Student ID (NOT your name):

Final Examination: QUESTION BOOKLET

Prof. Will Fithian

Fall 2023

- Do *NOT* open this question booklet until you are told to do so.
- Write your Student ID number (**NOT** your name) at the top of this page.
- Write your solutions in this booklet.
- No electronic devices are allowed during the exam.
- Be neat! If we can't read it, we can't grade it.
- You can treat any results from lecture or homework as "known," and use them in your work without rederiving them, but do make clear what result you're using. You do not need to explicitly check regularity conditions for the theorems from class that required them.
- For a multi-part problem, you may treat the results of previous parts as given (if you don't prove the result for part (a), you can still use it to solve part (b)).
- I have starred some parts which I believe are the most difficult, and which I expect most students won't necessarily be able to solve in the time allotted. They are generally not worth more points than the less difficult parts, so don't waste too much time on them until you're happy with your answers to the latter.
- Be careful to justify your reasoning and answers. We are primarily interested in your understanding of concepts, so show us what you know.
- Good luck!

1. Laplace Location Family (24 points, 4 points / part).

Some useful facts for this problem:

- The exponential distribution with scale parameter $\theta > 0$ is called $\text{Exp}(\theta)$ and has density
$$p_\theta(x) = \frac{1}{\theta} e^{-x/\theta}, \quad \text{for } x > 0.$$
The mean is θ and the variance is θ^2.

- The Gamma distribution with scale parameter $\theta > 0$ and shape parameter $k > 0$ is called $\text{Gamma}(k, \theta)$ and has density
$$p_{k,\theta}(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad \text{for } x > 0,$$
where $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t}\, dt$. The mean and variance are $k\theta$ and $k\theta^2$.

- A sum of k independent $\text{Exp}(\theta)$ random variables is $\text{Gamma}(k, \theta)$.

- If $Z \sim \text{Gamma}(k, \theta)$ then $aZ \sim \text{Gamma}(k, a\theta)$, for any $a > 0$.

Suppose that we observe an i.i.d. sample from the *Laplace scale family* with parameter $\theta > 0$:

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Laplace}(0, \theta) = \frac{1}{2\theta} e^{-|x|/\theta}, \quad \text{for } x \in \mathbb{R}.$$

Note the density is supported on the entire real line. This is not the same as the Laplace location family that we have used as a running example in class.

(a) Show that $|X_i| \sim \text{Exp}(\theta)$ for $i = 1, \ldots, n$.

Solution:
For $0 \le a \le b < \infty$), we have

$$\begin{aligned}
\mathbb{P}(|X_i| \in [a, b]) &= \mathbb{P}(X_i \in [-b, -a]) + \mathbb{P}(X_i \in [a, b]) \\
&= \frac{1}{2}\left(\int_{-b}^{-a} \frac{1}{\theta} e^{x/\theta}\, dx + \int_a^b \frac{1}{\theta} e^{-x/\theta}\, dx\right) \\
&= \int_a^b \frac{1}{\theta} e^{-x/\theta}\, dx \\
&= \mathbb{P}(Y \in [a, b]),
\end{aligned}$$

where $Y \sim \text{Exp}(\theta)$.

(b) Find a minimal sufficient statistic for this model. Is it complete?

Solution:
The likelihood is

$$p(x) = \left(\frac{1}{2\theta}\right)^n \exp\left\{-\frac{1}{\theta}\sum_i |x_i|\right\},$$

which we can recognize as an exponential family model with sufficient statistic $\sum_i |X_i|$ and natural parameter $1/\theta$. Because $1/\theta$ ranges over the entire interval $(0, \infty)$, the sufficient statistic is complete and therefore minimal.

(c) Find the maximum likelihood estimator for θ and give its asymptotic distribution.

Solution:
Let $T(X) = \sum_i |X_i|$. The log-likelihood is

$$\ell_n(\theta; X) = -n\log(2\theta) - \frac{1}{\theta}T(X),$$

and its derivative is

$$\dot{\ell}_n(\theta; X) = -\frac{n}{\theta} + \frac{1}{\theta^2}T(X) = \frac{n}{\theta^2}(T(X)/n - \theta),$$

so the likelihood is maximized by setting $\dot{\ell}_n(\hat{\theta}; X) = 0$, giving

$$\hat{\theta} = T/n = \frac{1}{n}\sum_i |X_i|.$$

Moreover, we can easily see from the last expression for $\dot{\ell}_n(\theta; X)$ that its sign is the same as the sign of $T/n - \theta$, so T/n is the global maximizer. The Fisher information is therefore

$$J_n(\theta) = \text{Var}_\theta(T(X)/\theta^2) = n\theta^{-4}\text{Var}_\theta(X_i) = n\theta^{-2}.$$

As a result, the asymptotic distribution of the MLE is

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \theta^2).$$

(d) Show the estimator from the previous part is unbiased. Does it achieve the Cramér-Rao Lower Bound?

Solution:
The estimator is unbiased because $\mathbb{E}_\theta|X_i| = \theta$, since it is exponentially distributed. So T/n, which is an average of n random variables each having expectation θ, also has expectation θ and is therefore unbiased. Its variance is θ^2/n, again because it is an average of n independent random variables each with variance θ^2. This matches the Cramér-Rao Lower Bound based on the variance calculated above.

(e) Now, suppose that we are concerned the variance might be gradually shrinking. Specifically, we are concerned that the ith random variable has parameter $\theta_i = \theta_0(1-\delta)^i$. That is, we consider an alternative model with an additional parameter $\delta \in [0, 1)$, where

$$X_i \overset{\text{ind.}}{\sim} \text{Laplace}(0, \theta_0(1-\delta)^i), \qquad \text{for } i = 1, \ldots, n.$$

Assume (for this part **only**) that the value of θ_0 is known.

Suppose that we want to test our original model (which has $\delta = 0$) against the alternative that $\delta > 0$. Suggest a score test, giving an explicit expression for the score statistic and a cutoff based on its asymptotic distribution. You do **not** need to justify why the score statistic (calculated in the usual way and appropriately normalized) is asymptotically Gaussian in this non-i.i.d. model; you can just assume that it is.

Solution:
Now the log-likelihood is

$$\ell_n(\delta; X) = \sum_i -\log(2\theta_0) - i\log(1-\delta) - \frac{|X_i|/\theta_0}{(1-\delta)^i},$$

and its derivative is

$$\dot{\ell}_n(\delta; X) = \sum_i \frac{i}{1-\delta} - \frac{i|X_i|/\theta_0}{(1-\delta)^{i+1}}$$

The score evaluated at $\delta = 0$ is then

$$\dot{\ell}_n(0; X) = \sum_{i=1}^n i(1 - |X_i|/\theta_0),$$

and the Fisher information at $\delta = 0$ is

$$J_n(0) = \text{Var}_0(\dot{\ell}_n(\delta; X)) = \sum_{i=1}^n i^2 \text{Var}(|X_i|/\theta_0) = \sum_{i=1}^n i^2.$$

Thus, the normalized test statistic is

$$Z = J_n^{-1/2} \dot{\ell}_n(0; X) = \frac{\sum_{i=1}^n i(1 - |X_i|/\theta_0)}{\left(\sum_{i=1}^n i^2\right)^{-1/2}} \Rightarrow N(0, 1),$$

and since we are doing a one-sided test we reject if it is larger than z_α.

(f) (*) Now, drop the assumption that θ_0 is known, so that now both θ_0 and δ are unknown. Assume we want to test the same hypothesis, $H_0 : \delta = 0$ against $H_1 : \delta > 0$, with θ_0 as a nuisance parameter. How can we modify the test from the previous part so that it has finite-sample control of the Type I error rate? You do not need to give an explicit cutoff, but you should give a sufficient explanation of how you would find it without knowing the value of θ_0.

Solution:
A natural idea is to condition on the value of $T(X)$, which is a sufficient statistic for the null submodel. Then, we can calculate the same statistic from the previous part, plugging in $\hat{\theta}_0 = T/n$ for θ_0:

$$W = \frac{\sum_{i=1}^n i(1 - |X_i|n/T)}{\left(\sum_{i=1}^n i^2\right)^{-1/2}}.$$

Then we can reject for large values of W, which is equivalent up to an affine transformation to rejecting for small values of $\sum_i i|X_i|/T$.

The statistic W may have smaller than unit variance but it doesn't really matter since we can just calculate its conditional distribution by Monte Carlo or other numerical integration techniques, and reject when W is larger than some quantile. In fact we can simulate directly fron the conditional distribution of W, or of $\sum_i i|X_i|/T$, by simulating $D = (|X_1|, \ldots, |X_n|)/T$ from the Dirichlet(1_n) distribution, but this is not necessary to get full credit.

Alternative solution: A natural idea is to condition on the value of $T(X)$, which is a sufficient statistic for the null submodel. Rejecting for large Z is equivalent to rejecting for small values of $\sum_i i|X_i|$, so we can just simulate from the conditional distribution given $T(X)$ and reject when the statistic is above its conditional upper α quantile. It so happens this is equivalent to the first approach because D is independent of T, so the conditional distribution of $(|X_1|, \ldots, |X_n|)$ given $T = t$ is just $t \cdot D$.

2. Multivariate normal means (20 points, 5 points / part).

Suppose that we observe two multivariate normal random vectors in $\mathbb{R}^d$, for $d \geq 3$:

$$X^{(i)} \overset{\text{ind.}}{\sim} N_d(\theta^{(i)}, \sigma^2 I_d), \quad \text{for } i = 1, 2.$$

where $\theta^{(1)}, \theta^{(2)} \in \mathbb{R}^d$ and $\sigma^2 > 0$, and I_d is the $d \times d$ identity matrix.

For all parts below, if you refer to quantiles of a t, χ^2, or F distribution, you will need to **give the relevant degrees of freedom** in order to receive full credit.

(a) Assume (for this part **only**) that σ^2 is known but $\theta^{(1)}, \theta^{(2)}$ are unknown, and suggest a test of the hypothesis $H_0: \theta^{(1)} = \theta^{(2)}$ (that $\theta_j^{(1)} = \theta_j^{(2)}$ for *every* $j = 1, \ldots, d$) against the hypothesis that $\theta^{(1)} \neq \theta^{(2)}$ (that $\theta_j^{(1)} \neq \theta_j^{(2)}$ for *at least one* $j = 1, \ldots, d$). Give your test statistic and a rejection cutoff in terms of a quantile of a χ^2 distribution.

Solution:
We can define the variable $Y = X^{(2)} - X^{(1)} \sim N_d(\theta_j^{(2)} - \theta_j^{(1)}, 2\sigma^2 I_d)$. Under the null hypothesis, $\frac{1}{2\sigma^2}\|Y\|^2 \sim \chi_d^2$, so we can reject when this statistic is above the upper-α quantile of that distribution.

(b) Now drop the assumption that σ^2 is known (i.e. now it is **unknown**), and assume (for this and the next part **only**) that $\theta_j^{(2)} = \theta_j^{(1)} + \delta$, for some $\delta \in \mathbb{R}$ (i.e. every coordinate is shifted by the same amount δ), but apart from this assumption, both $\theta^{(1)}$ and $\theta^{(2)}$ are unknown. Propose a finite-sample test of $H_0: \delta = 0$ against the two-sided alternative $H_1: \delta \neq 0$. Give a test statistic and cutoffs in terms of a quantile of a specific distribution.

Note: You do *not* need to prove any optimality properties for your test, but you won't receive full credit if you trivialize the problem by giving an inefficient test, even if the test is valid in the Type I error sense.

Solution:
Working with the same Y, we have $Y \sim N_d(\delta 1_d, 2\sigma^2 I_d)$. This is just the setup for a one-sample t-test, so we reject if $\frac{|\overline{Y}|}{\sqrt{S^2/d}}$ is above the upper-$\alpha/2$ quantile of a t_{d-1} distribution, where $\overline{Y} = \frac{1}{d}\sum_j Y_j \sim N(\delta, 2\sigma^2/d)$ and $S^2 = \frac{1}{d-1}\sum_j (Y_j - \overline{Y})^2 \sim \frac{2\sigma^2}{d-1}\chi_{d-1}^2$, independently.

(c) Under the same assumptions as in part (b), propose a confidence interval for δ. If you didn't solve part (b), or if you are not confident in your answer, you may assume that there is a valid test statistic from part (b) of the form $T(X^{(1)}, X^{(2)})$, and (non-data-dependent) cutoff values $c_1(\alpha)$ and $c_2(\alpha)$ (so the test rejects if $T < c_1$ or $T > c_2$), and give your answer in terms of these.

Solution:
If we want to test the point null $\delta = \delta_0$, we can just shift the problem to get $Y - \delta_0 1_d \sim N_d((\delta - \delta_0)1_d, 2\sigma^2 I_d)$. So we want to invert the test that rejects when $\frac{|\overline{Y} - \delta_0|}{\sqrt{S^2/d}} > t_{d-1}(\alpha/2)$. In other words, our interval should be

$$\overline{Y} \pm \sqrt{S^2/d} \cdot t_{d-1}(\alpha/2).$$

(d) (*) Now, drop the assumption about δ from the previous parts, so $\theta^{(1)}$ and $\theta^{(2)}$ are again completely unknown. And now assume (for this part **only**) that $\sigma^2 = 1$. Also, suppose that we believe $\theta^{(1)} \approx \theta^{(2)}$ as vectors in $\mathbb{R}^d$, but we do not have any other strong priors about it. Suggest an estimator that will have MSE less than $2d$ for all values of $\theta^{(1)}, \theta^{(2)}$, but which will have MSE $d + 2$ whenever $\theta^{(1)} = \theta^{(2)}$.

Note: You do not need to prove that your estimator has these properties, it is sufficient to give a correct functional form for the estimator.

Solution:
Let $Z = Y/\sqrt{2}$ so it has an identity covariance. We want to use a James-Stein estimator for the mean of Z (which will be 0 if we're lucky) and the MLE for the mean of $W = (X^{(1)} + X^{(2)})/\sqrt{2} \sim N_d(\theta^{(1)} + \theta^{(2)}, I_d)$. Note Z and W are independent, which we can verify by noting that they represent two orthogonal projections of $(X^{(1)}, X^{(2)})$. Let μ denote the mean of Z, and ν the mean of W. Then $\hat{\nu} = W$ and

$$\hat{\mu} = \left(1 - \frac{d-2}{\|Z\|^2}\right) Z = \left(1 - \frac{2d-4}{\|X^{(2)} - X^{(1)}\|^2}\right) \frac{X^{(2)} - X^{(1)}}{\sqrt{2}}.$$

Since $\theta^{(2)} = (\mu + \nu)/\sqrt{2}$ and $\theta^{(1)} = (-\mu + \nu)/\sqrt{2}$, we have

$$\hat{\theta}^{(2)} = \frac{\hat{\mu} + W}{\sqrt{2}} = \frac{X^{(1)} + X^{(2)}}{2} + \left(1 - \frac{2d-4}{\|X^{(2)} - X^{(1)}\|^2}\right) \frac{X^{(2)} - X^{(1)}}{2},$$

and

$$\hat{\theta}^{(1)} = \frac{-\hat{\mu} + W}{\sqrt{2}} = \frac{X^{(1)} + X^{(2)}}{2} - \left(1 - \frac{2d-4}{\|X^{(2)} - X^{(1)}\|^2}\right) \frac{X^{(2)} - X^{(1)}}{2},$$

The MSE for estimating $(\theta^{(1)}, \theta^{(2)})$ is the sum of the MSEs for the two estimators $\hat{\mu}$ and $\hat{\nu}$. The second has MSE d and the first has MSE strictly less than d, equaling 2 if $\mu = 0$.

3. Nonparametric two-sample problem (20 points, 5 points / part).

Some useful facts you may assume are true for this problem:

- In the *one-sample* model with $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P$, with the X_i observations taking values in $\mathbb{R}$ and no further assumptions on the distribution P, the order statistics
$$S(X) = (X_{(1)}, \dots, X_{(n)})$$
are complete sufficient.

- If $Z_n \Rightarrow Z$ and $W_n \Rightarrow W$, and Z_n is independent of W_n for every n, then $(Z_n, W_n) \Rightarrow (Z, W)$ where Z is independent of W.

Assume we have a non-parametric two-sample problem of the form

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P, \quad \text{and} \quad Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} Q,$$

independently, where all X_i and Y_i take values in $\mathbb{R}$. You may assume, without proving, that $(S(X), S(Y))$ is complete sufficient for the full model.

(a) Define the estimand $g(P, Q) = \mathbb{P}_{X\sim P, Y\sim Q}(X > Y)$, i.e. the probability that an observation from P is larger than an independent observation from Q. Find the UMVU estimator for $g(P, Q)$ and explain why it is UMVU.

Solution:
We have an unbiased estimator in $1\{X_1 > Y_1\}$, so all we need to do is Rao-Blackwellize it conditional on $S(X)$ and $S(Y)$. Conditional on these statistics, X_1 and Y_1 are independently uniform draws from $S(X)$ and $S(Y)$, so the UMVU estimator is
$$\frac{1}{n^2} \sum_{i,j=1}^{n} 1\{X_i > Y_j\}.$$

Common errors: Note that $\frac{1}{n}\sum_i 1\{X_i > Y_i\}$ would not be correct; it is unbiased but cannot be calculated from the complete sufficient statistic. And $\frac{1}{n}\sum_i 1\{X_{(i)} > Y_{(i)}\}$ is not even unbiased. However, the estimator $\frac{1}{n^2}\sum_{i,j=1}^n 1\{X_{(i)} > Y_{(j)}\}$ is correct; students who gave this answer or another equivalent answer got full credit as long as each of the n^2 pairs of X and Y values are represented once.

(b) Define $\mu = \mathbb{E}_P X$, $\nu = \mathbb{E}_Q Y$, $\sigma^2 = \text{Var}_P(X)$, and $\tau^2 = \text{Var}_Q(Y)$. Show that $T(X, Y) = (\overline{X}/\overline{Y})^2$ is a consistent estimator for $\theta = (\mu/\nu)^2$ as $n \to \infty$, assuming $\nu > 0$ and $\sigma^2, \tau^2 \in (0, \infty)$.

Solution:
This follows from the continuous mapping theorem after observing that $\overline{X} \to \mu$ and $\overline{Y} \to \nu$ in probability, by the law of large numbers, and the function $f(x, y) = (x/y)^2$ is continuous everywhere except where $y = 0$.

Common "error": The function f is not continuous everywhere. If you said or implied in your answer that the function f was continuous, without the caveat that the denominator must be nonzero (i.e. that $\nu > 0$), I deducted 1 point out of 5. I made exceptions in rare cases where it seemed that something else in the answer made implicit reference to $\nu > 0$ being necessary.

(c) Give the asymptotic distribution of $T(X, Y)$ as $n \to \infty$, appropriately normalized so that the error has a nondegenerate distribution, and justify your answer. Your answer should be given as a distribution whose parameters are explicit functions of $\mu, \nu, \tau^2, \sigma^2$, and θ.

Solution:
For this, we have $\sqrt{n}(\overline{X}_n - \mu) \Rightarrow N(0, \sigma^2)$ and $\sqrt{n}(\overline{Y}_n - \nu) \Rightarrow N(0, \tau^2)$, so we can apply the delta method to $\sqrt{n}\left(\binom{\overline{X}_n}{\overline{Y}_n} - \binom{\mu}{\nu}\right) \Rightarrow N_2(0, D)$ for $D = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix}$. The function is $f(x, y) = (x/y)^2$, whose gradient (for $y \neq 0$) is $\nabla f(x, y) = (2x/y^2, -2x^2/y^3)$. As a result we have

$$\sqrt{n}(T - \theta) \Rightarrow N(0, \omega^2),$$

where

$$\omega^2 = \nabla f(\mu, \nu)' D \nabla f(\mu, \nu) = 4\left(\frac{\mu^2\sigma^2}{\nu^4} + \frac{\mu^4\tau^2}{\nu^6}\right) = \frac{4}{\nu^2}(\theta\sigma^2 + \theta^2\tau^2).$$

Grading note: If you missed that we need $\nu > 0$ in (b) I didn't take points off again. If you *didn't* miss it in (b), then I assumed you knew it in (c). So this detail played no role in grading this part.

(d) (*) If $\mu = \nu = 0$, give the asymptotic distribution of $T(X, Y)$ as $n \to \infty$, normalized appropriately if necessary. Justify your answer.

Solution:
In this case we have $Z = \sqrt{n}\binom{\overline{X}_n}{\overline{Y}_n} \Rightarrow N_2(0, D)$. By the continuous mapping theorem, then, $T = (Z_1/Z_2)^2 \Rightarrow \frac{\sigma^2}{\tau^2} F_{1,1}$.

Note: This part was graded more leniently than part (b) since I essentially gave full marks to anyone who got to the right answer, but there is a slight subtlety about applying the continuous mapping theorem here, which is worth mentioning. Why shouldn't we care about the discontinuity point anymore? The condition for the theorem can be relaxed to say just that the set of discontinuity points of the function f has measure zero in the limiting probability distribution (this would not be true in part (b) where we were appealing to the fact that $\overline{X}_n \to \nu$ in probability. If you noticed this issue (and I don't think anyone mentioned it for this part) and wanted to show convergence more directly, you could replace T with a truncated version $T_B(X, Y) = \min(T, B)$, since $f(x, y) = \min((x/y)^2, B)$ is continuous. Then we would have $T_B \Rightarrow \min(F_{1,1}, B)$ for every B, which means the cdf of T is converging everywhere to the (continuous) $F_{1,1}$ cdf. But again, this level of detail was not necessary for full marks.

4. Bayes estimation for Uniform Scale family (20 points, 5 points / part).

Some useful facts for this problem:

- The Unif$[0, \theta]$ distribution for $\theta > 0$ has density

$$p_\theta(x) = \frac{1}{\theta}, \quad \text{for } x \in [0, \theta].$$

  Its mean and variance are $\theta/2$ and $\theta^2/12$.

- The Pareto distribution with minimum value $x_0 > 0$ and shape parameter $\alpha > 0$ is called Pareto(x_0, α) and has density

$$p_{x_0,\alpha}(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}, \quad \text{for } x \geq x_0.$$

  Its mean is $\frac{\alpha x_0}{\alpha - 1}$ if $\alpha > 1$ and is infinite otherwise, and its variance is $\frac{\theta_0^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ if $\alpha > 2$ and infinite otherwise.

Assume that we observe a uniformly distributed random variable

$$X \sim \text{Unif}[0, \theta].$$

Assume for parts (a) - (b) below that the relevant loss is the standard squared error loss $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$.

(a) Show that $\theta \sim$ Pareto(θ_0, α) is conjugate to this family and find the posterior distribution and Bayes estimator for θ.

Solution:
The prior is $\lambda(\theta) \propto_\theta \frac{1}{\theta^{\alpha+1}} 1\{\theta \geq \theta_0\}$, and the likelihood is $p_\theta(x) = \frac{1}{\theta} 1\{x \leq \theta\}$, so the posterior distribution is

$$\begin{aligned} \lambda(\theta \mid x) &\propto_\theta \frac{1}{\theta^{\alpha+2}} 1\{\theta \geq \theta_0\} 1\{\theta \geq x\} \\ &= \frac{1}{\theta^{\alpha+2}} 1\{\theta \geq \max(x, \theta_0)\} \\ &\propto_\theta \text{Pareto}(\max(x, \theta_0), \alpha + 1). \end{aligned}$$

As a result the posterior mean (which is the Bayes estimator for squared error loss) is $\hat{\theta} = \frac{1+\alpha}{\alpha} \max(\theta_0, X) = (1 + 1/\alpha) \max(\theta_0, X)$.

Common mistake: The posterior is not Pareto$(\theta_0, \alpha+1)$ (if it were, it wouldn't depend on the data). A good number of students forgot to mind the indicators.

I think people made that mistake because we have often been lackadaisical in class and on homework about keeping explicit track of the support of distributions. It is usually fine not to worry about the support, since there is usually a base measure for the family that determines the support for all densities in the problem, and so it goes without saying that all densities we work with for that problem have the same support. But for both the uniform and Pareto families in this problem, the support depends on the parameter so we have to keep track of it if we want to get the calculations right.

(b) Next consider the prior $\lambda(\theta) = 2\theta \cdot 1\{0 \leq \theta \leq 1\}$. Find the Bayes estimator and Bayes risk.

Solution:
The posterior is

$$\lambda(\theta \mid x) \propto_\theta 2\theta \cdot 1\{\theta \leq 1\} \cdot \frac{1}{\theta} 1\{x \leq \theta\} = 2 \cdot 1\{x \leq \theta \leq 1\} \propto_\theta \text{Unif}[x, 1].$$

The Bayes estimator is therefore $(1 + X)/2$, and the MSE is

$$\begin{aligned} \text{MSE}(\theta) &= \left(\mathbb{E}_\theta \left[\frac{1+X}{2} \right] - \theta \right)^2 + \text{Var}_\theta \left(\frac{1+X}{2} \right) \\ &= \left(\frac{1}{2} - \frac{3\theta}{4} \right)^2 + \theta^2/48 \\ &= \left(\frac{9}{16} + \frac{1}{48} \right) \theta^2 - \frac{3}{4}\theta + \frac{1}{4} \\ &= \frac{7}{12}\theta^2 - \frac{3}{4}\theta + \frac{1}{4}, \end{aligned}$$

and the Bayes risk is

$$\begin{aligned} \int_0^1 2\theta \left(\frac{7}{12}\theta^2 - \frac{3}{4}\theta + \frac{1}{4} \right) d\theta &= \int_0^1 \left(\frac{7}{6}\theta^3 - \frac{3}{2}\theta^2 + \frac{1}{2}\theta \right) d\theta \\ &= \frac{7}{24} - \frac{1}{2} + \frac{1}{4} = \frac{1}{24}. \end{aligned}$$

Common mistake: Many of the same people who got part (a) wrong got this wrong too, giving $\text{Unif}[0, 1]$ as the posterior. I felt bad taking points off again for a similar mistake so I gave partial credit, but not too much; it is a failure of statistical intuition to think that the data is not going to determine the posterior in this problem.

(c) (*) Is the minimax risk for this problem finite? Show that it is infinite or find an upper bound on the minimax risk.

(**Hint:** It might help to consider a subproblem where θ is bounded above by $B > 0$.)

Solution:
Consider the prior $\frac{2\theta}{B}1\{0 \leq \theta \leq B\}$. We can repeat essentially the same calculation as in (b) to get that the estimator is $(X + B)/2$.

But if $Y = X/B$ and $\zeta = \theta/B$, then we have $\zeta \sim 2\zeta \cdot 1\{\zeta < 1\}$, and $Y \mid \zeta \sim \text{Unif}[0, \zeta]$, and the Bayes risk is

$$\mathbb{E}[((X + B)/2 - \theta)^2] = B^2\mathbb{E}[((Y + 1)/2 - \zeta)^2] = B^2/24.$$

Because any Bayes risk is a lower bound for the minimax risk, the minimax risk must be infinite.

(d) Now consider instead the squared relative error loss $L(\hat{\theta}, \theta) = \left(\frac{\hat{\theta}-\theta}{\theta}\right)^2$. Find the best linear estimator; i.e. if we take our estimator as aX, for $a > 0$, find the a that minimizes the corresponding risk and give the risk as a function of θ.

Solution:
The risk is

$$\begin{aligned} R(\theta) &= \frac{1}{\theta^2}\mathbb{E}\left[(aX - \theta)^2\right] \\ &= \frac{1}{\theta^2}\left[(a\mathbb{E}_\theta X - \theta)^2 + a^2\text{Var}_\theta(X)\right] \\ &= (a/2 - 1)^2 + a^2/12 \\ &= (1/4 + 1/12)a^2 - a + 1 \\ &= a^2/3 - a + 1. \end{aligned}$$

Differentiating, we find the minimum is at $a = 3/2$, giving constant risk of $1/4$.