

MATH 241 Fall 2016 Homework #6

Professor Adam Kapelner

Due *in class*, Friday, November 11, 2016

(this document last updated Tuesday 1st November, 2016 at 6:33pm)

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 the section about the Geometric and Negative Binomial r.v.’s, expectation and variance as well as expectation and variance of linear transformations and sums of r.v.’s in Ross.

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 L^AT_EX. Links to installing L^AT_EX and program for compiling L^AT_EX is found on the syllabus. You are encouraged to use overleaf.com. If you are handing in homework this way, upload `hwxx.tex` and `preamble.tex`, 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. If not using L^AT_EX, print this document and write in your answers. I do not accept homeworks not on this printout. Keep this first page printed for your records. Write your name and section below (A or B).

NAME: _____ SECTION (A or B): _____

More Random Variables We will continue to look at the binomial, geometric and its generalization, the negative binomial. And we will cover expectation and variance as well.

Problem 1

This is the fun part of the homework. You're going to repeat the experiments we did in class. Let's make some data!



- (a) [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 “ $X \sim \text{something}$ ” below.
- (b) [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 (a).
- (c) [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.
- (d) [easy] Find \bar{x} from the data you recorded in part (c).

- (e) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make \bar{x} closer to $\mathbb{E}[X]$?
- (f) [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.
- (g) [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 (f).
- (h) [easy] Shake the cup and count the number of heads. Repeat this five times. Record your data below.
- (i) [easy] Find \bar{x} from the data you recorded in part (h).
- (j) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make \bar{x} closer to $\mathbb{E}[X]$?
- (k) [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.
- (l) [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 (k).

- (m) [easy] Flip until you get a head. Repeat this five times. Record your data below.
- (n) [easy] Find \bar{x} from the data you recorded in part (m).
- (o) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make \bar{x} closer to $\mathbb{E}[X]$?
- (p) [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.
- (q) [E.C.] $X \sim \text{NegBin}(r, p)$. Verify $\mathbb{E}[X] = \frac{r}{p}$. It is similar to the expectation proof for the Geometric.
- (r) [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 (p).
- (s) [easy] Flip until you get two heads. Repeat this five times. Record your data below.
- (t) [easy] Find \bar{x} from the data you recorded in part (s).
- (u) [easy] Is $\bar{x} \approx \mathbb{E}[X]$? If not, what could you change in the experiment to make \bar{x} closer to $\mathbb{E}[X]$?

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

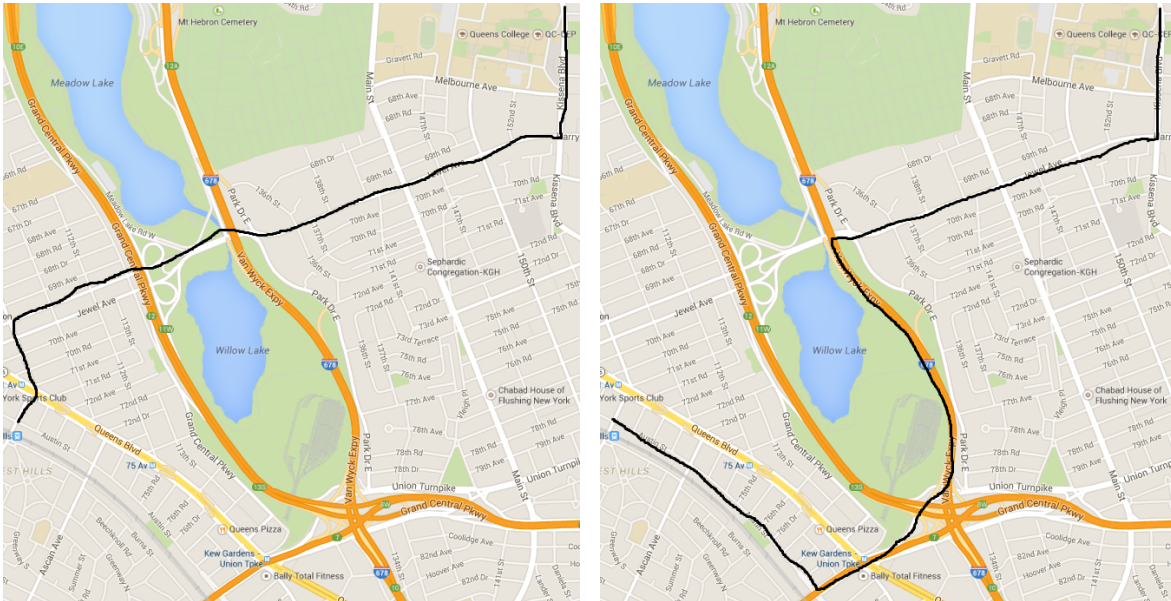
(w) [E.C.] $X \sim \text{Hypergeometric}(n, K, N)$. Verify $\mathbb{E}[X] = n \frac{K}{N}$ from the definition of $\mathbb{E}[X]$.

Problem 2

We are going to return to our in-class discussion of the my ride from QC to 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. Thus, some numbers have changed from our in-class discussion.



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 13 min and the Van Wyck route is 8 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 \$0.35/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 \$1.75/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 \$2.55. Let B be the r.v. which models the total bill for my uberX ride. Write B as a function of W and D .

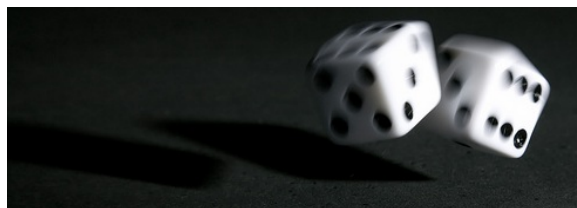
(p) [harder] We didn't really cover this in class, but you should be able to do it. W and D are one-to-one so the scaled W and scaled D sum is really one r.v. Find $\mathbb{E}[B]$ based also on the formula we learned in class about the expectation of a r.v. with a constant added.

(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 3

Imagine rolling two fair dice (no sorcery). Let X_1 be the r.v. corresponding to the first die and let X_2 be the r.v. corresponding to the second die. Let the outcome results be \$1 if you roll a 1, \$2 if you roll a 2, ..., and \$6 if you roll a six.



(a) [easy] What brand name r.v. is X_1 distributed as? Write $X \sim$ something and make sure the parameters are correct.

- (b) [easy] Does $X_1 \stackrel{d}{=} X_2$? Yes or no is fine.
- (c) [easy] Are X_1 and X_2 independent? Yes or no is fine.
- (d) [easy] Compute $\mathbb{E}[X_1]$ from first principles.
- (e) [easy] Compute $\text{Var}[X_2]$ from first principles.
- (f) [easy] The standard deviation is also called “standard error” and it sometimes denoted “SE.” Use your answer in (e) to find $\text{SE}[X_i]$ for $i \in \{1, 2\}$. Please just use the square root and do not rederive the variance again from scratch.
- (g) [easy] Draw the PMF for X_i for $i \in \{1, 2\}$ and mark $\mathbb{E}[X_i]$ and $\text{SE}[X_i]$ on the graph similar to how we did in class.

- (h) [easy] Imagine the game where you just double the winnings of a single roll. This would be equivalent to just multiplying the r.v. by a scale factor of 2. Calculate $\mathbb{E}[2X_i]$, $\text{Var}[2X_i]$ and $\text{SE}[2X_i]$ from the formulas we learned in class.
- (i) [easy] Draw the PMF for $2X_i$ for $i \in \{1, 2\}$ and mark $\mathbb{E}[2X_i]$ and $\text{SE}[2X_i]$ that you calculated in (h) on the graph.
- (j) [difficult] Draw the PMF for $X_1 + X_2$. This involves taking a convolution. Since convolution won't be on the midterm or final, I'm going to give a hint. There is 1 way to get 2 or 12, 2 ways to get 3 or 11, 3 ways to get 4 or 10, 4 ways to get 5 or 9, 5 ways to get 6 or 8 and 6 ways to get 7.

(k) [easy] Calculate $\mathbb{E}[X_1 + X_2]$, $\text{Var}[X_1 + X_2]$ and $\text{SE}[X_1 + X_2]$ from the formulas we learned in class. Do not use the PMF from the last question; use the formulas from class.

(l) [difficult] Why are the standard errors in (h) and (k) different and why is (h) larger? This involves a lot of thinking and I want a few sentences *in English*.

(m) [easy] Imagine the general case of $X_1, \dots, X_n \stackrel{iid}{\sim}$ with mean μ and variance σ^2 . Define \bar{X} as we did in class. Redo the derivation of $\mathbb{E}[\bar{X}] = \mu$.

(n) [easy] Imagine the general case of $X_1, \dots, X_n \stackrel{iid}{\sim}$ with mean μ and variance σ^2 . Define \bar{X} as we did in class. Redo the derivation of $\text{SE}[\bar{X}] = \sigma/\sqrt{n}$.

- (o) [easy] Calculate $\mathbb{E}[\bar{X}_n]$, $\text{Var}[\bar{X}_n]$ and $\text{SE}[\bar{X}_n]$ using the definition of \bar{X}_n we learned in class only as a function of n . Hint: use the formula you just derived.
- (p) [difficult] Imagine n rolls of the dice to produce n r.v.'s denoted X_1, \dots, X_n which of course are still $\overset{iid}{\sim}$. Calculate $\mathbb{E}[X_1 + \dots + X_n]$, $\text{Var}[X_1 + \dots + X_n]$ and $\text{SE}[X_1 + \dots + X_n]$ only as a function of n .
- (q) [difficult] What does it mean that $\mathbb{E}[\bar{X}_n]$ is an unbiased estimator for μ ? Explain in a few sentences *in English* why this is a good thing
- (r) [easy] If $n = 1000$, what is $\text{SE}[\bar{X}_n]$? Does that mean it's getting really close to $\mathbb{E}[\bar{X}_n]$? Why or why not.
- (s) [difficult] Now you have the choice between game A — where you roll n times and average the winnings (*i.e.* you collect \bar{X}_n dollars at the end) or game B — where you roll one die and collect the amount you make on just one roll. Use your answers to the relevant previous questions (I won't tell you which ones explicitly) to explain why you would choose game A over B or vice versa. I want multiple sentences *in English*. You must convince me you understand the tradeoff that game A and B are making.

- (t) [easy] Redo the derivation of the rule $\text{Var}[aX] = a^2 \text{Var}[X]$ for any discrete r.v. X and any constant $a \in \mathbb{R}$.

- (u) [harder] Let Z be the standardized r.v. for \bar{X}_n . Standardization of a r.v. X is defined as subtracting its mean and dividing by its standard error. For \bar{X} this would be:

$$Z := \frac{\bar{X} - \mathbb{E}[\bar{X}]}{\text{SE}[\bar{X}]} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Prove from the formulas in class that $\mathbb{E}[Z] = 0$ and $\text{Var}[Z] = \text{SE}[Z] = 1$. Hint: use those two rules about $\text{Var}[aX]$ and $\text{Var}[X + c]$ you just rederived

- (v) [easy] Why is “standardization” called “standardization”?

- (w) [harder] If $n = 1000$ and you made $\bar{x} = \$4.00$, what is the z -score of this \bar{x} ? That is if \bar{X}_n was standardized into the r.v. Z (as in the previous question), what would be the corresponding realization of z that corresponds to this \bar{x} ?

- (x) [easy] We will learn later in Math 241 that $z \notin [-3, 3]$ are very strange and smack of something being awry. Is something awry with making \$4.00 on average? Explain using a sentence *in English*.

- (y) [harder] Returning to X , the dice game in the beginning of the problem (the outcome results being \$1 if you roll a 1, \$2 if you roll a 2, ..., and \$6 if you roll a six), you calculated variance using the definition $\text{Var}[X] := \mathbb{E}[(X - \mu)^2]$ assuming the classic squared error loss: $e(x, \mu) := (x - \mu)^2$. Imagine we defined a new variance metric using absolute loss, $e(x, \mu) := |x - \mu|$. We'll denote this "new variance" with a big squiggly symbol, $\widetilde{\text{Var}}[X] := \mathbb{E}[|X - \mu|]$ just to make sure you don't confuse it with the standard definition of $\text{Var}[X]$. Calculate $\widetilde{\text{Var}}[X]$ and include units.

Problem 4

More simple r.v. practice.

- (a) [harder] Let $X_1 \sim \text{Bernoulli}(p)$. Derive an expression for $\text{Var}[X_1]$ as a function of the parameter as we did in class.
- (b) [easy] You know that T_n is the sum of $n \stackrel{iid}{\sim}$ bernoulli r.v.s with parameter p . Show that $\text{Var}[T_n]$ can be easily derived using the variance-sum formula we learned in class.

- (c) [harder] If you had complete control of both parameters n and p , what would be the easiest manipulation to make the variance of T_n as small as possible?
- (d) [easy] Prove $\text{Var}[X] = \mathbb{E}[X^2] - \mu^2$ for any r.v. X (this is in your notes).
- (e) [easy] Show from the definition of variance that $\text{Var}[X] \geq 0$ for any r.v. X .
- (f) [easy] Show that if $\text{Var}[X] = 0$, then X must be degenerate.
- (g) [E.C.] Show that if expected error for any $e(x, \mu)$ is 0, then X must be degenerate.
- (h) [difficult] Show for any two r.v.'s X and Y which are independent that $\text{Var}[X \times Y] = \mu_X^2 \sigma_Y^2 + \mu_Y^2 \sigma_X^2 + \sigma_X^2 \sigma_Y^2$. Remember, two r.v.'s multiplied together is a new r.v., $g(X, Y)$.

- (i) [easy] Let a_1, a_2, \dots, a_n be a sequence of constants. Let X_1, \dots, X_n be a sequence of r.v.'s which thereby share the same mean μ . Create a simplified expression for $\mathbb{E}[a_1X_1 + \dots + a_nX_n]$. I want the simplest combination of symbols a_1, a_2, \dots, a_n and μ .
- (j) [harder] Let a_1, a_2, \dots, a_n be a sequence of constants. Assume X_1, \dots, X_n are a sequence of $\stackrel{iid}{\sim}$ r.v.'s which thereby share the same variance σ^2 . Create a simplified expression for $\text{SE}[a_1X_1 + \dots + a_nX_n]$. I want the simplest combination of symbols a_1, a_2, \dots, a_n and μ .
- (k) [harder] Imagine a r.v. X with PMF $p(x) = c/x^2$ and $\text{Supp}[X] = \mathbb{N}$. What is the exact value of c which makes $p(x)$ a valid PMF? The answer can be found [here](#).
- (l) [E.C.] X is the same as in the previous problem. Show that $\text{Var}[X]$ does not exist (which means it's not a real number). You will need a fact from that same wikipedia page that you visited in the last problem. And now you've learned what the harmonic series is too.

(m) [difficult] Let $Y := X^2$ where X is a r.v. with PMF unknown. Use the notation μ for $\mathbb{E}[X]$, σ^2 for $\text{Var}[X]$, μ_3 for $\mathbb{E}[X^3]$ and μ_4 for $\mathbb{E}[X^4]$. Assume μ , σ^2 , μ_3 , μ_4 all exist in \mathbb{R} . Find $\text{Var}[Y]$ as a function of μ , σ^2 , μ_3 and μ_4 . This was the extra credit on last semester's midterm, thus the solution is online. Try to do it yourself first.

(n) [difficult] Consider $X \sim \text{NegBin}(r, p)$. Prove that $\text{Var}[X] = r(1-p)/p^2$ assuming that the variance of a geometric r.v. with parameter p is $(1-p)/p^2$.

(o) [difficult] Consider $X \sim \text{Hypergeometric}(n, K, N)$. Prove that $\mathbb{E}[X] = n\frac{K}{N}$ using the rules of expectation.

(p) [harder] Prove the memorylessness property of $X \sim \text{Geometric}(p)$ by proving it holds for the definition of memorylessness i.e. $\mathbb{P}(X > x) = \mathbb{P}(X > x_0 + x \mid X > x_0)$.