MATH 340/640 Fall 2025 Homework #7

Professor Adam Kapelner

Due by email on the date found on the homepage

(this document last updated Saturday 15th November, 2025 at 11:52pm)

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* — read about the concepts we discussed in class online. For this homework set, review the previous random variables and read about truncations, the Weibull, the ParetoI, the law of iterated expectation, the law of total variance, joint ch.f's, the multivariate normal distribution 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. "[MA]" are for those registered for 640 and extra credit otherwise.

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 5 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 acquaint ourselves with truncations.

(a) [harder] Let $Z \sim \mathcal{N}(0, 1)$ and $Y = Z \,\big|\, |Z| > 2$. Find $f_Y(y)$. Note: $\mathbb{P}(|Z| > 2) \approx 5\%$.

(b) [in the notes] If $X \sim \text{ParetoI}(k, \lambda)$ and $c > 0$. Prove $Y = X \,|\, X > k + c$ is also ParetoI-distributed and find its density.

Problem 2

We will derive some properties of the ParetoI distribution.

(a) [in the notes] Using the formula we derived in class for λ as a ratio of logs, find the ParetoI distribution corresponding to the 80-20 rule. The value of k doesn't matter; let $k = 1$.

(b) [easy] Let $X \sim \text{ParetoI}(\lambda, 1)$ where the value of λ was found in the previous question. Find $Q[X, 0.8]$, the 80th %ile of this distribution.

(c) [easy] Using the formula we derived in class for λ as a ratio of logs, find the ParetoI distribution corresponding to the 99-1 rule. The value of k doesn't matter so let $k = 1$.

(d) [easy] If human wealth follows the 99-1 rule, what do the top 1% own?

(e) [harder] If human wealth follows the 99-1 rule, what do the top top 1% of the top 1% own?

(f) [easy] Plot the PDF from (b) in the range $x \in [0, 20]$ and shade in the area where $x < a$ in blue and $x \geq a$ in red. Label the axes appropriately. This problem is done for you (see Figure 1 next page).

(g) [easy] Plot the PDF from (b) times x (this is the total amount of land owned by people up to land amount x) in the range $x \in [0, 20]$ and shade in the area where $x < a$ in blue and $x \geq a$ in red. Label the axes appropriately. This problem is done for you (see Figure 2 next page).

(h) [easy] Calculate the blue region's area in both plots and compare the blue region in these two plots. What is this relationship called? This problem is done for you.

Although the plot of the density has 80% of the total area in the region where $x < a$, the plot of the total amount of land owned has only 20% of the total area in the region where $x < a$. This is the 80-20 rule.

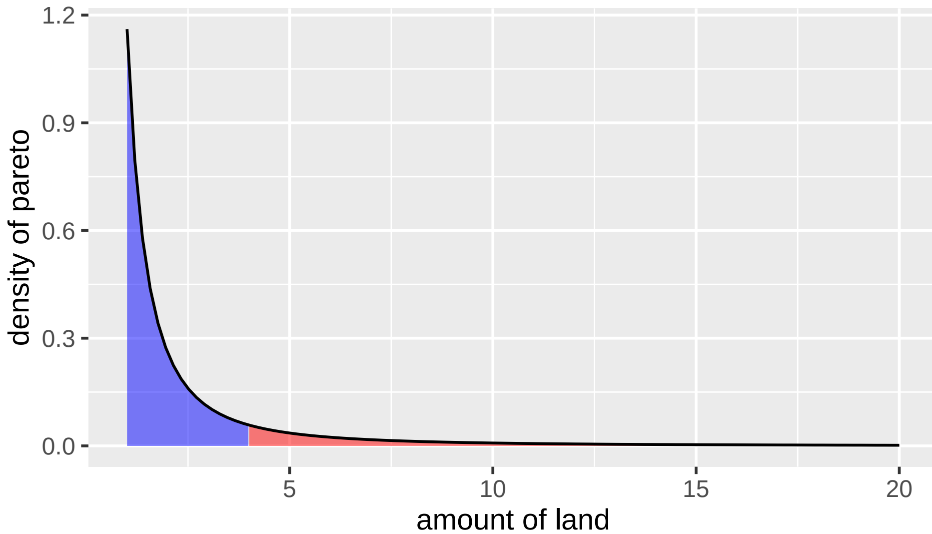

Figure 1: PDF of the Pareto from (a)

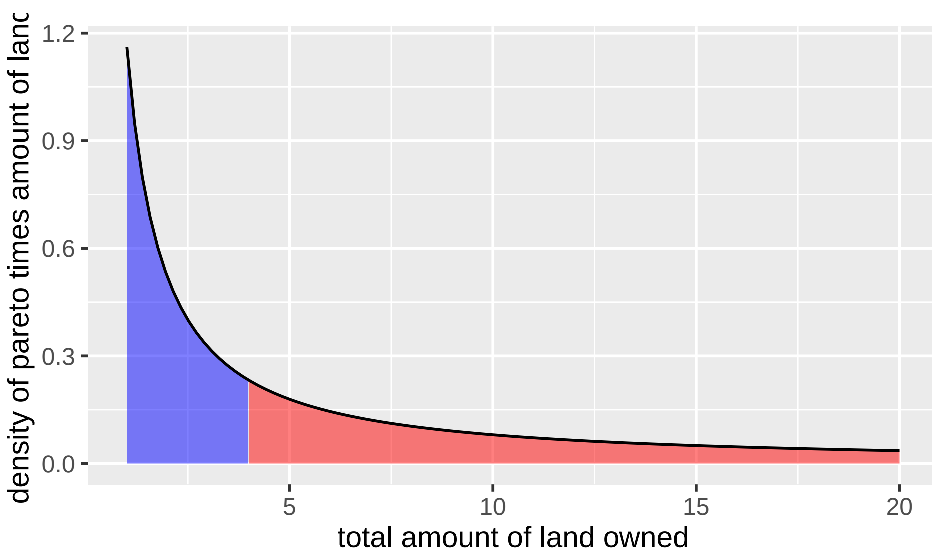

Figure 2: PDF of the integral of the Pareto PDF from (a) times x

Problem 3

We will acquaint ourselves with the Weibull modulus.

(a) [in the notes] Draw the densities of the three types of Weibull distributions and indicate the set of the Weibull modulus k below each graph.

(b) [in the notes] Let $X \sim \text{Weibull}(k, \lambda)$. If $k = 1$, prove that this Weibull is an exponential distribution with rate paramter λ and then prove that it exhibits the memorlessness property.

(c) [difficult] Let $X \sim \text{Weibull}(k, \lambda)$. If $k > 1$, prove that $\mathbb{P}(X > x + c \mid X > c) < \mathbb{P}(X > x)$ for all $x, c > 0$. This problem is done for you below.

We proved in class that $\mathbb{P}(X > x + c \mid X > c) = e^{\lambda^k (c^k - (x+c)^k)}$ and $\mathbb{P}(X > x) = e^{-(\lambda x)^k}$. This means what we want to show is equivalent to showing:

$$\frac{e^{\lambda^k (c^k - (x+c)^k)}}{e^{-(\lambda x)^k}} = e^{\lambda^k (c^k + x^k - (x+c)^k)} < 1$$

Taking the log of both sides we wish to show that $\lambda^k (c^k + x^k - (x+c)^k) < 0$. We know that $\lambda^k > 0$ since $\lambda, k > 0$ due to the parameter space of X. So now we need to show that

$$c^k + x^k < (x+c)^k$$

We now multiply both sides by x^{-k} and define $d := c/x$. We know $d > 0$ since both $x, c > 0$. This gives us:

$$1 + d^k < (1 + d)^k$$

Now we let $k = 1 + \beta$. Since $k > 1$, we know β is positive. Using this notation, we want to show

$$1 + dd^\beta < (1 + d)(1 + d)^\beta$$

Expanding gives us

$$1 + dd^\beta < (1 + d)^\beta + d(1 + d)^\beta \tag{1}$$

The inequality $a_1 + b_1 < a_2 + b_2$ if $a_1 < b_1$ and $a_2 < b_2$. This is true here as Since

$$1 < (1 + d)^\beta \quad \Rightarrow \quad 0 < \beta \ln (1 + d)$$

and $\beta > 0$ and $\ln (1 + d) > 0$ as $d > 0$. And also

$$dd^\beta < d(1 + d)^\beta \quad \Rightarrow \quad d^\beta < (1 + d)^\beta \quad \Rightarrow \quad \beta \ln (d) < \beta \ln (1 + d) \quad \Rightarrow \quad \ln (d) < \ln (1 + d) \, .$$

(d) [in the notes] Let $X \sim \text{Weibull}(k, \lambda)$. If $k \in (0, 1)$, prove that $\mathbb{P}(X > x + c \mid X > c) > \mathbb{P}(X > x)$ for all $x, c > 0$. This proof follows from the previous question. Begin at Inequality 1 and flip the inequality.

(e) [harder] Let $X \sim \text{Weibull}(k, \lambda)$ where $k > 1$. Provide one real life example that could be modeled by this rv that was not the one we said in class.

(f) [harder] Let $X \sim \text{Weibull}(k, \lambda)$ where $k \in (0, 1)$. Provide one real life example that could be modeled by this rv that was not the one we said in class.

Problem 4

We will acquaint ourselves with the law of iterated expectation and the law of total variance.

(a) [easy] State the law of iterated expectation.

(b) [harder] If $X \sim \text{Gamma}(\alpha, \beta)$ and $Y \mid X = x \sim \text{Exp}(x)$, find $\mathbb{E}[Y]$ using the law of iterated expecation.

(c) [harder] This is a canonical problem in econometrics. If $X \sim \mathcal{N}(\mu, \tau^2)$ and $Y \mid X = x \sim \mathcal{N}(a + bx, \sigma^2)$, where $a, b \in \mathbb{R}$, find $\mathbb{E}[Y]$ using the law of iterated expecation.

(d) [easy] State the law of total variance (which is also called the "variance decomposition formula").

(e) [harder] [MA] If $X \sim \text{Gamma}(\alpha, \beta)$ and $Y \mid X = x \sim \text{Exp}(x)$, find $\mathbb{V}\text{ar}[Y]$ using the law of total variance.

(f) [harder] This is a canonical problem in econometrics. If $X \sim \mathcal{N}(\mu, \tau^2)$ and $Y \mid X = x \sim \mathcal{N}(a + bx, \sigma^2)$, where $a, b \in \mathbb{R}$, find $\mathbb{V}\text{ar}[Y]$ using the law of total variance.

Problem 5

We will learn to sample a multinomial iteratively. Imagine a bag of fruits with 4 apples, 7 bananas, 6 cantaloupes.

(a) [easy] Review for the final: imagine sampling from the bag 37 times with replacement and tallying the number of apples as x_1, the number of bananas as x_2 and the number of cantaloupes as x_3. How is the rv $\boldsymbol{X} := [X_1 \; X_2 \; X_3]^\top$ distributed?

(b) [easy] What is the distribution of X_1?

(c) [easy] What is the distribution of $X_2 \mid X_1 = x_1$?

(d) [harder] What is the distribution of $X_3 \mid X_2 = x_2, X_1 = x_1$?

(e) [easy] Using your answers to parts (b-d), write an iterative algorithm that can draw samples from $\boldsymbol{X}$.

Problem 6

These questions are about joint characteristic functions and the multinomial distribution.

(a) [easy] State the definition of the joint ch.f.

(b) [easy] State properties 0-9 for joint ch.f's.

(c) [in the notes] Let $\boldsymbol{X} \sim \text{Mult}_K(n, \boldsymbol{p})$. Find $\phi_{\boldsymbol{X}}(\boldsymbol{t})$.

(d) [easy] Find the distribution of X_{17} using $\phi_{\boldsymbol{X}}(\boldsymbol{t})$.

(e) [harder] [MA] Find $\mathbb{C}\text{ov}\left[X_{17}X_{37}, X_{53}\right]$ using $\phi_{\boldsymbol{X}}(\boldsymbol{t})$.

(f) [difficult] Let $\boldsymbol{X}$ be the multivariate Cauchy distribution with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$. It's characteristic function is $\phi_{\boldsymbol{X}} = e^{i\boldsymbol{t}^\top\boldsymbol{\mu} - \sqrt{\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}}}$ (see 3.13 of this paper if you're interested). Find the distribution of X_{17} using $\phi_{\boldsymbol{X}}$.

Problem 7

These questions are about the multivariate normal.

(a) [easy] Let $\boldsymbol{X} \sim \mathcal{N}_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$. Write the jdf of $\boldsymbol{X}$ below. What are the restrictions on $\boldsymbol{\Sigma}$?

(b) [harder] Let $Z_1, \dots, Z_n \stackrel{iid}{\sim} \mathcal{N}(0,\, 1)$ and let $\boldsymbol{Z} = [Z_1 \,\dots\, Z_n]^\top$. Show that $\boldsymbol{Z}$ is multivariate normal with $\boldsymbol{\mu} = \mathbf{0}_n$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_n$ using the jdf of the $\mathcal{N}_n(\boldsymbol{\mu},\, \boldsymbol{\Sigma})$ rv.

(c) [easy] Find $\phi_{\boldsymbol{Z}}(\boldsymbol{t})$. Hint: use the iid property and the known chf of the standard normal.

(d) [easy] Let $A \in \mathbb{R}^{m \times n}$ and $\mu \in \mathbb{R}^n$. Let $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}$. Let $\boldsymbol{\Sigma} := AA^\top$, Find $\phi_{\boldsymbol{X}}(\boldsymbol{t})$.

(e) [easy] Let $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu},\, \boldsymbol{\Sigma})$. Find the distribution of X_{17} using $\phi_{\boldsymbol{X}}(\boldsymbol{t})$.

(f) [difficult] Let $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Find the distribution of $[X_{17}\, X_{37}]^\top$ using $\phi_{\boldsymbol{X}}(\boldsymbol{t})$.

(g) [harder] Assuming for any vector rv $\boldsymbol{Y}$ with dimension n that $\mathbb{V}\text{ar}[\boldsymbol{\mu} + \boldsymbol{Y}] = \mathbb{V}\text{ar}[\boldsymbol{T}]$ and $\mathbb{V}\text{ar}[A\boldsymbol{Y}] = A\mathbb{V}\text{ar}[\boldsymbol{Y}]A^\top$. Let $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Show that $\mathbb{V}\text{ar}[\boldsymbol{X}] = \boldsymbol{\Sigma}$.

(h) [in the notes] Define the Mahalanobis distance. How is it distributed?

(i) [difficult] Let $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $B \in \mathbb{R}^{m \times n}$ and $\boldsymbol{c} \in \mathbb{R}^m$. Find the distribution of $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{c}$.

(j) [difficult] [MA] Are there any restrictions on $\boldsymbol{\mu}, \boldsymbol{c}, B, \boldsymbol{\Sigma}$ in the previous question to have the result you found?

Problem 8

We will acquaint ourselves with finding modes of distributions.

(a) [harder] Let $X \sim \text{Gamma}(\alpha, \beta)$. Find Mode$[X]$. Any restrictions on α, β besides the parameter space?

(b) [harder] Let $X \sim T_k$. Let $Y = \mu + \sigma X \sim T_k(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma > 0$ which is called the "location-scale T-distribution". Find Mode$[Y]$. We can use the result from a previous year's midterm II:

$$f_Y(y) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k \sigma^2}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{(y-\mu)^2}{k\sigma^2}\right)^{-\frac{k+1}{2}}$$

Problem 9

These exercises will give you practice with the Poisson process and the analogous Binomial-Negative Binomial relationship.

(a) [easy] Write the assumptions and the main result of the Poisson process (an equivalence of two probability statements and then an equivalence using the CDF's of the Erlang and the Poisson models).

(b) [easy] Assume $X_1, X_2, X_3, \ldots \overset{iid}{\sim} \text{Exp}(\lambda)$. Calculate $\mathbb{P}(X_1 + X_2 + X_3 + X_4 + X_5 < 1)$ using the two different ways (i.e. via the Poisson Process relationship).

(c) [harder] Let $N \sim \text{Poisson}(\lambda)$. Describe a way to use the realizations from the r.v.'s $X_1, X_2, X_3, \ldots \overset{iid}{\sim} \text{Exp}(\lambda)$ to create a realization n from the Poisson model.

(d) [difficult] Assume $X_1, X_2, X_3, \ldots \overset{iid}{\sim} \text{Exp}(\lambda)$. Calculate $\mathbb{P}(X_1 + X_2 + X_3 + X_4 + X_5 < m)$ where $m \in \mathbb{N}$ using two different ways (i.e. via the Poisson Process relationship).

(e) [E.C.] Prove the analgous Binomial-Negative Binomial relationship (an equivalence of two probability statements and then an equivalence using the CDF's of the Binomial and the Negative Binomial models).