

MATH 241 Fall 2014 Homework #5

Professor Adam Kapelner

Due 5PM in my office, Tues Oct 14, 2014

(this document last updated Sunday 12th October, 2014 at 2:25pm)

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**. This one is a short workout.

Reading is still *required*. For this homework set, please read the hypergeometric variable section of Chapter 2.

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 up to a maximum of three days.

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 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. You are free to handwrite on separate paper if you wish but 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): _____

Hypergeometric Distribution Since we haven't covered much else, this majority of this assignment will be about this distribution.

Problem 1

The hypergeometric is sampling “without replacement.” Imagine you have this bag of marbles with 37 marbles and 17 of them are black. We will define a “success” as drawing a black marble.



- (a) [easy] Let's say you draw one marble. Call this r.v. X . Is it hypergeometric?
- (b) [easy] The hypergeometric distribution has three parameters. What are the parameters for X ?
- (c) [easy] Write, but do not draw, the PDF, $f(x)$ for the r.v. X where x is the number of successes.
- (d) [easy] What is the support of this r.v.?

- (e) [harder] There is another variable we learned about in class with this same support. Show that X is distributed as this type of r.v. and find its parameter(s).
- (f) [easy] Now imagine you draw 4 marbles without replacement. Call this r.v. X (and forget about the previous r.v. X from this question, parts a-e). How is X distributed? Use the notation in class and find its parameters.
- (g) [easy] What is the support of X ?
- (h) [easy] Write, but do not draw, the PMF of X .
- (i) [easy] Draw the PMF of X .

(j) [easy] Draw the CDF of X .

(k) [easy] What is the probability of getting 4 successes in a row? Use the PMF.

(l) [easy] What is the probability of getting 4 successes in a row? Use conditional probability. This should yield the same answer.

(m) [easy] Now imagine you draw 27 marbles without replacement. Call this r.v. X (and forget about the previous r.v. X). How is X distributed? Use the notation in class and find its parameters.

(n) [easy] What is the support of X ? Why is $0 \notin \text{Supp}(X)$?

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

(p) [difficult] Find the mode of this distribution. “Mode” is defined as the most likely outcome result.

(q) [E.C.] Why is this distribution called “hypergeometric?” Supposedly Euler came up with the word. But why is it called this? Is it related to the geometric distribution? This is easy extra points for those who are good with Google and have a soft spot for history.

Problem 2

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 $f(x)$ which you wrote in (b) and it designates the number of successes. Show that successes and failure are essentially the same thing by finding $f(n - x)$ and replacing K with $N - K$. What does this teach you?

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

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

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

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

Problem 3

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.

(a) [easy] Imagine a bag of 1,000 marbles with 500 black marbles where black signifies success once again. You draw 3 marbles without replacement. What is the probability of getting all successes? Use the PMF of a hypergeometric distribution you create.

- (b) [easy] Now draw 3 marbles with replacement. What is the probability you get all successes?
- (c) [easy] Calculate the percent difference in the answer you got in (a) and the answer you got in (b). Why is it small?
- (d) [E.C.] Derive a function of N and n which gives this percentage difference for N and n generally when the number of successes $K = \frac{1}{2}N$.
- (e) [easy] We will now begin deriving the binomial in pieces. Parameterize a hypergeometric by setting $K = pN$. What is the parameter space for p ?

- (f) [easy] Write the PMF $f(x)$ for this r.v. using the p parameterization using x as the free variable.
- (g) [easy] What limit do we take and why are we taking this limit?
- (h) [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.
- (i) [easy] Which three terms can you factor out from the limit expression? Show that they are equivalent to $\binom{n}{x}$.

- (j) [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 $6!/10! = 1/(10 \times 9 \times 8 \times 7)$. This is difficult because you have to get the indexing right.
- (k) [harder] How many terms are in the numerator? How many terms are in the denominator.
- (l) [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 $N \rightarrow \infty$.
- (m) [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 $N \rightarrow \infty$ as well as a bunch of $N(1 - p) - c_i$ where the c_i 's are all constants which are negligible as $N \rightarrow \infty$.

- (n) [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 $p \times p \times \dots$ for a total of x times, i.e. p^x .
- (o) [harder] Match each $N(1 - p) - c_i$ 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 $(1 - p) \times (1 - p) \times \dots$ for a total of $n - x$ times, i.e. $(1 - p)^{n-x}$.
- (p) [easy] Using your answers from parts (i), (n) and (o) write the binomial r.v.'s PMF.
- (q) [easy] What is the support of this binomial r.v. you just derived?
- (r) [harder] What is a binomial r.v.? Use your own words.