the parameter space of r and p? Be careful not to allow degenerate cases.

Problem 3

You are testing RAM. The manufacturing process is near perfect. The probability of finding faulty RAM is about 1 in 300. We assume all RAM chips are independent with respect to whether they are faulty.
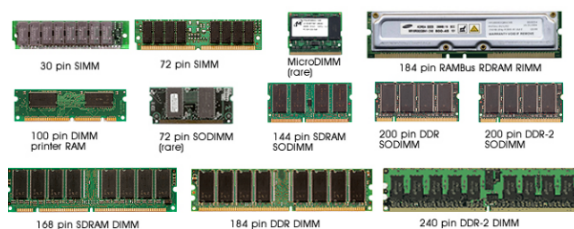
(a) [easy] What is the probability you get three faulty RAM chips in a row?

(b) [harder] What is the probability you have to investigate 100 RAM chips in order to find exactly 3 faulty chips? Compute explicitly.

(c) [harder] What is the probability you have to investigate 500 RAM chips in order to find exactly 3 faulty chips? You can leave in choose notation and use exponents as well.

(d) [harder] What is the probability you have to investigate 6 faulty RAM chips in order to find exactly 500 working chips? Compute explicitly. You may want to use the other parameterization from class.

(e) [difficult] What is the probability you have to investigate more than 500 RAM chips to see exactly 3 faulty chips? You can leave in choose notation and use exponents as well. Leave in terms of the binomial CDF.

Problem 4

We are going to return to our in-class discussion of the my ride from my office to my apartment in Forest Hills using the Uber Taxi service. Since Queens is in New York City, I will be modeling based on Uber NYC rates. The current rates are posted here. I have extra incentive now to provide you with a realistic model.

For the purposes of this exercise, assume there are only two routes in which to drive back. This is close to realistic. There is the "Van Wyck" (outlined in black on the right below) and "Jewel Ave" which is the Q64 bus route (outlined in black on the left below).
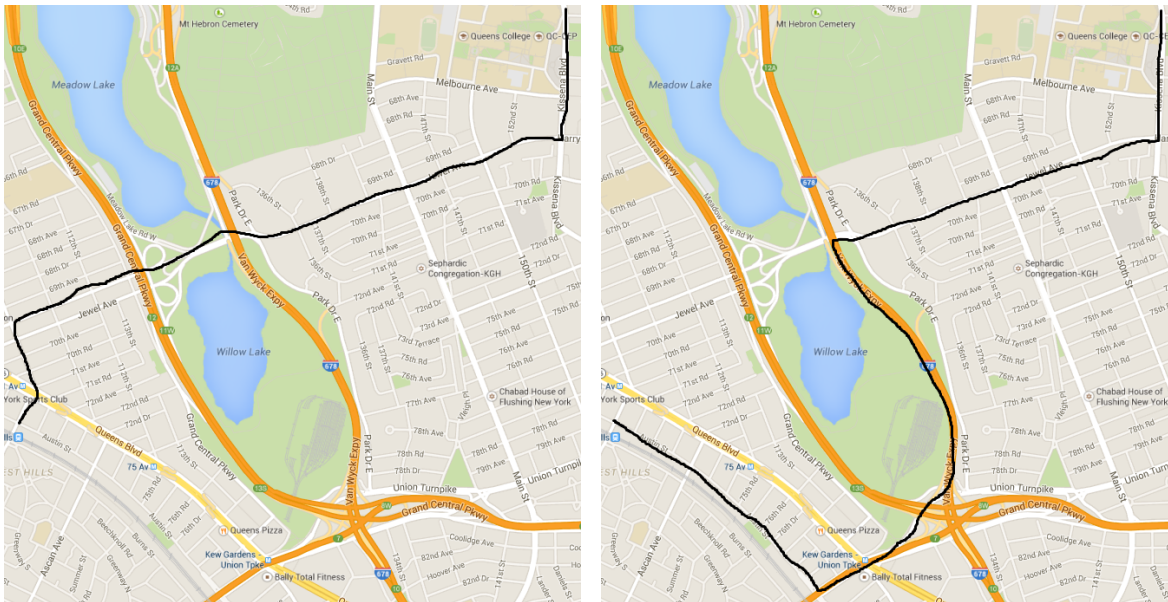
The only determinant of route selection is whether or not there is traffic on the Van Wyck. If there is traffic, I take Jewel Ave route; if not, I take the Van Wyck route. The probability of traffic on the Van Wyck is 30%. The Jewel Ave route is 2.3 miles and takes 10 min and the Van Wyck route is 6 min and is 3.6 miles.

(a) [easy] Let W be the r.v. which models the time I travel in the Uber Taxi. What is its distribution? Use the notation we used in class.

(b) [easy] What is $\text{Supp}\,[W]$?

(c) [easy] Compute $\mathbb{E}\,[W]$ from the definition of expectation.

(d) [easy] Write a sentence that synthesizes what part (c) means.

(e) [easy] Let D be the r.v. which models the distance I travel in the Uber Taxi. What is its distribution? Use the notation we used in class.

(f) [easy] Compute $\mathbb{E}\,[D]$.

(g) [difficult] Are the r.v.'s W and D dependent? Justify your answer *in English*.

(h) [easy] Write a sentence that synthesizes what part (f) means.

(i) [easy] UberX charges \$0.40\min. Let M be the r.v. which is what I pay for time on my trip home. Find the distribution of M.

(j) [easy] Write M as a function of W.

(k) [easy] Calculate $\mathbb{E}[M]$ based on the formula we learned in class about expectations of r.v.'s scaled by a constant.

(l) [easy] UberX charges \$2.15\mi of distance covered. Let L be the r.v. which is what I pay for mileage on my trip home. Find the distribution of L.

(m) [easy] Write L as a function of D.

(n) [easy] Calculate $\mathbb{E}[L]$ based on the formula we learned in class about expectations of r.v.'s scaled by a constant.

(o) [easy] Uber also includes a base fare of \$3. Let B be the r.v. which models the total bill for my uberX ride. Write B as a function of W and D.

(p) [harder] We didn't really cover this in class, but you should be able to do it. W and D are one-to-one so the scaled W and scaled D sum is really one r.v. Find $\mathbb{E}[B]$ based also on the formula we learned in class about the expectation of a r.v. with a constant added.

(q) [easy] Write a sentence that synthesizes what part (p) means.

(r) [difficult] UberBLACK is the original Uber taxi service. They dispatch a luxury black sedan to pick me up. The base fare is $7 and they charge $0.65\min and $3.75\mi. Calculate $\mathbb{E}[B]$ where B is now the total bill for UberBLACK.

Problem 5

This is the fun part of the homework. You're going to repeat the experiments we did in class.

(a) [easy] Write the definitions of "datum" and "data" from class. Not from the dictionary!

(b) [easy] Why are r.v.'s also called "data generating processes?"

(c) [easy] Grab a cup and 8 pennies (or nickels, or dimes, etc). Use a magic marker to mark four of them (front and back). If you shake the cup and pull out three coins, let X be the r.v. for how many marked coins you pull out? How is X distributed? Write "$X \sim$ something" below.

(d) [easy] Using as fact that $\mathbb{E}[X] = n\frac{K}{N}$ when $X \sim \text{Hypergeometric}(n, K, N)$ (see Problem 6), calculate $\mathbb{E}[X]$ for the r.v. you constructed in part (c).

(e) [easy] Shake the cup and take out 3 coins. How many were marked? Repeat this five times. Record your data below. That is, just write down the five numbers separated by commas.

(f) [easy] Find $\bar{x}$ from the data you recorded in part (e).

(g) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make $\bar{x}$ closer to $\mathbb{E}[X]$?

(h) [easy] Now forget that the coins are marked. If you shake the cup and flip all 8 coins, let X be the r.v. for how many heads are flipped. How is X distributed? Write "$X \sim$ something" below.

(i) [easy] Using the fact we proved in class that $\mathbb{E}[X] = np$ when $X \sim \text{Binomial}(n, p)$, calculate $\mathbb{E}[X]$ for the r.v. you constructed in part (i).

(j) [easy] Shake the cup and count the number of heads. Repeat this five times. Record your data below.

(k) [easy] Find $\bar{x}$ from the data you recorded in part (l).

(l) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make $\bar{x}$ closer to $\mathbb{E}[X]$?

(m) [easy] Now imagine one coin in the cup and success is defined as getting a head. Further imagine that you don't stop flipping this coin until you get a head. Let X be the r.v. for how many flips you make. How is X distributed? Write "$X \sim$ something" below.

(n) [easy] Using the fact we proved in class that $\mathbb{E}[X] = 1/p$ when $X \sim \text{Geometric}(p)$, calculate $\mathbb{E}[X]$ for the r.v. you constructed in part (o).

(o) [easy] Flip until you get a head. Repeat this five times. Record your data below.

(p) [easy] Find $\bar{x}$ from the data you recorded in part (r).

(q) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make $\bar{x}$ closer to $\mathbb{E}[X]$?

(r) [easy] Now imagine one coin in the cup and success is defined as getting a head. Further imagine that you don't stop flipping this coin until you get two heads on at least two independent tosses. We did not have time to do this one in class. Make sure you actually do it! Let X be the r.v. for how many flips you make. How is X distributed? Write "$X \sim$ something" below.

(s) [easy] Using as fact that $\mathbb{E}[X] = r/p$ when $X \sim \text{NegBin}(r, p)$ (see Problem 6), calculate $\mathbb{E}[X]$ for the r.v. you constructed in part (u).

(t) [easy] Flip until you get two heads. Repeat this five times. Record your data below.

(u) [easy] Find $\bar{x}$ from the data you recorded in part (x).

(v) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make $\bar{x}$ closer to $\mathbb{E}[X]$?

Problem 6

Googlable but more fun if you try on your own first. Do on a separate piece of paper.

(a) [E.C.] $X \sim \text{Hypergeometric}(n, K, N)$. Verify $\sum_{x \in \text{Supp}[X]} p(x) = 1$.

(b) [E.C.] $X \sim \text{Hypergeometric}(n, K, N)$. Verify $\mathbb{E}[X] = n\frac{K}{N}$.

(c) [E.C.] $X \sim \text{NegBin}(r, p)$. Verify $\mathbb{E}[X] = \frac{r}{p}$.