

# MATH 241 Spring 2015 Homework #7

Professor Adam Kapelner

Due 5PM, Wednesday, Apr 1, 2015

(this document last updated Tuesday 31<sup>st</sup> March, 2015 at 11:37am)

## 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*. But for this homework set, read the relevant parts of Chapter 4 about the geometric r.v. and the expectation operator of r.v.’s.

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]”, **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. If the problem asks you for a computation, **round to two or three decimals (do not answer in an exact fraction).**

This homework is worth 100 points but the point distribution will not be determined until after the due date. Late homework will be penalized **20 points for the first day and 50 points thereafter due to spring break.**

Between 1–15 points are arbitrarily given as a bonus (conditional on quality) if the homework is typed using L<sup>A</sup>T<sub>E</sub>X (15 points only if it is perfect L<sup>A</sup>T<sub>E</sub>X). Links to installing L<sup>A</sup>T<sub>E</sub>X and program for compiling L<sup>A</sup>T<sub>E</sub>X are found on the syllabus. You are encouraged to use [overleaf.com](http://overleaf.com) (make sure you upload both the hwxx.tex and the preamble.tex file). 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. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert as a figure 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 L<sup>A</sup>T<sub>E</sub>X, print this document and write in your answers. **Handing it in without this printout is NOT ACCEPTABLE.** You will have to redo it.

Keep this page printed for your records. Write your name and section below where section A is if you’re registered for the 9:15AM–10:30AM lecture and section B is if you’re in the 12:15PM–1:30PM lecture.

NAME: \_\_\_\_\_ SECTION (A or B): \_\_\_\_\_

## Problem 1

We will do a few short questions asking which r.v. to use and why. We will be investigating this 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 avocados 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.
- (b) [easy] You buy 10 avocados. Assume the draws of avocados 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 avocados) many times, what would be the average number of rotten avocados you get?
- (h) [easy] If you did this activity once, what is the expected number of rotten avocados 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 (f).
- (k) [easy] Write the support for the r.v. you created in (f).

- (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 avocados 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 didn't cover the Poisson r.v. in class, so here's an introduction to it. The Poisson will not be covered on the exam, but I may give you a new r.v. with its PMF and ask you to make do just like I do here.

" $X \sim \text{Poisson}(\lambda)$ " means random variable  $X$  is Poisson-distributed with parameter  $\lambda$ . The Poisson is the limit of the binomial as  $n$  gets large and  $p$  gets small where  $\lambda = n \times p$ . Thus as  $n$  increases without bound, the support must be the binomial's support with  $n \rightarrow \infty$ . Thus,  $\text{Supp}[X] = \{0, 1, \dots\} = \mathbb{N}_0$ . The Poisson can be thought of as infinite  $\overset{iid}{\sim}$  Bernoulli trials where  $p \approx 0$ . The PMF is given below:

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

We will use this probability model to model traffic accidents on Queens Boulevard. Queens Blvd. is known by some as the “Boulevard of Death” (you can read the Wikipedia information here).

Fatal pedestrian accidents on Queens Blvd. since 2001 is presented in the table below:

2001	2002	2003	2004	2005	2006	2007
4	2	5	1	2	2	1

Table 1: Number of pedestrian accidents by year.

- (a) [easy] Calculate the sample average  $\bar{x}$  from Table 1.
- (b) [easy] Consider the data in Table 1 above to be  $\stackrel{iid}{\sim}$  draws from the distribution discussed in the problem header. Thus, the accidents for 2001 – 2007 can be thought of as:

$$X_{2001}, X_{2002}, \dots, X_{2007} \stackrel{iid}{\sim} \text{Poisson}(\lambda)$$

Let:

$$\bar{X} := \frac{X_{2001} + X_{2002} + \dots + X_{2007}}{7}$$

Is  $\bar{x}$ , the sample average using the data in Table ?? (your answer to part a) a realization from  $\bar{X}$  the “average r.v.?” Yes or no is fine.

- (c) [easy] If we had many more years of data than the 7 years, what would  $\bar{X}$  converge to eventually? You can use notation from class. Do not answer using a number.
- (d) [easy] What is the law we used in (c)? Look in your notes. State the law and write a sentence about what it means to you *in English*.

- (e) [easy] In problem 5 you prove that  $\mathbb{E}[X] = \lambda$  if  $X \sim \text{Poisson}(\lambda)$ . Assume this as fact. Now pretend 7 years of data is as good as infinite years of data and use your answer to (a) to approximate  $\lambda$  using the fact from (c). Write “ $\lambda \approx \text{something}$ ” below.
- (f) [easy] Now, build a general model for number of annual pedestrian accidents. “ $X \sim \text{something}$ ” is all you need to write below but you need to use your answer from (e).
- (g) [harder] Fill in the following table for the r.v.  $X$  you built in (j). Use three decimal places ONLY. (That goes **double** for Ilya, Joe and others). If you are feeling really lazy, check out “`dpois(0:8, --)`” and “`ppois(0:8, --)`” in R where the dash is your answer to (e).

$x$	$p(x)$	$F(x)$
0		
1		
2		
3		
4		
5		
6		
7		
8		

- (h) [easy] Find the Quantile  $[X, 0.10]$ .

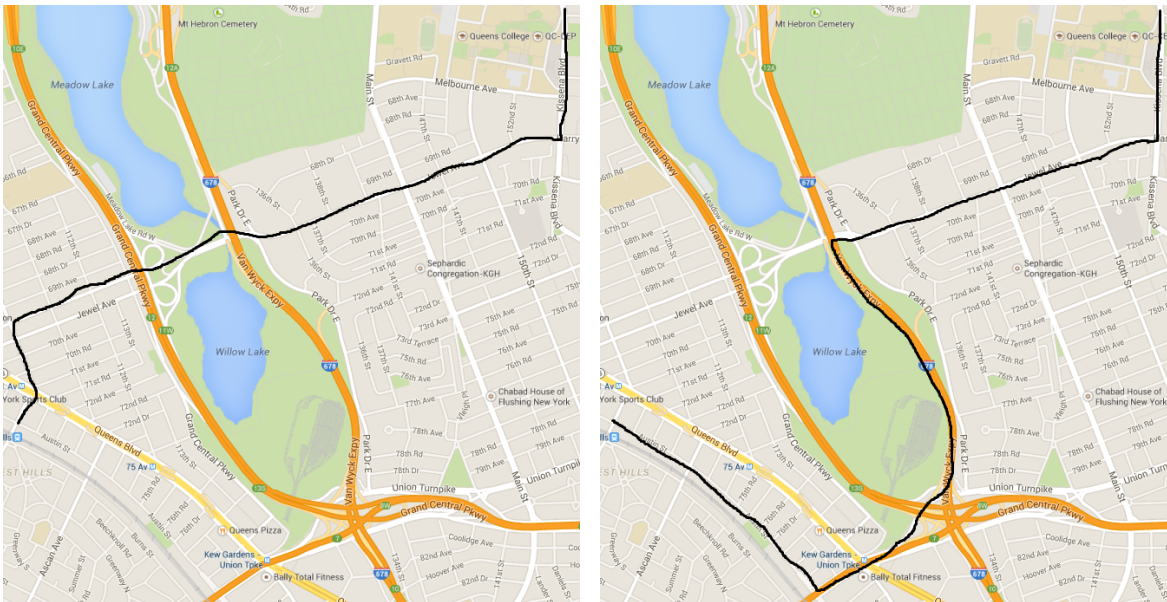
- (i) [easy] Find the 99%ile of  $X$ .
- (j) [easy] Find the median  $[X]$ .
- (k) [easy] Find the IQR  $[X]$ .
- (l) [easy] Find the Mode  $[X]$ .
- (m) [easy] Estimate the  $\mathbb{P}(X > 8)$  based on your calculation of  $F(8)$ .
- (n) [easy] Is it fair to say for a few draws of  $X$  the support essentially is just  $\{0, \dots, 8\}$ ?  
Write yes/no plus a sentence explanation.
- (o) [difficult] Using your answer from (m), what is the probability in *100 years* of data you see *more than 1 year* with the number of pedestrian accidents greater than 8?
- (p) [difficult] This is a preview of what's to come when we get to Statistics. Consider the hypothetical scenario where Queens Borough President Melinda Katz institutes new safety precautions for Queens Blvd. In 2015, the number of accidents drop to 0. Based on the table you completed in (k), can you say with assurance that this drop was due to President Katz's efforts? Why or why not? Justify using a number and a sentence *in English*.

### Problem 3

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. Thanks for helping me plan my budget!



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 4

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 5), 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 " $X \sim \text{something}$ " 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. Let  $X$  be the r.v. for how many flips you make. How is  $X$  distributed? Write " $X \sim \text{something}$ " below.
- (s) [easy] Using as fact that  $\mathbb{E}[X] = r/p$  when  $X \sim \text{NegBin}(r, p)$  (see Problem 5), 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  $\mathbb{E}[X]$ ? Note: I tried to make another question, but L<sup>A</sup>T<sub>E</sub>X crashed after (z).

## Problem 5

Extra credits as promised... Googable 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.]  $X \sim \text{NegBin}(r, p)$ . Verify  $\sum_{x \in \text{Supp}[X]} p(x) = 1$ .
- (c) [E.C.]  $X \sim \text{Poisson}(\lambda)$ . Verify  $\sum_{x \in \text{Supp}[X]} p(x) = 1$ .
- (d) [E.C.]  $X \sim \text{Binomial}(n, p)$ . Verify  $\mathbb{E}[X] = np$ .
- (e) [E.C.]  $X \sim \text{Hypergeometric}(n, K, N)$ . Verify  $\mathbb{E}[X] = n \frac{K}{N}$ .
- (f) [E.C.]  $X \sim \text{NegBin}(r, p)$ . Verify  $\mathbb{E}[X] = \frac{r}{p}$ .
- (g) [E.C.]  $X \sim \text{Poisson}(\lambda)$ . Verify  $\mathbb{E}[X] = \lambda$ .