Student ID (NOT your name):

Final Examination: QUESTION BOOKLET

Prof. Will Fithian

Fall 2020

- 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. Poisson minimax estimation (24 points, 4 points / part).

Some useful facts for this problem:

- For $\theta > 0$, the Poisson density for $X \sim \text{Pois}(\theta)$ is $\frac{\theta^x e^{-\theta}}{x!}$ on $x = 0, 1, \ldots$. The mean and variance are both θ.
- The Gamma density for $X \sim \text{Gamma}(k, \beta)$, where $\beta > 0$ is the *rate* parameter, is

$$\frac{\beta^k}{\Gamma(k)} x^{k-1} e^{-\beta x}, \qquad \text{on } x > 0,$$

  where $\Gamma(k) = \int_0^\infty z^{k-1} e^{-z}\, dz$. The mean and variance are k/β and k/β^2, respectively.
- If $X \sim \text{Gamma}(k, \beta)$ (in the rate parameterization) with $k > 1$, then $\mathbb{E}[X^{-1}] = \beta/(k-1)$.

Consider estimating θ given a single Poisson observation $X \sim \text{Pois}(\theta)$ using the loss function

$$L(d, \theta) = \frac{(d - \theta)^2}{\theta}.$$

Throughout this problem, unless otherwise specified, the risk of a given estimator is always calculated using this loss.

(a) Find the MLE and calculate its risk function.

(b) Show that $\theta \sim \text{Gamma}(k, \beta)$ is a conjugate prior for this problem and give the posterior distribution.

(c) Find the Bayes estimator for the prior from part (b) and the loss L defined above.

(d) (*) Show that the Bayes risk of the Bayes estimator from part (c) is $1/(1 + \beta)$

(e) Show that the MLE is minimax relative to the loss L.

(f) Show that the minimax risk for the usual squared error loss — i.e., $L_{\text{SE}}(d, \theta) = (d - \theta)^2$ — is infinite (this motivates changing the loss function to our L, which "adjusts" for the hardness of the problem).

Problem 1 answers continued (1):

Problem 1 answers continued (2):

Problem 1 answers continued (3):

2. ANOVA with random effects (25 points, 5 points / part).

Some useful facts for this problem:

- For $\sigma > 0$ and $\mu \in \mathbb{R}$, the Gaussian density for $X \sim N(\mu, \sigma^2)$ is $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$. If $\sigma^2 = 0$ then $X = 0$ almost surely.
- For a positive integer k, the χ^2_k density for $X \sim \chi^2_k$ is

$$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}.$$

  Its mean and variance are k and $2k$, respectively.

Assume we observe X_{ij} for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, and consider the hierarchical Gaussian model

$$\alpha_i \overset{\text{i.i.d.}}{\sim} N(0, \tau^2)$$
$$X_{ij} \mid \alpha \overset{\text{ind.}}{\sim} N(\mu + \alpha_i, \sigma^2).$$

Define the following quantities for the purposes of this problem:

$$\overline{X}_i = \frac{1}{n} \sum_j X_{ij},$$
$$S_i^2 = \frac{1}{n-1} \sum_j (X_{ij} - \overline{X}_i)^2,$$
$$\overline{X} = \frac{1}{nm} \sum_{i,j} X_{ij}, \quad \text{and}$$
$$S_B^2 = \frac{1}{m-1} \sum_i (\overline{X}_i - \overline{X})^2 \quad \text{(the } B \text{ stands for "between groups").}$$

The parameters $\mu \in \mathbb{R}, \tau^2 \geq 0$, and $\sigma^2 > 0$ are unknown. $\alpha_1, \ldots, \alpha_m$ are unobserved random variables but they are not parameters, and the model could be rewritten without them.

In this problem, unless otherwise stated, you do **NOT** need to show tests and confidence intervals are UMP(U) or UMA(U). Where I ask you to give an explicit formula, it is fine for the formula to be in terms of quantiles of one or more distributions from class.

(a) Show that $S_B^2, S_1^2, \ldots, S_m^2$ are mutually independent and give their distribution.

(b) Find a finite-sample, equal-tailed confidence interval for μ. Give an explicit formula.

(c) Give a finite-sample test of the null hypothesis $H_0 : \ \tau^2 = 0$ vs $H_1 : \tau^2 > 0$. Give an explicit formula for the test statistic and the critical value.

(d) (*) Find a finite-sample, equal-tailed confidence interval for τ^2/σ^2. Give an explicit formula.

(e) (*) Show that the model (with the additional restriction that $\tau^2 > 0$) is a three-parameter exponential family and $\left(\overline{X},\, S_B^2,\, \sum_i S_i^2\right)$ is a complete sufficient statistic.

Problem 2 answers continued (1):

Problem 2 answers continued (2):

Problem 2 answers continued (3):

3. "And if you ever saw it..." (24 points, 6 points / part).

Some useful facts for this problem:

- For $n \in \{0, 1, \ldots\}$ and $p \in [0, 1]^d$ with $\sum p_i = 1$, the multinomial density for $X \sim \text{Multinom}(n, p)$ is

$$p_1^{x_1} \cdots p_d^{x_d} \frac{n!}{x_1! \cdots x_d!}, \qquad \text{on } x \in \{0, \ldots, n\}^d \text{ with } \sum_i x_i = n.$$

An ecologist is interested in estimating the total population of reindeer in a wildlife preserve near the North Pole. She makes two visits to the preserve on two consecutive days and looks for reindeer. Each time she finds a reindeer she marks it with a unique identifying tag, so she can tell if she sees the same reindeer twice (in ecology this type of study is called a *capture-recapture* or *mark-recapture* study).

Assume that the same population of n of reindeer is present in the preserve on both days, and each reindeer on each day has the same probability $\pi \in (0, 1)$ of being seen by her, independently across the reindeer and the days (so the detections / non-detections are like $2n$ i.i.d. "coin flips" each with success probability π). Note that n is the unknown parameter of interest and π is an unknown nuisance parameter.

Let N_{11} denote the number of reindeer she sees both days, N_{10} the number she sees the first day not the second, and N_{01} the number she sees the second day but not the first. (Note that N_{00}, the number of reindeer she sees on neither day, is not observed.)

(a) Write down the likelihood for the model as a function of N_{01}, N_{10}, and N_{11} and show that $T = (N_{01} + N_{10}, N_{11})$ is a sufficient statistic for the model.

You do **NOT** need to show a sufficiency reduction from the Bernoulli model of detected/non-detected "coin flips" for each reindeer-day; after all we do not really get to observe the data for that model because we don't know how many reindeer went undetected on both days. Just start with N_{01}, N_{10}, N_{11} as the data and n and π as the parameters.

(b) (*) Show that T is minimal sufficient (for this part you may assume we already know it is sufficient).

(c) Define the estimator

$$\hat{n} = \frac{(N_{01} + N_{10} + 2N_{11})^2}{4N_{11}}.$$

Show that $\hat{n}$ is consistent in the sense that $\hat{n}/n \xrightarrow{p} 1$ as $n \to \infty$ with π fixed.

(d) Find the asymptotic distribution of $\hat{n}$ from part (c) as $n \to \infty$ with π fixed. You should center and scale appropriately so that it has a non-degenerate limiting distribution (that is, after centering and scaling it shouldn't converge in probability to a constant).

Problem 3 answers continued (1):

Problem 3 answers continued (2):

Problem 3 answers continued (3):

4. Nonlinear regression (24 points, 6 points / part).

Note the Gaussian density is printed in the preamble of Problem 2.

We are given a sample of n pairs (x_i, Y_i) where $x_1, \ldots, x_n \in \mathbb{R}$ are fixed real numbers and

$$Y_i = g(\alpha + \beta x_i) + \varepsilon_i, \quad \text{where } \varepsilon_i \overset{\text{ind.}}{\sim} N(0, \sigma^2 h(x_i)).$$

Assume (except where otherwise specified) that:

- $g : \mathbb{R} \to \mathbb{R}$ is a known function which is strictly increasing and infinitely differentiable.
- $h : \mathbb{R} \to (0, \infty)$ is a known continuous function.
- $\alpha, \beta \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown

Finally, let $r_i = Y_i - g(\hat{\alpha} + \hat{\beta} x_i)$ denote the ith residual. Throughout the problem, assume we are estimating the parameter vector $(\alpha, \beta, \sigma^2)$ jointly by maximum likelihood; let $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ denote the joint MLE.

(a) Show that the MLE for α and β is found by setting weighted averages of the residuals to 0:

$$\sum_{i=1}^{n} w_i r_i = \sum_{i=1}^{n} w_i r_i x_i = 0,$$

and give explicit expressions for the weights w_i in terms of the data, the functions g and h, and the maximum likelihood estimators $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$.

(b) Give an explicit expression for the MLE for σ^2, i.e. $\hat{\sigma}^2$, in terms of the data, the functions g and h, and the maximum likelihood estimators $\hat{\alpha}, \hat{\beta}$.

(c) (*) Now assume (for this part **ONLY**) that instead of fixed numbers we observe i.i.d. random variables $X_1, \ldots, X_n$, which are continuous and bounded random variables ($|X_i| \leq B$ almost surely, for some $B > 0$.) Give the asymptotic distribution of the maximum likelihood estimators $(\hat{\alpha}, \hat{\beta})$ in terms of the functions g and h, and expectations of suitable random variables. The limit is taken as $n \to \infty$ with the other parameters fixed.

You may assume without proof that $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ are consistent for the true population values, and that all of the regularity conditions from class for our theorem on the asymptotic distribution of the MLE hold (the log-likelihood and its derivatives are well-behaved in the required sense). You do not need to write down what the conditions are, either.

(**Hint:** it might be easier to do the problem assuming σ^2 is known, and then explain why the answer doesn't change when σ^2 is unknown.)

(d) (*) We now go back to assuming the x_i values are fixed. Now assume $h(z) \equiv 1$ but g is completely unknown (apart from the restrictions described in the preamble: strictly increasing and infinitely differentiable). Give a finite-sample test of $H_0: \ \beta \leq 0$ vs $H_1: \ \beta > 0$. You should provide a test statistic and describe how to calculate the critical value. For full credit you must show your test controls the rejection probability throughout the composite null hypothesis (that is, for all valid choices of g, α, and σ^2.)

Since we are already using the letter α for the intercept, I suggest using a to denote the significance level in your answer.

Problem 4 answers continued (1):

Problem 4 answers continued (2):

Problem 4 answers continued (3):