STT 872, 867-868 Fall Preliminary Examination
Tuesday, August 18, 2020
12:30 - 5:30 pm

INSTRUCTIONS:

1. This examination is closed book. Every statement you make must be substantiated. You may do this either by quoting a theorem/result and verifying its applicability or by proving things directly. You may use one part of a problem to solve the other part, even if you are unable to solve the part being used. A complete and clearly written solution of a problem will get a more favorable review than a partial solution.

2. You must start solution of each problem on a separate page. Be sure to put the number assigned to you on the top left corner of every page of your solution. Also please number the pages with n/m (top right corner), where n is the current page number and m is the total number of pages, to keep the ordering and to avoid missing any pages during scanning.

3. In ZOOM, the video must be turned on for the whole duration of the exam, while the microphone must be muted for the whole duration of the exam. There should be no other people present in the room during the exam. DO NOT use virtual background. The camera should show a wide angle with you and the desk where your work is visible.

4. If you have questions during the exam (e.g. bathroom break requests) you can send a chat message in ZOOM to the host. Email/cell phone communication with Tami would be a back-up method to ZOOM/ D2L if they fail.

5. The exam will last 5 hours. Additional 30 minutes will be allowed to organize the paper solution (write your assigned number and the page number (n/m) on each page), scan it and upload to D2L. Submit your solution as a PDF file. Before the submission, make sure the PDF is clearly readable and it contains all your answers (check on your laptop). Failing to do so may result in substantial loss of points. Keep your paper solution until the examination result is out. If you run into any upload issues, email your solutions to Tami directly.

6. Please refrain from discussing the exam in any way before the results are made available.

1. Let $X_1, \cdots, X_n$ are i.i.d. $B(a,b)$, i.e. the common density has the support $[0, 1]$, where it is proportional to $x^a(1-x)^b, a > 0, b > 0$. Suppose $a = b$.

(a) (3 pts) Find a minimal sufficient statistic for a.

(b) (3 pts) Is the statistic in part 1(a) complete?

(c) (3 pts) Find the MLE of a.

(d) (3 pts) Find the asymptotic distribution of the MLE.

2. Let $X_1, \cdots, X_n$ be a sample from the Poisson(λ) distribution truncated on the left at 0.

(a) (4 pts) Find the UMVU of λ. Hint: $(e^\lambda - 1)^n = \sum_{k=0}^{\infty} \frac{n!\lambda^k}{k!} C_{k,n}$ where $C_{k,n}$ are Stirling's number of the second kind.

(b) (4 pts) Find the MLE of λ and its asymptotic variance.

(c) (3 pts) Find the Cramer-Rao lower bound for the variance of unbiased estimators of λ.

(d) (2 pts) Do the UMVU or the MLE attain their information lower bound?

3. Consider estimation of unknown parameters $\theta_1, \cdots, \theta_p$ based on data $X_1, \cdots, X_p$ that are independent with $X_i \sim U(0, \theta_i), i = 1, \cdots, p$ under the squared error loss $L(\theta, d) = \sum_{i=1}^{p} (\theta_i - d_i)^2$.

(a) (4 pts) Following a Bayesian approach, model the unknown parameters as random variables $\Theta_1, \cdots, \Theta_p$ which are i.i.d. and absolutely continuous with a common density

$$x\lambda^2 \mathbf{I}_{(0,\infty)}(x) e^{-\lambda x}.$$

Find the Bayes estimators for $\Theta_1, \cdots, \Theta_p$.

(b) (3 pts) Suggest an empirical estimate of λ based on the sample average $\bar{X}$.

(c) (4 pts) Compute the risk of the empirical Bayes estimate.

4. Consider testing for $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ based on a single observation X from $N(\theta, 1)$. Using the apparent symmetry of this testing problem, it seems natural to base a test on $Y = |X|$.

(a) (3 pts) Find densities q_θ for Y and show that the distribution for Y depends only on $|\theta|$.

(b) (3 pts) Show that the densities $q_\theta, \theta \geq 0$, have monotone likelihood ratios.

(c) (3 pts) Find the uniformly most powerful level α test of H_0 versus H_1 based on Y.

(d) (4 pts) The uniformly most powerful test $\psi^*(Y)$ in part 4(c) is not most powerful compared with tests based on X. Find a level α test $\psi(X)$ with a better power at $\theta = -1$: $E_{-1}\psi(X) > E_{-1}\psi^*(Y)$. What is the difference in power at $\theta = -1$ if $\alpha = 5\%$?

5. Consider the linear regression model

$$Y = X\beta + \epsilon,$$

where $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}, \beta \in \mathbb{R}^p, \epsilon \in \mathbb{R}^n$, and $\epsilon_i' s$ are independent with $\mathbb{E}(\epsilon_i) = 0, \mathrm{var}(\epsilon_i) = \sigma^2$. (Parts (a)(b)(c) are separate problems and not related)

(a) Suppose the observed data is $(\tilde{X}, Y)$ instead of (X, Y), where $\tilde{X} = X + \Delta$ with $\Delta \in \mathbb{R}^{n\times p}$ being the rounding errors when recording the values of the covariates. These errors Δ are determined by the actual values X and some consistent rounding rule, thus are (unknown) constants rather than random variables. Assume both X and $\tilde{X}$ are of full column rank. The least square estimate is denoted by $\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}Y$.

(a1) (3 pts) Show that the usual MSE $= \frac{1}{n-p}\|Y - \tilde{X}\tilde{\beta}\|_2^2$ can be biased for estimating σ^2.

(a2) (3 pts) For an unknown linear functional $\ell'\beta$ with $\ell \in \mathbb{R}^p$, the relative bias of $\ell'\tilde{\beta}$ as an estimator of $\ell'\beta$ is defined as $\mathrm{RB}(\ell'\tilde{\beta}) = |\mathrm{Bias}(\ell'\tilde{\beta})|/\sqrt{\mathrm{var}(\ell'\tilde{\beta})}$. Prove that $\mathrm{RB}(\ell'\tilde{\beta}) \leq \frac{1}{\sigma}\sqrt{\beta'\Delta'\Delta\beta}$. How do you interpret the result?

(b) For a given nonempty subset $\mathcal{M} \subseteq \{1, 2, \ldots, p\}$, let $\hat{\beta}_{\mathcal{M}}$ be the OLS based on the variables indexed by $\mathcal{M}$.

(b1) (4 pts) Compute the expected in-sample prediction error of $\hat{\beta}_{\mathcal{M}}$, denoted by $\Gamma_{\mathcal{M}}$. Compute the corresponding leave-one-out cross validation estimate of the prediction error, denoted by $\hat{\Gamma}^{cv}_{\mathcal{M}}$ (Hint: The Woodbury matrix identity is $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, where $A \in \mathbb{R}^{n\times n}, U \in \mathbb{R}^{n\times k}, C \in \mathbb{R}^{k\times k}, V \in \mathbb{R}^{k\times n}$).

(b2) (3 pts) Assume as $n \to \infty$, $\frac{1}{n}X'X \to \Sigma \succ 0$ and $\max_{1\leq i\leq n} x_i'(X'X)^{-1}x_i \to 0$ where x_i is the i^{th} row of X. For any given $\mathcal{M} \subseteq \{1, 2, \ldots, p\}$, prove that $\hat{\Gamma}^{cv}_{\mathcal{M}} \xrightarrow{P} \Gamma_{\mathcal{M}}$, as $n \to \infty$.

(c) The Lasso estimator is given by

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1,$$

where $\lambda > 0$ and $\|\cdot\|_1$ is the usual ℓ_1 norm. It is further assumed that $\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I}_n)$ and σ^2 is known.

(c1) (3 pts) We focus on a special orthogonal design case where $X'X = \mathbf{I}_p, n = p$, and will use Lasso estimates to solve the following hypothesis testing problem

$$H_0 : \beta = 0, \quad H_1 : \beta \neq 0.$$

Consider the Lasso estimate $\hat{\beta}(\lambda^*)$ with λ^* defined as

$$\lambda^* = \sup\{\lambda > 0 : \hat{\beta}(\lambda) \text{ has at least two nonzero components}\}.$$

Construct a test statistic based on the SSE difference $\|Y\|_2^2 - \|Y - X\hat{\beta}(\lambda^*)\|_2^2$, and give the corresponding rejection region $\big($Hint: Suppose $z_1, \ldots, z_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ and let $v_1 \geq v_2 \geq \cdots \geq v_n$ be the order statistics of $|z_1|, |z_2|, \ldots, |z_n|$. Then $v_1(v_1 - v_2) \xrightarrow{d} \mathrm{Exp}(1)\big)$.

(c2) (4 pts) We focus on the orthogonal design case in which $X'X = \mathbf{I}_p$. For a given $\lambda > 0$, the Lasso estimate $\hat{\beta}(\lambda)$ selects a set of variables $\hat{\mathcal{M}} = \{1 \leq j \leq p : [\hat{\beta}(\lambda)]_j \neq 0\}$ where $[\hat{\beta}(\lambda)]_j$ denotes the j^{th} component of $\hat{\beta}(\lambda)$. An important problem is to construct confidence interval after variable selection. Adjusting for the variable selection process, derive a confidence interval $\mathcal{C}$ for $\beta_1 + \beta_2$ with $1 - \alpha$ level *conditional* coverage (conditioning on the selected variables):

$$\mathbb{P}\Big(\beta_1 + \beta_2 \in \mathcal{C} \mid \hat{M} = \{1, 2\} \text{ and } [\hat{\beta}(\lambda)]_1 < 0, [\hat{\beta}(\lambda)]_2 > 0\Big) = 1 - \alpha.$$

(Hint: Denote $F^{[a,b]}_{\mu,\tau^2}(x) = \mathbb{P}(z \le x \mid a \le z \le b)$ where $z \sim \mathcal{N}(\mu, \tau^2)$; namely, $F^{[a,b]}_{\mu,\tau^2}(\cdot)$ is the CDF of a $\mathcal{N}(\mu, \tau^2)$ random variable truncated to the interval $[a, b]$. Use $F^{[a,b]}_{\mu,\tau^2}(\cdot)$ and the OLS estimate $\hat{\beta}_1^{ols} + \hat{\beta}_2^{ols}$ to construct a *conditional* pivot)

6. Consider the linear mixed model,

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i, \quad i = 1, 2, \ldots, N,$$

where $Y_i \in \mathbb{R}^{n_i}$, $X_i \in \mathbb{R}^{n_i \times m}$, $Z_i \in \mathbb{R}^{n_i \times k}$, and all the $b_i's, \epsilon_i's$ are mutually independent. (Parts (a)(b) are separate problems and not related)

(a) Suppose $m = 1, k = 1$, and $n_i = 1, X_i = Z_i = 1, b_i \sim \mathcal{N}(0, \tau^2), \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for all $i = 1, 2, \ldots, N$. The variances $\{\sigma_i^2\}$ are assumed to be known. This specified model is called a meta-analysis model and has important applications for pooling studies in life science.

  (a1) (4 pts) Construct a $100(1-\alpha)\%$ confidence interval for τ^2.

  (a2) (4 pts) Use Wald test to solve the hypothesis $H_0 : \beta = 0, \quad H_1 : \beta \neq 0$.

(b) Suppose $\epsilon_i \sim N(0, \sigma^2 \mathbf{I}_{n_i})$. The classical linear mixed model further assumes $b_i \sim \mathcal{N}(0, \tilde{D})$. However, in many applications like longitudinal study, the data has latent clustering structures so that it is more appropriate to model the random effects by a mixture of Gaussian distributions.

  (b1) (4 pts) Suppose b_i follows a two-component Gaussian mixture distribution with pdf:

$$f(b) = \alpha \cdot g_{\mu_1, \tilde{D}}(b) + (1-\alpha) \cdot g_{\mu_2, \tilde{D}}(b),$$

  where $g_{\mu,\Sigma}(b) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(b-\mu)'\Sigma^{-1}(b-\mu)}$ denotes the pdf of $\mathcal{N}(\mu, \Sigma)$, and $\alpha \in (0, 1)$ is the proportion of the first Gaussian component in the distribution. The parameters of this mixture distribution is $(\alpha, \mu_1, \mu_2, \tilde{D})$. With the Gaussian mixture random effects, is this linear mixed model identifiable? How does it compare with the classical linear mixed model in terms of model identifiability?

  (b2) (4 pts) A growth curve data about 20 girls measures the height of each girl on a yearly basis from age 6 to 12. Suppose the 20 girls can be clustered into two categories according to whether the mother is tall or short. Without their mothers' information, such a clustering structure is latent. Propose and explain a linear mixed model from (b1) for this data.

  (b3) (4 pts) Detail out the EM algorithm for computing the MLE's of the parameters under the linear mixed model specified in (b1).