MATH 241 Spring 2015 Homework #5

Professor Adam Kapelner

Due 5PM, Tuesday, Mar 17, 2015

(this document last updated Friday 13th March, 2015 at 3:52pm)

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*. But for this homework set, I can't find anything from the 7th edition of Ross except the first few pages of Chapter 4 that are "worth it" for you to read.

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]", 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. Late homework will be penalized 10 points per day.

Between 1–15 points are arbitrarily given as a bonus (conditional on quality) 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` (make sure you upload both the hwxx.tex and the preamble.tex file). 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 as a figure or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

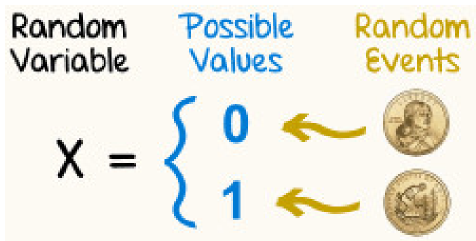
The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. **Handing it in without this printout is NO LONGER ACCEPTABLE.** Keep this page printed for your records. Write your name and section below where section A is if you're 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): ______

Random Variables We now begin question about the second unit of this class: r.v.'s.

Problem 1

In class we spoke about how random variables map outcomes from the sample space to a number *i.e.* $X : \Omega \to \mathbb{R}$. That is they are set functions, just like the probability function which is $\mathbb{P} : 2^{\Omega} \to [0, 1]$. We will be investigating this concept here.



(a) [easy] Here is a way to produce $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ using the Ω from a roll of a die. Map outcomes 1,2,3 to 0 and outcomes 4,5,6 to 1. This works because

$$\mathbb{P}(X = 0) = \mathbb{P}(\{\omega : X(\omega) = 0\}) = \mathbb{P}(\{1\} \cup \{2\} \cup \{3\}) = 1/2 \text{ and}$$
$$\mathbb{P}(X = 1) = \mathbb{P}(\{\omega : X(\omega) = 1\}) = \mathbb{P}(\{4\} \cup \{5\} \cup \{6\}) = 1/2.$$

Describe three other scenarios or devices that produce their own Ω's that also result in $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

(b) [harder] We talked about in class how the sample space no longer needs to be considered once the random variable is described. Why? Use your answer to (a) to inspire this answer. Write it *in English* below.

(c) [difficult] Back to philosophy... Let's say X models the price difference that IBM stock moves in one day of trading. For instance, if the stock closed yesterday at \$56.24 and today it closed at \$57.24, the random variable would be \$1 for today. According to our definition of a random variable, there is a sample space with outcomes being drawn ($\omega \in \Omega$) that is "controlling" the value of X. Describe it the best you can *in English*. There are no right or wrong answers here, but your answer must be coherent and demonstrate you understand the question.

Problem 2

We will now study probability mass functions (PMF's) denoted as $p(x)$ and cumulative distribution functions (CDF's) denoted as $F(X)$ and review the r.v.'s we did in class.

(a) [easy] Draw the PMF for $X \sim \text{Bernoulli}(p)$.

(b) [easy] Draw the CDF for $X \sim \text{Uniform}(\{1, 3, 4, 9\})$.

(c) [harder] Using the r.v. from the previous question, what is $\mathbb{P}(X \in (3,9))$? I am trying to trick you here.

(d) [difficult] Take a r.v. X with $\text{Supp}[X] = [0,1]$. Is this a "discrete r.v.?" Yes / no and explain. Hint: is $|[0,1]| \leq |\mathbb{N}|$?

(e) [difficult] In class we defined the Bernoulli r.v. as:

$$X \sim \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

and put its PMF on the board. Write $p(x)$ for $X \sim \text{Bernoulli}(p)$ that is only valid for not only all values in the $\text{Supp}[X]$ but all values in $\mathbb{R}$. Use the indicator function and set theory notation.

(f) [difficult] What is the parameter space of X where $X \sim \text{Bernoulli}(p)$ and why?

(g) [difficult] Sometimes knowing the Ω matters a little bit. Let's say $X_1 \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ is generated from one coin and $X_2 \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ is generated from another coin independently tossed. Create a new r.v. $T = X_1 + X_2$. Describe its PMF using the $\sim$

notation like in the previous problem. Thus write "$T \sim$" something.

(h) [difficult] Consider the PMF we discussed for $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. Does $\int p(x)\,\mathrm{d}x = F(x) + C$ where the constant $C \in \mathbb{R}$? Explain. Think carefully about what integration really means.

(i) [difficult] How about the opposite? Consider the CDF we discussed for $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. Does $\mathrm{d}/\mathrm{d}x[F(x)] = p(x)$? Explain. Think carefully about what differentiation really means.

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

Problem 3

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, $p(x)$ 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.

Problem 4

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 $n \leq K$ 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?

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

Problem 5

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] 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.

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