MATH 241 Fall 2016 Homework #5

Professor Adam Kapelner

Due 5PM KY604, Tuesday, October 25, 2016

(this document last updated Thursday 13th October, 2016 at 8:20pm)

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 Bernoulli, Binomial, Geometric and Negative Binomial 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 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, 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 LATEX, 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 look at the binomial, geometric and its generalization, the negative binomial

Problem 1

We will continue to look at the binomial.

(a) [easy] Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (p). Give a real-life example of this situation

(b) [easy] Let $T_n = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (p). How is T_n distributed?

(c) [difficult] Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (p) and $T_n = \sum_{i=1}^{n} X_i$. Derive the distribution of T_n from first principles just like we did in the notes.

(d) [difficult] Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (p) and $T_n = \sum_{i=1}^{n} X_i$. Derive the distribution of T_n from first principles just like we did in the notes.

Problem 2

Imagine two Bernoulli r.v.'s X_1 and X_2 which model two fair coin flips where Heads is mapped to 1 and tails is mapped to 0. The probability of heads is $1/2$.

(a) [easy] Given no other information, explain using the definition of r.v. independence why these two r.v.'s are independent.

(b) [easy] Given no other information, explain using the definition of equality in distribution why $X_1 \overset{d}{=} X_2$.

(c) [easy] Are $X_1, X_2 \overset{iid}{\sim}$ Bernoulli (p)?

(d) [harder] Now imagine these two coins were linked using some sort of sorcery. They are flipped at the same time but are guaranteed to flip the same way. That is, if the first coin goes heads, the second coin must go heads (and if the first coin goes tails, the second coin must go tails).

Draw a probability tree of this scenario.

(e) [harder] Explain using the definition of r.v. independence why these two r.v.'s are *dependent*.

(f) [difficult] Using the same two sorcery-controlled coins, explain using the definition of equality in distribution why or why not $X_1 \overset{d}{=} X_2$.

(g) [easy] Are $X_1, X_2 \overset{iid}{\sim}$ Bernoulli (p) if they are modeled by these two sorcery-controlled coins? Yes / no.

(h) [easy] Let $T_2 = X_1 + X_2$. Is $T_2 \sim$ Binomial $\left(2, \frac{1}{2}\right)$? Why or why not?

Problem 3

Imagine you are flipping the same bundle of coins from the practice midterm. The probability of the coin bundle landing on its side is $\mathbb{P}(S) = 1/11$. Heads and tail probability are $5/11$. Let's call landing on its side a "success."

(a) [easy] I flip the coin bundle once. Model a success as a "1." Show that the r.v. modeling this event outcome is Bernoulli and define its parameter. Write "$X \sim$" something below.

(b) [easy] Let's say we flip 10 times. What is the probability that we get one (and only one) success? I want to see a probability model. Write "$X \sim$" something below. Then I want to see a probability statement. Then I want to see a computation. Answer then in decimal rounded to two digits.

(c) [easy] Let's say we flip 10 times. What is the probability that we get 5 (and only 5) successes?

(d) [easy] Let's say we flip 10 times. What is the probability that we get 8 (and only 8) successes?

(e) [harder] Let's say we flip 10 times. What is the probability we get one or two successes?

(f) [difficult] Let's say we flip 10 times. What is the probability we get 3 or less successes?

Problem 4

Imagine you are playing roulette again this time in America. The probability of winning a bet on black is 18/38. Call this a "success."

(a) [easy] Let's say we spin 15 times. What is the probability that we get 10 successes?

(b) [harder] Let's say we spin 30 times. Write a summation expression for getting 15 or more successes. Do not compute the answer explicitly.

(c) [difficult] Preview of statistics. You are now the casino floor manager for roulette. You witness 40 spins and it comes out black 18 times. Is this a "weird" or "unexpected" outcome? Explain using a calculation and a few sentences *in English*.

(d) [difficult] You witness 40 spins and it comes out black 38 times. Is this a "weird" or "unexpected" outcome? Explain using a calculation and a few sentences *in English*.

(e) [E.C.] You witness 40 spins. How many times should black occur "normally?" At which large values of number of blacks do get concerned by? At which small values of number of blacks do you get concerned by?

Problem 5

Now that we understand both the binomial and the concept of $\overset{iid}{\sim}$, we will ask some conceptual questions.

(a) [harder] The human mouth has 32 teeth. If the probability of a cavity at some point in a lifetime is 5%, is it possible to calculate the probability of 7 cavities during a lifetime using a binomial r.v. model $X \sim \text{Binomial}\,(32,\,5\%)$ and computing $\mathbb{P}\,(X = 7)$? Why or why not?

(b) [harder] Bob drives his car twice a day during the workweek, 50 weeks per year for a total of 200 work days and thus, 400 commuting rides. If the probability of getting into an accident is about 1/1000, could we say the number of accidents per year can be modeled as a r.v. $X \sim \text{Binomial}(400, 1/1000)$? Why or why not?

Problem 6

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 good (c.f. a bad avocado which may have brown inside because it's partially rotten). Call this probability of good p. Model the number of *good* avocados you have using a random variable. All you need to write is $X \sim$ something. You do not need to write the PMF, draw the PMF, draw the CDF, etc.

(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. 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. $\sum$) 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) [difficult] 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. Be careful between "rotten" and "good".

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

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

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

(k) [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.

(l) [harder] Let's say at some point in your avocado shopping that you learned how to detect rotten avocadoes by using the "squeeze test" and you used this learning to select new avocados. What assumption(s) would be violated?

(m) [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(s) would be violated?

(n) [harder] Grocery stores usually put the old avocados on the top of the avocado basked and thus the new avocados on the bottom. If your strategy was just to pick the "top" avocado each time, what assumption(s) would be violated?

(o) [harder] Sometimes avocados are sold in mesh bags together in a 6-pack which are taken from the same farm even the same tree. If your strategy for picking is picking a basket all at once, examining each of the 6 avocados and then moving on to a new bunch, what assumption(s) would be violated?

Problem 7

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.

(k) [harder] How did the negative binomial r.v. get its name?

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.
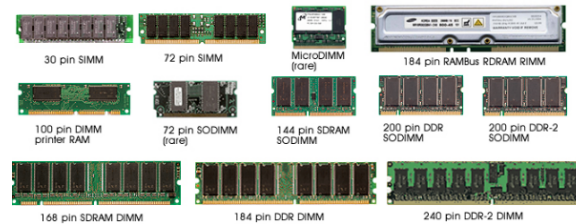
(a) [easy] What is the probability you get three faulty RAM chips in a row?

(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 9

Some simple definitional questions.

(a) [easy] What is a realization of a r.v.?

(b) [easy] Define "datum".

(c) [easy] Define "data".

(d) [easy] In a given r.v. model X, what set do they data belong to?

(e) [easy] Define "iid data".