MATH 368/621 Fall 2021 Homework #4

Professor Adam Kapelner

Due by email 11:59PM November 14, 2021

(this document last updated Wednesday 3rd November, 2021 at 7:32pm)

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 univariate and multivariate transformation of continuous rv's, kernels, order statistics, the beta distribution, mixture / compound distributions.

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: ____

Problem 1

We will now explore a couple of extreme distributions.

(a) [harder] Let $X \sim \text{Exp}(1)$ and $Y = -\ln(X) \sim \text{Gumbel}(0, 1)$. Find the PDF of this standard Gumbel distribution. Make sure you include the indicator function throughout your proof.

(b) [easy] Find the CDF of Y.

(c) [easy] Let $G = \beta Y + \mu \sim \text{Gumbel}(\mu, \beta)$. Find the PDF of G, the general Gumbel distribution.

(d) [harder] Show that for any r.v. X, if $Y = aX + b$, then $F_Y(y) = F_X\left(\frac{y-b}{a}\right)$.

(e) [easy] Using the answer in the previous question, find the CDF of $G \sim \text{Gumbel}(\mu, \beta)$.

Problem 2

These exercises will give you practice with the Weibull distribution.

(a) [easy] If $X \sim \text{Exp}\left(1\right)$ then show that $Y = \frac{1}{\lambda} X^{\frac{1}{k}} \sim \text{Weibull}\left(k, \lambda\right)$ where $k, \lambda > 0$.

(b) [harder] Find $\text{Med}\left[Y\right]$.

(c) [difficult] [MA] Prove that if $k > 1$ then $\mathbb{P}(Y \geq y + c \mid Y \geq c) < \mathbb{P}(Y \geq y)$ for $c > 0$.

(d) [difficult] If $X \sim \text{Exp}(\lambda)$ then show that $Y = X^{\beta} \sim$ Weibull where $\beta > 0$. Find the resulting Weibull's parameters in terms of the parameterization we learned in class (i.e. your answer in part a).

(e) [easy] Using Y, the Weibull in terms of the parameterization we learned in class (i.e. your answer in part a), find the PDF of $W = Y + c \sim \text{Weibull}(k, \lambda, c)$ which is known as the "translated Weibull" or "3-parameter Weibull model".

Problem 3

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] If $k(x) = e^{-\lambda x} x^{k-1} \mathbb{1}_{x>0}$ how would you know if the r.v. X was an Erlang (k, λ) or a Gamma (k, λ)?

(d) [harder] If $k(x) = x e^{-x^2} \mathbb{1}_{x>0}$, how is X distributed?

Problem 4

We will now practice using order statistics concepts.

(a) [easy] If $X_1, \ldots, X_n \stackrel{iid}{\sim} f(x)$ 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] [MA] If discrete $X_1, \ldots, X_n \overset{iid}{\sim}$ some discrete distribution, why would the formulas in (a-c) not be accurate?

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

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

(g) [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.

(h) [difficult] [MA] If $X \sim \text{Binomial}\,(n,\, p)$, show that $F(x) = I_{1-p}(n-k, k+1)$. You will need to assume the property $I_x(a,b) = 1 - I_{1-x}(b,a)$.

This answer is done for you. You should all read it and understand it.

Assume $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{U}\,(0,\, 1)$. We know then that $F_X(x) = x$. So by the CDF formula for an order statistic k and the fact that the CDF is a probability we have:

$$\begin{aligned} F_{X_{(k)}}(x) &= \sum_{j=k}^{n} \binom{n}{j} F_X(x)^j \left(1 - F_X(x)\right)^{n-j} \\ &= \sum_{j=k}^{n} \binom{n}{j} x^j \left(1 - x\right)^{n-j} \end{aligned}$$

$$= \quad 1 - \sum_{j=0}^{k-1} \binom{n}{j} x^j \left(1-x\right)^{n-j}$$

From class we proved that the order statistics for the standard uniform are distributed beta and we also know its CDF:

$$X_{(k)} \sim \text{Beta}\left(k,\, n-k+1\right) = \text{Beta}\left(k,\, n-(k-1)\right) \quad \Rightarrow \quad F_{X_{(k)}}(x) = I_x(k, n-(k-1))$$

By the fact on wikipedia about the I function we have:

$$F_{X_{(k)}}(x) = I_x(k, n-(k-1)) = 1 - I_{1-x}(n-(k-1), k)$$

Thus we have the strange equality:

$$\sum_{j=0}^{k-1} \binom{n}{j} x^j \left(1-x\right)^{n-j} = I_{1-x}(n-(k-1), k)$$

Letting $y := k-1$,

$$\sum_{j=0}^{y} \binom{n}{j} x^j \left(1-x\right)^{n-j} = I_{1-x}(n-y, y+1)$$

Note that the lhs is the CDF for $Y \sim \text{Binomial}\left(n,\, x\right)$. $\blacksquare$

Problem 5

We will now practice multivariate change of variables where $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$ where $\boldsymbol{X}$ denotes a vector of k continuous r.v.'s and $\boldsymbol{g} : \mathbb{R}^k \to \mathbb{R}^k$ and is 1:1.

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

(b) [harder] 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 $R = \frac{X_1}{X_2}$.

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

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

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

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

(h) [harder] State the formula for the PDF of $R = \frac{X_1}{X_1+X_2}$ if X_1 and X_2 are independent and have positive supports. This should be a simpler expression than the previous.

(i) [difficult] Find a formula for the PDF of $E = X_1^{X_2}$ where $X_1, X_2 \overset{iid}{\sim} f(x)$.

(j) [difficult] Find the simplest formula you can for the PDF of $Q = \frac{X_1}{X_2} e^{X_3}$ where X_1, X_2, X_3 are dependent continuous r.v.'s.

(k) [easy] Show that $R = \frac{X_1}{X_2} \sim \beta'(\alpha, \beta)$, the beta prime distribution, if $X_1 \sim \text{Gamma}(\alpha, 1)$ independent of $X_2 \sim \text{Gamma}(\beta, 1)$. Marked easy since we did it in class. Be careful to include the indicator function for r in the final result.

Problem 6

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's 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 7

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

(a) [difficult] Show that if $Y \mid X = x \sim \text{Poisson}(x)$ and $X \sim \text{Gamma}(\alpha, \beta)$ then $Y \sim \text{ExtNegBinomial}\left(\alpha, \frac{\beta}{1+\beta}\right)$. Hint: find the kernel of the $\text{ExtNegBinomial}\left(\alpha, \frac{\beta}{1+\beta}\right)$ before you begin. And during the derivation of $f_Y(y)$, keep simplifying the kernel by removing multiplicative constants.

(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.