MATH 241 Fall 2014 Homework #7

Professor Adam Kapelner

Due 5PM in my office, Tues Oct 28, 2014

(this document last updated Wednesday 22$^{\text{nd}}$ October, 2014 at 12:17 Noon)

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.**

Reading is still *required*. For this homework set, please read the binomial, negative binomial variable, Poisson and expectation section of Chapter 2. Avoid the parts that deal with "moment generating functions."

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.

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 may also use `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. 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): ______

Review Everyone needs more practice with the r.v.'s we've been studying.

Problem 1

We will do a few short questions asking which r.v. to use and why. We will be investigating this by imagining a trip the grocery store to buy ingredients for guacamole.

(a) [easy] You buy an avocado at the grocery store which is mushy. Thus, it may have brown inside because it's partially rotten. Call this probability of rottenness p. Model the number of good avocados you have using a random variable. Call this r.v. X. Hint: the number of good avocadoes is either zero or one since you buy one and if it's good, you have one; if it's bad you have zero. All you need to write is $X \sim$ something. You do not need to write the PMF, draw the PMF, draw the CDF, nor contemplate the meaning of life in the next centimeter of white space.

(b) [easy] You buy 10 such mushy 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.

(c) [easy] Write the PMF for the r.v. you created in (b).

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

(e) [easy] Use the sigma notation for summing (e.g. $\sum_{i=1}^{5}$) 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.

(f) [easy] 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.

(g) [easy] Write the PMF for the r.v. you created in (f).

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

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

(j) [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.

(k) [easy] Now we keep picking avocados and are waiting until we get 4 rotten ones. Model the number of avocados you cut open using a r.v. Call this r.v. X.

(l) [easy] Write the PMF for the r.v. you created in (k).

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

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

(o) [easy] What is the probability you stop when looking at the second avocado? If this probability is zero, comment on why this is impossible.

(p) [easy] Use the sigma notation for summing (e.g. $\sum_{i=1}^{5}$) to calculate the probability that you stop after 8 avocados (not including 8). Since you don't know p you cannot actually compute a numerical value for this probability. Leave it in sigma notation.

(q) [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).

(r) [easy] Comment on why the r.v. you created in (k) is the sum of many $\overset{iid}{\sim}$ r.v.'s you modeled in (f).

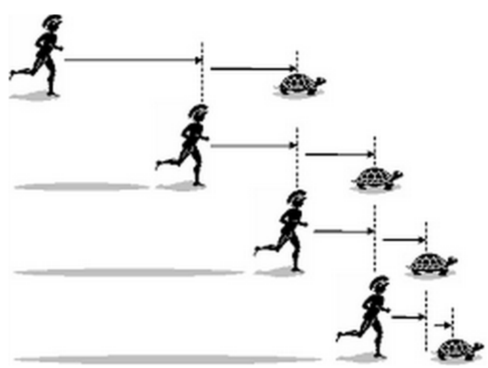
(s) [harder] Let's say you learned how to detect rotten avocadoes and you used this learning to select new avocados. What assumption is violated?

(t) [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 is violated now?

(u) [harder] If either (s) or (t) occurs, are any answers in your solutions to questions (b) - (r) correct?

Problem 2

A big problem is verifying that the PMF's add up to one over the support.



(a) [easy] $X \sim \text{Bernoulli}(p)$. Verify $\sum_{x \in \text{Supp}(X)} f(x) = 1$.

(b) [difficult] $X \sim \text{Binomial}(n, p)$. Verify $\sum_{x \in \text{Supp}(X)} f(x) = 1$. Use the trick from the class notes or something you demonstrated in a previous homework assignment. You do not need to write more than a couple sentences.

(c) [difficult] $X \sim \text{Geometric}(p)$. Verify $\sum_{x \in \text{Supp}(X)} f(x) = 1$. Use the proof in the class notes. No need to prove when the geometric series converges. Just write a comment about it.

(d) [E.C.] $X \sim \text{NegBin}(r, p)$. Verify $\sum_{x \in \text{Supp}(X)} f(x) = 1$. Hard but googlable.

(e) [E.C.] $X \sim \text{Hypergeometric}\,(n,\ K,\ N)$. Verify $\sum_{x \in \text{Supp}(X)} f(x) = 1$. I believe this must be done in cases.

(f) [E.C.] $X \sim \text{Poisson}\,(\lambda)$. Verify $\sum_{x \in \text{Supp}(X)} f(x) = 1$. The answer is in the last page of my scanned lecture PDF on github. We didn't get time to cover it in class, hence it will not be covered on any exams.

The Poisson We will be studying this r.v. now.

Problem 3

We will be investigating the situation where there are $n = 1500$ students who are Math, Economics, Accounting, Math Education and CS majors at Queens College Sophomore level and above. Assume these students qualify to take Math 241 given by yours truly.

> *MATH 241.* Introduction to Probability and Mathematical Statistics. 3 hr.; 3 cr.
> *Prereq. or coreq.:* MATH 132 or 143 or 152.
> An introduction to the basic concepts and techniques of probability and statistics with an emphasis on applications. Topics to be covered include the axioms of probability, combinatorial methods, conditional probability, discrete and continuous random variables and distributions, expectations, confidence interval estimations, and tests of hypotheses using the normal, t-, and chi-square distributions. *Students taking this course may not receive credit for MATH 114, except by permission of the chair. Not open to students who are taking or who have received credit for MATH 611.*

(a) [easy] Assume each student is ruggedly individualistic about their choice of classes. What assumption is this? One word is good.

(b) [easy] Assume each students has the same interest level in Probability and Statistics. What assumption is this? One word is good.

(c) [easy] The assumptions in (a) and (b) combine into which necessary assumption for using the binomial r.v. to model the number of students? One acronym is good.

(d) [easy] Assume the probability they enroll in my Math 241 is $p = 2\%$. Calculate the probability of 30 students enrolling in my Math 241 *exactly*. Do not use an approximation. This is the expected number of students to enroll (i.e. $1500 \times 2\% = 30$).

(e) [easy] This is a case where n is high and p is low. Let U be the r.v. that models the approximation we learned about in class. How is U distributed? Indicate your parameters clearly.

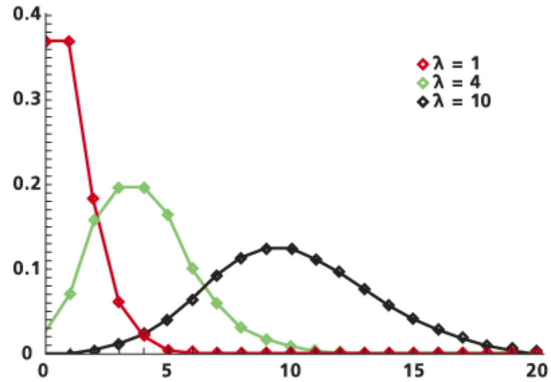
(f) [easy] The R code below will graph the PMF of $X \sim \text{Binomial}\,(1500, 2\%)$ in red and the PMF of U in green. Copy and paste the code into the console. If this doesn't work, download the PDF, open it with Adobe PDF reader and try again. I've tested this on Windows, MAC and Linux and it works on all three.

```
n = 1500;
p = 0.02;
support_max = 300;
bin = array(NA, support_max);
pois = array(NA, support_max);
for (i in 1 : support_max){
  bin[i] = dbinom(i, n, p);
  pois[i] = dpois(i, n * p);
}
plot(bin[1:60],
  pch = 16,
  col = "red",
  xlab = "Number of students enrolled in Math 241",
  ylab = "probability",
  main = "Class Enrollment Probabilities for Math 241");
points(pois[1:60], col = "green", pch = 16);
#placeholder for last line
```

Is U a good approximation of X? Determine visually from the plot as to why or why not? Print out the plot and attach it to your homework.

Problem 4

We will be deriving the Poisson PMF from the ground up using the class notes.



(a) [easy] What is the Poisson r.v.? Write a paragraph in your own words describing it.

(b) [easy] Use $\lambda = np$ to pin the relationship of $n \times p$. What limit do we take on n? What limit do we take on p?

(c) [easy] We need a limit to respect the relationship of $\lambda = np$. Use $n \to \infty$ for now and set $p = \lambda/n$. Write the PMF of the binomial using this substitution for p.

(d) [easy] Use your answer from (c), place the limit notation in front. Why are we taking this limit?

(e) [easy] Now, break open the combination term and the $(\lambda/n)^x$ term into atomic parts. Anything that is not a function of n factor out and write it in front of the limit operator.

(f) [harder] Now expand the $n!/(n-x)!$ term into many terms using the $\ldots$ notation. How many of these terms are there?

(g) [easy] Expand the n^x term into many terms. How many terms are there?

(h) [harder] Match the first term from (f) with the first term from (g). Match the second term from (f) with the second term from (g), etc through all the terms in (f) and (g). Take the limit of each of these terms.

(i) [easy] Now expand the $(1-\lambda/n)^{n-x}$ into two terms.

(j) [harder] Take the limit of the $(1-\lambda/n)^{-x}$ term. Do so carefully in stages.

(k) [harder] Using the fact that we derived in class, that $e^c = \lim_{n\to\infty}(1+c/n)^n$, solve for the limit of the $(1-\lambda/n)^n$ term.

(l) [easy] Combine your answers to parts (e), (h), (j) and (k) and write the PMF for the r.v. $X \sim \text{Poisson}(\lambda)$ below.

(m) [harder] What is the support of X?

(n) [harder] What is the paramter space of λ?