MATH 368/621 Fall 2020 Homework #5

Professor Adam Kapelner

Due by email noon, Friday, Nov 20, 2020

(this document last updated Wednesday 11$^{\text{th}}$ November, 2020 at 2:32pm)

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, review from math 241 about conditional probability, expectation and variance then read on your own about PMF transformations, the family of gamma functions, the negative binomial, poisson, exponential, Erlang, uniform rv's and the Poisson process.

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 7 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. 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.

NAME: ______________________ SECTION: ____ CLASS: 368 | 621

Problem 1

According to the Pew Research Center's demographic survey of Americans, "religious" people have more children than "non-religious" people. As an example, Mormons have on average 3.4 children and others have on average 2.1 children. We will model both groups' number of children as Poisson rv'ss where N_M denotes the model for Mormons and N_O denotes the model for Atheists:

$$\begin{aligned} N_M &\sim \text{Poisson}\,(3.4) \\ N_O &\sim \text{Poisson}\,(2.1) \end{aligned}$$

(a) [difficult] [MA] Comment on the appropriateness of the Poisson model here.

(b) [harder] If we are to only consider Mormons vs everyone else, there about 7M Mormons in the American population which is about 330 million. Create a r.v. X which models sampling one American at random and the realization 1 denotes Mormon and the realization 0 denotes non-Mormon. Find its PMF.

(c) [harder] If you call Y the number of children an American has, draw the two-stage tree (like in class) and then find the distribution of Y where atheist/Mormon status is unknown.

(d) [easy] Is Y a "mixture distribution" or "compound distribution"?

(e) [difficult] If you know someone has 5 children, what is the probability they are Mormon according to our model?

Problem 2

We will now practice multilevel models, mixture distributions and compound distributions.

(a) [harder] Show that if $Y \mid X = x \sim \text{Poisson}(x)$ and $X \sim \text{Gamma}(\alpha, \beta)$ then $Y \sim \text{ExtNegBinomial}\left(\alpha, \frac{\beta}{1+\beta}\right)$. Copy from what we did in class first. This fill in the left-out step. Hint: find the kernel of the $\text{ExtNegBinomial}\left(\alpha, \frac{\beta}{1+\beta}\right)$ before you begin.

(b) [harder] Show that if $Y \mid X = x \sim \text{Exp}(x)$ and $X \sim \text{Gamma}(\alpha, \beta)$ then $Y \sim \text{Lomax}(\alpha, \beta)$. You will need to look up the Lomax distribution on wikipedia.

(c) [harder] Draw a tree of the folowing multilevel hierarchical model.

$$\begin{aligned} X_1 &\sim \text{Gamma}(\alpha_1, \beta_1) \quad \text{independent of} \\ X_2 &\sim \text{Gamma}(\alpha_2, \beta_2) \\ Y \mid X_1 = x_1,\ X_2 = x_2 &\sim \text{Beta}(x_1,\, x_2) \end{aligned}$$

(d) [difficult] [MA] Get as far as you can when finding the PDF of the compound distribution Y.

Problem 3

Moment generating functions (mgf's) and characteristic functions (ch.f.'s)!

(a) [easy] Find a piecewise function that can compute i^n where $i := \sqrt{-1}$ and $n \in \mathbb{N}$. Hint: use the "mod" function (modulus division) to express the cases (see section B's lecture).

(b) [easy] Prove that $\left|e^{i\theta}\right| = 1$ for all θ.

(c) [easy] Give one example function f where you show conclusively that $f \notin L^1$.

(d) [easy] Prove that all PDF's are $\in L^1$.

(e) [harder] Given the Fourier inversion theorem, prove that if $\phi_X(t) \in L^1$ then

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{itx}\phi_X(t)dt.$$

Hint: use the work found in my lecture from 2017.

(f) [easy] Find the ch.f. of $X \sim \text{Bernoulli}\,(p)$.

(g) [harder] Find the ch.f. of $T \sim \text{Binomial}(n, p)$. Hint: use the binomial theorem.

(h) [easy] Using ch.f.'s, find $\mathbb{E}[T]$.

(i) [harder] Using ch.f.'s, find $\mathbb{V}\text{ar}[T]$.

(j) [harder] Using ch.f.'s, show that if $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, then $T = X_1 + \ldots + X_n \sim \text{Binomial}(n, p)$.

(k) [easy] Define the mgf and prove properties 0, 2, 3 and 4 for mgf's. (You cannot prove property 1 without advanced math).

Problem 4

The central limit theorem (CLT) and a corollary.

(a) [easy] State the setup / assumptions of the CLT.

(b) [easy] Prove the CLT (copy from the notes if you get stuck).

(c) [difficult] Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ some distribution with mean μ_X and variance $\sigma_X^2 < \infty$ and let $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim}$ some distribution with mean μ_Y and variance $\sigma_Y^2 < \infty$ which are independent of the X_i's. Prove the following central limit theorem corollary:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}}} \xrightarrow{d} \mathcal{N}(0,\, 1)\,.$$

We need this fact to do two-sample testing in statistics. This looks harder than it is. Trace through the proof in (b), use algebra to simplify expressions and make substitutions in the appropriate places.

Problem 5

Introducing the king: the normal distribution $\mathcal{N}$ and his princes/sses: the lognormal distribution Log$\mathcal{N}$, chi-squared distribution χ^2_k, Student's T distribution T_k and Fisher-Snecodor's distribution F_{k_1,k_2}.

(a) [easy] Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ independent of $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Prove $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ using ch.f.'s.

(b) [E.C.] Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ independent of $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Prove $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ using the definition of convolution on a separate page. This is a lot of boring algebra but it will hone your skills. You can find it in the book or on the Internet (but try not to look at the answer).

(c) [easy] Let $X \sim \text{Log}\mathcal{N}(\mu, \sigma^2)$ and $Y = \ln(X)$. How is Y distributed? Use a heuristic argument. No need to actually change variables.

(d) [harder] Let $X_1 \sim \text{Log}\mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \text{Log}\mathcal{N}(\mu_2, \sigma_2^2), \ldots, X_n \sim \text{Log}\mathcal{N}(\mu_n, \sigma_n^2)$ all independent of each other and $Y = \prod_{i=1}^n X_i$. How is Y distributed? Use a heuristic argument. No need to actually change variables.

(e) [easy] Let $X \sim \chi^2_k$, find $\mathbb{E}[X]$ using the fact that $X = Z_1^2 + Z_2^2 + \ldots + Z_k^2$ where $Z_1, Z_2, \ldots, Z_k \overset{iid}{\sim} \mathcal{N}(0, 1)$.

(f) [easy] Let $X \sim \chi^2_k = \text{Gamma}\left(\frac{k}{2}, \frac{1}{2}\right)$. Find the PDF of X by making the correct substitutions in the gamma PDF and simplifying.

(g) [easy] Using $Z_1, Z_2, \ldots \overset{iid}{\sim} \mathcal{N}(0, 1)$, the function g s.t. $g(Z_1, Z_2, \ldots) \sim \chi^2_k$ where $k \in \mathbb{N}$ is a constant is given below:

$$g(Z_1, Z_2, \ldots) = Z_1^2 + Z_2^2 + \ldots + Z_k^2 \sim \chi^2_k$$

Following this example, find a function g s.t. $g(Z_1, Z_2, \ldots) \sim F_{k_1,k_2}$ where $k_1, k_2 \in \mathbb{N}$ are constants.

(h) [easy] Let $X \sim F_{k_1,k_2}$, find the kernel of $f_X(x)$.

(i) [easy] Using $Z_1, Z_2, \ldots \overset{iid}{\sim} \mathcal{N}(0, 1)$, find a function g s.t. $g(Z_1, Z_2, \ldots) \sim T_k$ where $k \in \mathbb{N}$ is a constant.

(j) [easy] Let $X \sim T_k$, find the kernel of $f_X(x)$.

(k) [E.C.] Derive the PDF of the T_k distribution using the ratio formula where you first find the distribution of the denominator explicitly. Do on a separate piece of paper.

(l) [E.C.] Show that the PDF of $X \sim T_k$, converges to the PDF of $Z \sim \mathcal{N}(0, 1)$ when $k \to \infty$. Hint: use Stirling's approximation. Do on a separate piece of paper.

(m) [easy] Let $X \sim \text{Cauchy}(0, 1)$, find the kernel of $f_X(x)$.

(n) [easy] Using $Z_1, Z_2, \ldots \overset{iid}{\sim} \mathcal{N}(0, 1)$, find a function g s.t. $g(Z_1, Z_2, \ldots) \sim \text{Cauchy}(0, 1)$.

(o) [easy] Let $X \sim \text{Cauchy}(0, 1)$, prove that $\mathbb{E}[X]$ does not exist without using its ch.f.