MATH 241 Fall 2015 Homework #6

Professor Adam Kapelner

Due 5PM in my mail slot, Friday, October 23, 2015

(this document last updated Monday 19th October, 2015 at 5:57pm)

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 bernoulli, hypergeometric and binomial .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): _____

Hypergeometric Distribution   This is a very interesting random variable and we are covering it thoroughly between this homework and the previous.

Problem 1

Generally, the hypergeometric has three parameters. We will solve for its support here under several disjoint conditions and then in class we will generalize it. Call X a hypergeometric r.v. with all its parameters free - meaning they can take on any value, so please use the notation $n,\ K,\ N$ in your answers as we did in class.

(a) [easy] Using the usual parameterization of the hypergeometric, describe the parameter space. You need to say what sets each of the parameters "lives" in.

(b) [easy] Write, but do not draw, the PMF of X.

(c) [harder] x is the free variable in $p(x)$ which you wrote in (b) and it designates the number of successes. Show that successes and failure are essentiall the same thing by finding $p(n-x)$ and replacing K with $N-K$. What does this teach you?

(d) [harder] Let's say $n < K$ and $n < N - K$. What is the support of X in this situation?

2

(e) [harder] Let's say $n < K$ and $n \geq N - K$. What is the support of X in this situation?

(f) [harder] Let's say $n \geq K$ and $n < N - K$. What is the support of X in this situation?

(g) [difficult] Let's say $n \geq K$ and $n \geq N - K$. What is the support of X in this situation?

(h) [E.C.] Describe the CDF of the general hypergeometric r.v.

Problem 2

We will look at hypergeometric distributions with large N. If N is really large, sampling

without replacement can be approximated by sampling with replacement. In the limit, it is sampling with replacement.

Parts (a-e) are from HW #5 so I've filled in the answers already for you. Other answers are filled in as well

(a) [easy] We will now begin deriving the binomial in pieces. Parameterize a hypergeometric by setting $K = pN$. What is the parameter space for p?

(b) [easy] Write the PMF $p(x)$ for this r.v. using the p parameterization using x as the free variable.

(c) [easy] What limit do we take and why are we taking this limit?

(d) [easy] Rewrite the PMF without choose notation using only factorials and simplify the fraction by moving the factorial terms from denominator, $\binom{N}{n}$, to the numerator.

$$p(x) = \lim_{N \to \infty} \frac{(pN)!}{x!(pN-x)!} \frac{((1-p)N)!}{(n-x)!((1-p)N-n+x)!} \frac{n!(N-n)!}{N!}$$

(e) [easy] Which three terms can you factor out from the limit expression? Show that they are equivalent to $\binom{n}{x}$.

(f) [difficult] Within the limit, you now have three ratios. Write these ratios by canceling out the common terms. For instance $10!/6! = 10 \times 9 \times 8 \times 7$ and $6!/10! = 1/(10 \times 9 \times 8 \times 7)$. You have to get the indexing right.

(g) [harder] How many terms are in the numerator? How many terms are in the denominator.

There are x terms in the first piece of the numerator and $n-x$ terms in the second piece of the denominator for a total of n terms. Then there are n terms in the denominator.

(h) [harder] Reason in English that the denominator looks like a bunch of $N - c_i$ where the c_i's are all constants which are negligible as $N \to \infty$.

(i) [harder] Reason in English that the numerator looks like a bunch of $Np - c_i$ where the c_i's are all constants which are negligible as $N \to \infty$ as well as a bunch of $(1-p)N - c_i$ where the c_i's are all constants which are negligible as $N \to \infty$.

(j) [harder] Match each $Np - c_i$ term in the numerator to one $N - c_i$ term in the denominator and take the limit of each one individually. Show that you wind up with

5

$p \times p \times \ldots$ for a total of x times, i.e. p^x.

(k) [harder] Match each $(1 - p)N - c_i$ term in the numerator to one $N - c_i$ term in the denominator and take the limit of each one individually. Show that you wind up with $(1 - p) \times (1 - p) \times \ldots$ for a total of $n - x$ times, i.e. $(1 - p)^{n-x}$.

(l) [easy] Using your answers from the previous parts to write the binomial r.v.'s PMF.

(m) [E.C.] Imagine you are drawing n successes with and without replacement. Derive a function of N and n which gives this percentage difference for N and n generally when the number of successes $K = \frac{1}{2}N$.
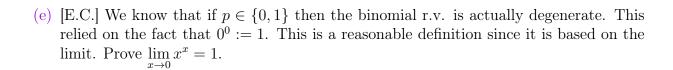
We will now look at the binomial in general.

(a) [easy] Show that success and failure is arbitrary by letting the number of successes x equal the number of failures $n-x$ and the probability of success p equals the probability of failure $1 - p$ using the PMF of the binomial distribution.

(b) [harder] Show using the definition of equals in distribution that $X_1 \overset{d}{=} X_2$ if $X_1 \sim$ Bernoulli (p) and $X_2 \sim$ Binomial $(1,\ p)$.

(c) [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?

(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.

(e) [E.C.] We know that if $p \in \{0, 1\}$ then the binomial r.v. is actually degenerate. This relied on the fact that $0^0 := 1$. This is a reasonable definition since it is based on the limit. Prove $\lim_{x \to 0} x^x = 1$.

Independence and equality of distribution of r.v.'s   Since we haven't covered much else, this majority of this assignment will be about this distribution.

Problem 4

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).

Explain using the definition of r.v. independence why these two r.v.'s are *dependent*.

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

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

Problem 5

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? That is, solve for $\mathbb{P}(X \leq 3) = F(3)$.
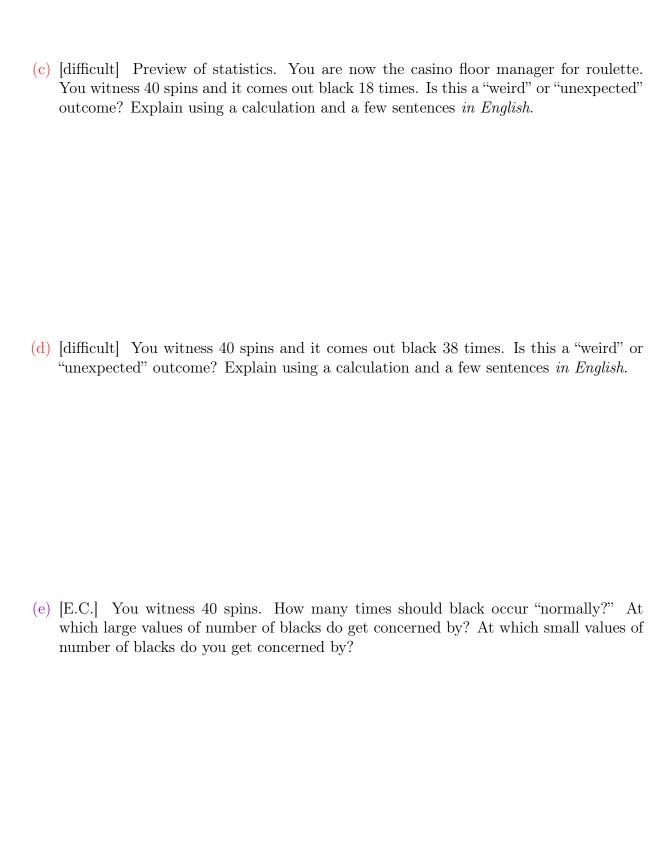
Problem 6

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?

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

(a) [harder] Recall X_1, X_2 from problem 1(d) which were the two sorcery-controlled coins. Let $T_2 = X_1 + X_2$. Is $T_2 \sim$ Binomial $\left(2, \frac{1}{2}\right)$? Why or why not?

(b) [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 $X \sim$ Binomial $(32, 5\%)$ and computing $\mathbb{P}(X = 7)$? Why or why not?