STT 872, 867-868 Spring Preliminary Examination
Wednesday, January 15, 2020
12:30 - 5:30 pm

NOTE: 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.

You must start solution of each problem on the given page. Be sure to put the number assigned to you on the right corner top of every page of your solution.

STT 872, 867-868 Spring Preliminary Examination
Wednesday, January 15, 2020
12:30 - 5:30 pm

1. Let $X_1, ..., X_n$ be i.i.d. from the density $\frac{\tau}{x^2}, x > \tau, \tau > 0$.

(a) (3 pts) Find a complete sufficient statistic for τ.

(b) (5 pts) Derive a UMP unbiased test of size $\alpha \in (0,1)$ for testing $H : \tau = 2$ vs $K : \tau \neq 2$ in as much detail as possible.

(c) (2 pts) Derive the power function of the test in (b).

(d) (2 pts) Construct a UMVUE of $\tau^2 + 2\tau$.

(e) (2 pts) Propose a consistent estimator of $\exp(\tau)$.

(f) (7 pts) Find the MRE estimator of τ under the loss function $L(\tau, d) = (d/\tau - 1)^2$.

(g) (4 pts) Let τ have density $e^{-\tau}, \tau > 0$. Find the Bayes estimator of τ under the squared loss. Explain in as much detail as possible.

2. Let X take on values 1, 2, 3, 4 with probabilities $p, 0.5 + p, p^2, 0.5 - 2p - p^2$ respectively.

(a) (2 pts) Describe all random variables $U(X)$ such that $E_p U = 0$ for all $p \in (0, 0.2)$.

(b) (3 pts) Describe all estimands $g(p)$ for which UMVUE exist.

(c) (8 pts) Consider $H : p = 0.05$ vs $K : p = 0.1$. Does a MP test at level $\alpha \in (0,1)$ based on X exist? Explain in as much detail as possible. Compute its power.

(d) (2 pts) Let $X_1, \ldots, X_n$ be iid random variables from the distribution of X. Consider $T = \frac{1}{n}\sum_{i=1}^n I(X_i = 2) - 0.5$. Describe conditions that would make T an admissible estimator of p under squared loss. Show it is inadmissible for $p \in (0, 0.2)$. Hint: also consider $S = \frac{1}{n}\sum_{i=1}^n I(X_i = 1)$.

3. This problem focuses on linear fixed effects models.

(a) Consider m simple linear regression models (of full rank),

$$y_{ki} = \alpha_k + \beta_k x_{ki} + \epsilon_{ki}, \quad i = 1, 2 \ldots, n_k, \quad k = 1, 2, \ldots, m,$$

where $\alpha_k \in \mathbb{R}$ is the intercept and $\beta_k \in \mathbb{R}$ is the slope in the kth model. Here, all the ϵ_{ki}'s are independently and identically distributed as $\mathcal{N}(0, \sigma^2)$.

(a1) (3 pts) Suppose we want to test whether the m regression lines are identical. Give the test statistic and α-level rejection region for the test.

(a2) (4 pts) Suppose the first regression line is parallel to and on the left side of the second one. Construct a $100(1-\alpha)\%$ confidence interval for the horizontal distance between the two lines.

(a3) (4 pts) Give two methods of constructing simultaneous confidence intervals for all the linear functions of the form $\{\alpha_k + \beta_k x, \forall x \in \mathbb{R}, k = 1, \ldots, m\}$, with joint coverage probability greater than or equal to $1-\alpha$.

(b) Consider the linear regression model,

$$Y = X\beta + \epsilon, \tag{1}$$

where $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}, \beta \in \mathbb{R}^p$. The matrix X is not necessarily of full rank.

(b1) (4 pts) In influence analysis, a useful measure of influence for the i^{th} data point is the Cook's distance defined as

$$D_i = \frac{\|X\hat{\beta} - X\hat{\beta}^{-i}\|_2^2}{p\hat{\sigma}^2},$$

where $\hat{\beta}$ is some estimator for β, $\hat{\beta}^{-i}$ is the same estimator computed with the i^{th} data point removed, and $\hat{\sigma}^2$ is the estimator for σ^2. Thus D_i measures the change in all of the fitted values when the i^{th} observation is deleted. For the Ridge estimator $\hat{\beta} = (X'X + \lambda I)^{-1}X'Y$ with $\lambda > 0$, prove that D_i can be simplified as

$$D_i = \frac{(y_i - x_i'\hat{\beta})^2}{p\hat{\sigma}^2} \cdot \frac{\sum_{j=1}^n h_{ij}^2}{(1-h_{ii})^2}.$$

Here, $y_i - x_i'\hat{\beta}$ is the i^{th} residual, h_{ij} is the $(i,j)^{\text{th}}$ entry of the hat matrix $X(X'X + \lambda I)^{-1}X'$. (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) (5 pts) Consider the Lasso estimator given by

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

where $\lambda > 0$ and $\|\cdot\|_1$ is the usual ℓ_1 norm. Let the j^{th} column of X be X_j. Define the equicorrelation set $\mathcal{S}$ by

$$\mathcal{S} = \Big\{ j \in \{1, 2, \ldots, p\} : |X_j'(Y - X\hat{\beta})| = \lambda \Big\}.$$

First show that the set $\mathcal{S}$ is unique (even when $\hat{\beta}$ is not unique). Suppose $X_{\mathcal{S}}$ (the submatrix of X consisting of the columns indexed by $\mathcal{S}$) is of full column rank, prove that the Lasso solution $\hat{\beta}$ is unique, and the solution has at most $\min\{n, p\}$ nonzero elements.

4. Growth curve modeling is a general framework for analyzing longitudinal data. Linear growth curve (LGC) model is one class of such models. The LGC model can be written in hierarchal fashion:

$$Y_i = Z_i d_i + \epsilon_i, \quad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}), \tag{2}$$

$$d_i = A_i \beta + b_i, \quad b_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma), \tag{3}$$

where $Y_i \in \mathbb{R}^{n_i}$, $Z_i \in \mathbb{R}^{n_i \times k}$, $A_i \in \mathbb{R}^{k \times m}$, $i = 1, 2, \ldots, N$, and $\{\epsilon_i\}_{i=1}^N$ are independent from $\{b_i\}_{i=1}^N$. The observations are $\{(Y_i, Z_i, A_i)\}_{i=1}^N$ where $\{(Z_i, A_i)\}_{i=1}^N$ are fixed. The parameters are $(\beta, \sigma^2, \Sigma)$. To ensure model identifiability, we assume $\sum_{i=1}^N A_i' A_i \succ 0$, $\{Z_i\}_{i=1}^N$ are of full column rank, and $\sum_{i=1}^N (n_i - k) > 0$.

(a1) (4 pts) Suppose $\sigma^{-2} \cdot \Sigma$ is known, give the closed form solution for the MLE of β.

(a2) (6 pts) [continued from (a1)] Suppose $\sigma^{-2} \cdot \Sigma$ is known, and consider an alternative estimator for β in the following way. According to (2), first construct OLS estimators $\hat{d}_i = (Z_i' Z_i)^{-1} Z_i' Y_i$ for the $d_i' s$. Then write out the likelihood for $\{\hat{d}_i\}_{i=1}^N$ and maximize it to obtain the MLE $\tilde{\beta}$. Prove that $\tilde{\beta}$ is equal to the MLE in (a1).

(a3) (6 pts) [continued from (a1)] Now suppose that $\sigma^{-2} \cdot \Sigma$ is unknown. Show that the MLE of β is unbiased.

(b) (4 pts) The LGC model can be considered as a special case of linear mixed models. Write down the minimum norm quadratic unbiased estimator (MINQUE) for σ^2, and prove that it is a consistent estimator as $\sum_{i=1}^N (n_i - k) \to \infty$. (Hint: The variance of a chi-squared distribution with r degrees of freedom equals $2r$)