MATH 621 Fall 2017 Homework #4 INCOMPLETE

Professor Adam Kapelner

Due *in review session* or KY604 9:30PM, Monday, November 13, 2017

(this document last updated Tuesday 7$^{\text{th}}$ November, 2017 at 10:00pm)

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 about univariate and multivariate transformations of r.v.'s (discrete and continuous), kernels of PMFs / PDFs, order statistics, the gamma and beta functions, mixture / compound distributions, the normal distribution and the χ^2 distribution.

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 10 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: ____

Problem 1

We will practice finding kernels and relating them to known distributions. The gamma function and the beta function will come up as well.

(a) [easy] Find the kernel of the negative binomial PMF.

(b) [easy] Find the kernel of the beta PDF.

(c) [easy] Find the kernel of the beta binomial PMF.

(d) [easy] If $k(x) = e^{\lambda x} x^{k-1}$ how would you know if the r.v. X was an Erlang (k, λ) or a Gamma (k, λ)?

(e) [harder] If $k(x) = x e^{-x^2}$, how is X distributed?

(f) [difficult] If $k(x) = x^{-d}$ where $d > 1$, how is X distributed?

(g) [difficult] Given only $k(x)$, would you be able find $\text{Supp}[X]$? Yes/no and explain.

(h) [difficult] Prove $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Using the method from class (i.e. the textbook) is not required.

Problem 2

We will now practice using order statistics concepts.

(a) [easy] If $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ where its CDF is denoted $F(x)$, express the CDF of the maximum X_i and express the CDF of the minimum X_i.

(b) [easy] If $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ where its CDF is denoted $F(x)$, express the PDF of the maximum X_i and express the PDF of the minimum X_i.

(c) [easy] If $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ where its CDF is denoted $F(x)$, express the PDF and the CDF of $X_{(k)}$ i.e. the kth smallest X_i.

(d) [difficult] If $X_1, \ldots, X_n \overset{iid}{\sim} p(x)$, why would the formulas in (a-c) not be accurate?

(e) [harder] If $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Exp}(\lambda)$, find the PDF and CDF of the maximum.

(f) [harder] If $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Exp}(\lambda)$, find the PDF and CDF of the minimum.

(g) [harder] If $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{U}(0,\, 1)$, show that $X_{(k)} \sim \mathrm{Beta}(k,\, n-k+1)$.

(h) [harder] Express $\binom{n}{k}$ in terms of the beta function.

(i) [E.C.] If $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{U}\,(a,\, b)$, show that $X_{(k)}$ is a linear transformation of the beta distribution and find its parameters.

(j) [harder] Show that $I_x(\alpha, \beta + 1) = I_x(\alpha, \beta) + \frac{x^\alpha (1-x)^\beta}{\beta B(\alpha,\beta)}$.

(k) [E.C.] If $X \sim \text{Binomial}\,(n,\, p)$, show that $F(x) = I_{1-p}(n-k, k+1)$

Problem 3

We will practice truncations of r.v.'s.

(a) [easy] Given r.v. X, restate the formulas for the PDF of X for (i) the arbitrary truncation to the set $X \in A$, (ii) the truncation for $X \geq x_0$ and (iii) the truncation for $X \leq x_0$.

(b) [harder] If $T \sim \text{Weibull}(k, \lambda)$ and it is known that $T \leq 120$ years, find the PDF of the truncated T.

(c) [harder] Using the notation from 2(i), find the PMF of $X \sim \text{Binomial}(n, p)$ where it is known that $X > n_0$.

Problem 4

We will now practice multivariate change of variables where $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$.

(a) [easy] State the formula for the PDF of $\boldsymbol{Y}$.

(b) [easy] Demonstrate that the formula for the PDF of $\boldsymbol{Y}$ reduces to the univariate change of variables formula if the dimensions of $\boldsymbol{Y}$ and $\boldsymbol{X}$ are 1.

(c) [easy] State the formula for the PDF of $Y = \frac{X_1}{X_2}$.

(d) [easy] State the formula for the PDF of $Y = \frac{X_1}{X_2}$ if X_1 and X_2 are independent.

(e) [easy] State the formula for the PDF of $Y = \frac{X_1}{X_2}$ if X_1 and X_2 are independent and have positive supports.

(f) [easy] State the formula for the PDF of $Y = \frac{X_1}{X_1+X_2}$.

(g) [easy] State the formula for the PDF of $Y = \frac{X_1}{X_1+X_2}$ if X_1 and X_2 are independent.

(h) [easy] State the formula for the PDF of $Y = \frac{X_1}{X_1+X_2}$ if X_1 and X_2 are independent and have positive supports.

(i) [difficult] Find a formula for the PDF of $Y = X_1^{X_2}$.

(j) [difficult] Find the PDF of $Y = \frac{X_1}{X_2}$ if $X_1 \sim \text{Gamma}\,(\alpha, 1)$ independent of $X_2 \sim \text{Gamma}\,(\beta, 1)$. This is known as the beta prime distribution $\beta'(\alpha, \beta)$.

Problem 5

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

(a) [easy] 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 Atheists have on average 1.6 children. Model both groups' number of children as Poissons.

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

(c) [easy] If we are to only consider atheists and Mormons, there are about 10M atheists in the American population and about 7M Mormons in the American population. Create a r.v. X which is 1 if Mormon and 0 if atheist.

(d) [harder] If you call Y the number of children someone has, find the distribution of Y where atheist/Mormon status is unknown. Draw a tree of this model.

(e) [easy] Is this a mixture or compound distribution?

(f) [difficult] If somone has 5 kids, what is the probability they are Mormon according to our model?

(g) [difficult] Find the mixture distribution for Y if $Y \mid X = x \sim \text{Binomial}\,(x, p)$ where $X \sim \text{Poisson}\,(\lambda)$. Draw a tree of this model. Get as far as you can.

(h) [difficult] Find the compound distribution for Y if $Y \mid X_1 = x_1,\ X_2 = x_2 \sim \text{Beta}\,(x_1,\ x_2)$ where $X_1 \sim \text{Gamma}\,(\alpha_1, \beta_1)$ independent of $X_2 \sim \text{Gamma}\,(\alpha_2, \beta_2)$. Draw a tree of this model. Get as far as you can.

(i) [easy] Can this be considered an "overdispersed" beta? Yes/no.

(j) [easy] Why does mixing / compounding give more "degrees of freedom" to the model of the phenomenom you care about (denoted Y in class and above). Discuss what this means and how it may be useful in the real world.

Problem 6

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 in the book but try not to look at it.

(c) [harder] Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = X\mathbb{1}_{X \geq a}$. Find $f_Y(y)$.

(d) [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.

(e) [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.

(f) [harder] The average return of the S&P 500 stock index since 1928 is 11.4% and the standard deviation is 19.7%. Assume for the purposes of this problem that percentage returns is normally distributed (even though it is not true in practice). If you put $1,000 into the stock market, what is the probability you have $5,000 after 10 years? The `R` function you need is `plnorm`.

(g) [easy] Using $Z_1, Z_2, \ldots \overset{iid}{\sim} \mathcal{N}(0,\, 1)$, find a function g s.t. $g(Z_1, Z_2, \ldots, k_1, \ldots) \sim \chi^2_k$ where $k_1, \ldots$ represents constants.

(h) [easy] Let $X \sim \chi^2_k$, 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. $g(Z_1, Z_2, \ldots, k_1, \ldots) \sim F_{k_1,k_2}$ where $k_1, \ldots$ represents constants.

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

(k) [easy] Using $Z_1, Z_2, \ldots \overset{iid}{\sim} \mathcal{N}(0,\, 1)$, find a function g s.t. $g(Z_1, Z_2, \ldots, k_1, \ldots) \sim T_k$ where $k_1, \ldots$ represents constants.

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

(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. $g(Z_1, Z_2, \ldots, k_1, \ldots) \sim \text{Cauchy}\,(0,\, 1)$ where $k_1, \ldots$ represents constants.

(o) [easy] Let $X \sim \text{Cauchy}\,(0,\, 1)$, prove that $\mathbb{E}\left[X\right]$ does not exist.

(p) [E.C.] Show that the PDF of $X \sim T_k$, converges to the PDF of $Z \sim \mathcal{N}(0,\, 1)$ when $k \to \infty$. Do not look at the notes.