

# MATH 241 Fall 2014 Homework #8

Professor Adam Kapelner

Due 5PM in my office, Tues Nov 4, 2014

(this document last updated Thursday 30<sup>th</sup> October, 2014 at 11:39am)

## Instructions and Philosophy

Once again, 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, please read the expectation section of Chapter 2. Avoid the parts that deal with “moment generating functions.”

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 for *extra credit* which are also marked “[E.C.]”. 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. Late homework will be penalized 10 points per day.

15 points are given as a bonus if the homework is typed using 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 is found on the syllabus. You may also use [writelatex.com](http://www.writelatex.com) which is a web service (you don’t have to install or configure anything on your local computer). 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 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. I STRONGLY recommend to write on a printout of this document since you will always have the questions handy to study from (and it is easier for me to grade accurately). 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

Queens Blvd. is known by some as the “Boulevard of Death” (you can read the Wikipedia information [here](#)).



From Wikipedia:

From 1993 to 2000, 72 pedestrians, were killed trying to cross the street, an average of 10 per year, with countless more injuries. Since 2000, at least partially in response to major news coverage of the dangerous road, the city government has taken measures to cut down on such incidents, including posting police along parts of the boulevard and doing spot-ticketing stings of jay-walkers against traffic, posting large signs proclaiming that “A Pedestrian Was Killed Crossing Here” at intersections where fatal accidents have occurred and installing more road-rule enforcement cameras.

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] Consider one year: 2001. Estimate the number of pedestrian crossings of Queens Blvd. in 2001. For those of you who ever interview for anything in business consulting, this is a good exercise. Note that this estimate does not use any of the data above nor any tools from my class that I can think of.

- (b) [easy] Based on your answer to (a) and the data in the first cell of Table 1, estimate the probability of a pedestrian accident on one crossing.
- (c) [easy] Considering the rather large answer in (a) and the rather small answer in (b), let  $X$  be the r.v. for number of pedestrian accidents in 2001. How is  $X$  distributed? What are its parameter(s)? I want the value of the parameter(s) not just a greek letter.
- (d) [harder] Write all the possible problems with the model you created in (c). Note: I will do something like this on the exam: build a model, tell me the problems with it, use the model.
- (e) [easy] We will now go a step further, consider the data above to be  $\overset{iid}{\sim}$  draws from one distribution — the distribution Poisson you created in (c). However, their parameters may now be different. Thus, the accidents for 2001 – 2007 can be thought of as:

$$X_{2001}, X_{2002}, \dots, X_{2007} \overset{iid}{\sim} f(x)$$

where the  $f(x)$  is the PMF of a model similar to the one you built in (c) but maybe not with the same  $\lambda$  parameter. Let:

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

Is  $\bar{x}$ , the sample average using the data in Table 1 a realization from the  $\bar{X}$  r.v.? Yes or no is fine.

- (f) [easy] If we had many more years of data than the 7 years, what would  $\bar{X}$  converge to eventually? You can use notation and not answer using a number.

- (g) [easy] What is the law we used in part (f)? Look in your notes. State the law and write a sentence about what it means to you *in English*.
- (h) [easy] Calculate  $\bar{x}$ , the sample average using the data in Table 1.
- (i) [E.C.] This may seem like a non-sequitor, but we'll need it for the next problem. Prove from first principles (*i.e.* from  $\mathbb{E}[X] := \sum_{x \in \text{Supp}[X]} x f(x)$ , the definition of expectation) that if  $X \sim \text{Poisson}(\lambda)$  then  $\mathbb{E}[X] = \lambda$ .
- (j) [easy] Using the fact that  $\mathbb{E}[X] = \lambda$  if  $X \sim \text{Poisson}(\lambda)$ , pretend 7 years of data is as good as infinite years of data and use your answer to (h) to approximate  $\lambda$  by setting  $\lambda = \bar{x}$ . Now, build a general model for number of annual pedestrian accidents. “ $X \sim$  something” is all you need to write below.

- (k) [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 (h).

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

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

- (m) [easy] Find the 99%ile of  $X$ .

- (n) [easy] Find the median  $[X]$ .

- (o) [easy] Find the IQR  $[X]$ .

- (p) [easy] Find the Mode  $[X]$ .

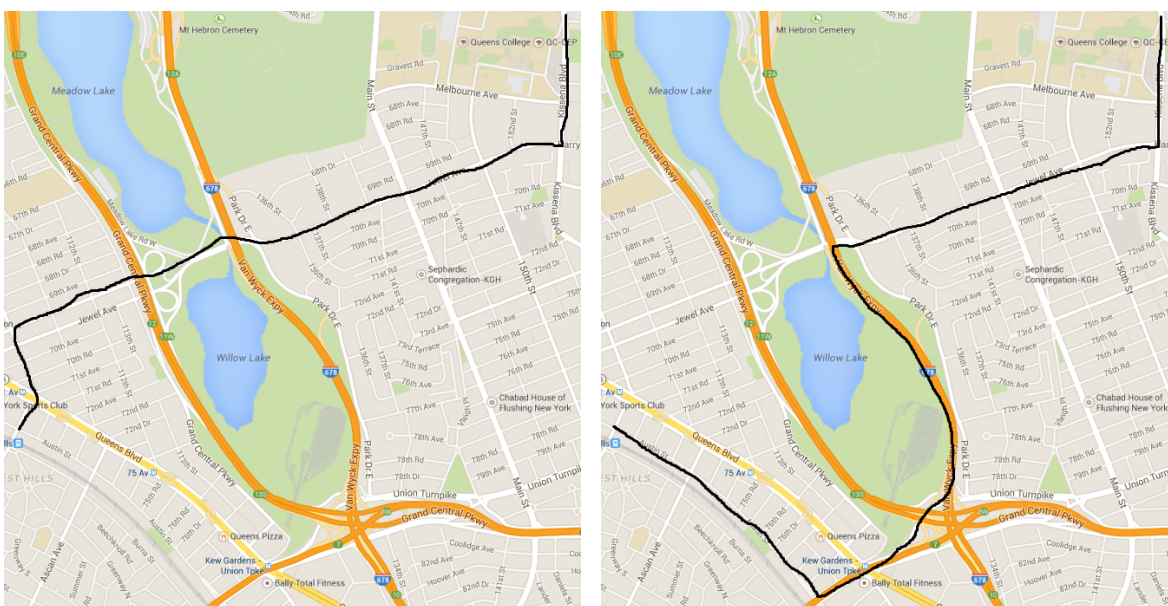
- (q) [easy] Estimate the  $\mathbb{P}(X > 8)$  based on your calculation of  $F(8)$ .
- (r) [easy] Is it fair to say for a few draws of  $X$  the support essentially is just  $\{0, \dots, 8\}$ ? Write yes/no plus a sentence explanation.
- (s) [difficult] What is the probability in 100 years of data you see more than 1 year with the number of pedestrian accidents greater than 8?
- (t) [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 2

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.
- (s) [harder] UberSUV dispatches an SUV and it is particularly useful if I'm transporting stuff between my office and apartment. The base fare is \$14 and they charge \$0.80/min and \$4.50/mi. Calculate  $\mathbb{E}[B]$  where  $B$  is now the total bill for UberSUV.

### Problem 3

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.”

- (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) [E.C.] For  $X \sim \text{Hypergeometric}(n, K, N)$ , show that  $\mathbb{E}[X] = n\frac{K}{N}$  from first principles (*i.e.* from  $\mathbb{E}[X] := \sum_{x \in \text{Supp}[X]} xf(x)$ , the definition of expectation). This is hard...
- (e) [easy] Using as fact that  $\mathbb{E}[X] = n\frac{K}{N}$  when  $X \sim \text{Hypergeometric}(n, K, N)$ , calculate  $\mathbb{E}[X]$  for the r.v. you constructed in part (c).

- (f) [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.
- (g) [easy] Find  $\bar{x}$  from the data you recorded in part (e).
- (h) [easy] Is  $\bar{x} \approx \mathbb{E}[X]$ ? If not, what could you change in the experiment to make  $\bar{x}$  closer to  $\mathbb{E}[X]$ ?
- (i) [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.
- (j) [difficult] For  $X \sim \text{Binomial}(n, p)$ , show that  $\mathbb{E}[X] = np$  from first principles (*i.e.* from  $\mathbb{E}[X] := \sum_{x \in \text{Supp}[X]} xf(x)$ , the definition of expectation). Try to do this yourself and if you give up, you can copy this from your notes.

- (k) [easy] Using as fact 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).
- (l) [easy] Shake the cup and count the number of heads. Repeat this five times. Record your data below.
- (m) [easy] Find  $\bar{x}$  from the data you recorded in part (l).
- (n) [easy] Is  $\bar{x} \approx \mathbb{E}[X]$ ? If not, what could you change in the experiment to make  $\bar{x}$  closer to  $\mathbb{E}[X]$ ?
- (o) [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.
- (p) [E.C.] For  $X \sim \text{Geometric}(p)$ , show that  $\mathbb{E}[X] = 1/p$  from first principles (*i.e.* from  $\mathbb{E}[X] := \sum_{x \in \text{Supp}[X]} xf(x)$ , the definition of expectation).

- (q) [easy] Using as fact 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).
- (r) [easy] Flip until you get a head. Repeat this five times. Record your data below.
- (s) [easy] Find  $\bar{x}$  from the data you recorded in part (r).
- (t) [easy] Is  $\bar{x} \approx \mathbb{E}[X]$ ? If not, what could you change in the experiment to make  $\bar{x}$  closer to  $\mathbb{E}[X]$ ?
- (u) [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.
- (v) [E.C.] For  $X \sim \text{NegBin}(r, p)$ , show that  $\mathbb{E}[X] = r/p$  from first principles (*i.e.* from  $\mathbb{E}[X] := \sum_{x \in \text{Supp}[X]} xf(x)$ , the definition of expectation). You need to use a trick. It is related to the trick from last week's extra credit problem.

- (w) [easy] Using as fact that  $\mathbb{E}[X] = r/p$  when  $X \sim \text{NegBin}(r, p)$ , calculate  $\mathbb{E}[X]$  for the r.v. you constructed in part (u).
- (x) [easy] Flip until you get two heads. Repeat this five times. Record your data below.
- (y) [easy] Find  $\bar{x}$  from the data you recorded in part (x).
- (z) [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).