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 (30 points, 5 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 $i = 1, \ldots, n$.

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

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

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

(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

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

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.

(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.

Problem 1 answers continued (1):

Problem 1 answers continued (2):

Problem 1 answers continued (3):

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.

(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.

(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.

(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.

Problem 2 answers continued (1):

Problem 2 answers continued (2):

Problem 2 answers continued (3):

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 $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P$, with the X_i observations taking values in $\mathbb{R}$ and no further assumptions on the distribution P, the order statistics
$$S(X) = (X_{(1)}, \ldots, X_{(n)})$$
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

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

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.

(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 $\sigma^2, \tau^2 \in (0, \infty)$.

(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.

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

Problem 3 answers continued (1):

Problem 3 answers continued (2):

Problem 3 answers continued (3):

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 $\frac{x_0^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ 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) - (c) 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.

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

(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 $B > 0$.)

(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.

Problem 4 answers continued (1):

Problem 4 answers continued (2):

Problem 4 answers continued (3):