STT 872, 867-868 Fall Preliminary Examination
Wednesday, August 21, 2024
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. Please refrain from discussing the exam in any way before the results are made available.

1. Let $X_1, \cdots, X_n$ be i.i.d. from a Geometric distribution with a parameter $p \in (0,1)$, i.e.

$$P(X_i = x) = p(1-p)^x, \;\; x = 0, 1, 2, \cdots.$$

   (1a) (4 pts) Find the MLE of $1/p^2$. Use delta-method to find its asymptotic variance, $n > 1$.

   (1b) (4 pts) Let $g(p) = 1/p^2$ and let the loss function be $\mathbb{E}(p\delta(X_1, \cdots, X_n) - 1/p)^2$. For $n > 1$, using a $\mathrm{Beta}(a, b)$ prior find a Bayes estimator of $g(p)$ where the density of Beta(a, b) prior is given by

$$\pi(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}, \;\; p \in (0,1)$$

   (1d) (3 pts) For $n = 1$, show that $I(X = 0)$ is a minimax estimator of p under the loss function $L(., p) = (. - p)^2/(p(1-p))$.

   (1c) (4 pts) Derive a UMP unbiased test of size $\alpha \in (0,1)$ for testing $H_0 : 1/3 \le p \le 2/3$ vs $H_1 : p < 1/3$ or $p > 2/3$ in the fullest possible detail, $n > 1$.

2. Let $X_1, \cdots, X_n$ follows $\mathrm{Exp}(\theta, 1)$ where the density of $\mathrm{Exp}(\theta, 1)$ is given by

$$f(x;\theta) = e^{-(x-\theta)}, x > \theta, \theta > 0.$$

   (2a) (3 pts) Show that $T_n = X_{(1)} - 1/n$ is the MRE estimator of θ under the squared error loss: $L(., \theta) = (. - \theta)^2$ for $n > 1$.

   (2b) (3 pts) Show that the inequality

$$\mathrm{var}(T_n) \ge \frac{1}{n\mathbb{E}\left[\partial/\partial\theta \log e^{-(X-\theta)}\right]^2}$$

   does not hold for any $\theta \in \mathbb{R}$ and $n > 1$. Does this conflict with Cramer-Rao lower bound?

3. Let $X_1, \cdots, X_n$ are i.i.d. from the normal distribution $N(0, \sigma^2)$ with $\sigma^2 > 0$ unknown.

   (3a) (4 pts) Show that the MLE of σ^2 is inadmissible for the squared error loss: $L(., \sigma^2) = (. - \sigma^2)^2$ for $n > 1$. (Hint: You may use $\mathbb{E}(\chi^2_{(n)}) = n$ and $\mathrm{var}(\chi^2_{(n)}) = 2n$).

   (3b) (4 pts) Find the uniformly most accurate confidence bound for σ^2 using the uniformly most powerful test for $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma > \sigma_0$ assuming $n > 1$.

4. Let a parameter $\theta \in \Theta = (1, 2)$ and $Y_1, \cdots, Y_n$ a random sample from the uniform distribution $U(\theta, 2\theta)$. Suppose that instead of $Y_1, \cdots, Y_n$, one observes $X_1, \cdots, X_n$ which are

$$X_i = \begin{cases} 2, & Y_i > 2 \\ Y_i, & Y_i \le 2 \end{cases}$$

   (4a) (3 pts) Denote the $\sigma-$finite measure $\nu = \delta + m$ in which δ is the point mass measure at $\{2\}$ and m is the Lebesgue measure. Show that the pdf of X_1 with respect to ν is

$$f_\theta(x) = \frac{2\theta - 2}{\theta} I_{\{2\}}(x) + \frac{1}{\theta} I_{(\theta,2)}(x)$$

(4b) (4 pts) Let $R = \sum_{i=1}^n I(X_i = 2)$, show that the UMVUE of $1 - \theta^{-1}$ based on $X_1, \cdots, X_n$ is $n^{-1}R/2$ for $n > 1$.

(4c) (4 pts) Based only on X_1, construct a MP test of level $\alpha \in (0, 1)$ for the hypotheses $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, where $1 < \theta_0 < \theta_1 < 2$ in the fullest possible detail.

5. Consider two **full-rank** linear regression models:

$$y_i = x_i'\beta + \epsilon_i, \quad i = 1, 2, \ldots, n,$$
$$z_i = x_i'\gamma + \xi_i, \quad i = 1, 2, \ldots, n,$$

where $y_i \in \mathbb{R}, z_i \in \mathbb{R}, x_i \in \mathbb{R}^p$. The two models share the same design matrix, but the error terms are assumed correlated. Specifically, for each $i = 1, 2, \ldots, n$, the vector (ϵ_i, ξ_i) is independently sampled from the bivariate Gaussian $\mathcal{N}(0, \Sigma)$ with the covariance matrix $\Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$. It is known that the MLE for Σ, denoted by $\hat{\Sigma} = \begin{pmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{pmatrix}$, takes the following form

$$\hat{a} = \frac{1}{n}Y'(I - X(X'X)^{-1}X')Y, \quad \hat{b} = \frac{1}{n}Y'(I - X(X'X)^{-1}X')Z,$$
$$\hat{c} = \frac{1}{n}Z'(I - X(X'X)^{-1}X')Z,$$

where $X = (x_1, \ldots, x_n)' \in \mathbb{R}^{n\times p}, \;\; Y = (y_1, \ldots, y_n)' \in \mathbb{R}^n, \;\; Z = (z_1, \ldots, z_n)' \in \mathbb{R}^n$. Further denote the OLS under the two models by $\hat{\beta}, \hat{\gamma}$, respectively.

(5a) (5 pts) Suppose $b = 0, a = 2c$ and c is unknown. Derive a α-level rejection region for the test $H_0 : \beta = \gamma, H_1 : \beta \neq \gamma$.

(5b) (5 pts) Suppose $\{a, b, c\}$ are all unknown. For a given vector $\ell \in \mathbb{R}^p$, construct a confidence interval for $\ell'\beta + 2\ell'\gamma$ with coverage probability equal to $1 - \alpha$ (Hint: build a pivot with F-distribution based on $\ell'\hat{\beta} + 2\ell'\hat{\gamma}$ and $\hat{\Sigma}$).

(5c) (5 pts) Suppose $\{a, b, c\}$ are all unknown. For a given vector $\ell \in \mathbb{R}^p$, prove that $\ell'\hat{\beta} + 2\ell'\hat{\gamma}$ is BLUE for $\ell'\beta + 2\ell'\gamma$, namely, for any linear estimator $t_1'Y + t_2'Z$ with $t_1, t_2 \in \mathbb{R}^n$ that is unbiased for $\ell'\beta + 2\ell'\gamma$, it holds that $\mathrm{Var}(t_1'Y + t_2'Z) \geq \mathrm{Var}(\ell'\hat{\beta} + 2\ell'\hat{\gamma})$.

6. Consider the model,

$$y_i = \underbrace{\beta_1 x_i^2 + \beta_2 x_i + \beta_3}_{:=f(x_i)} + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $x_i \in \mathbb{R}, \epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and the unknown parameters are $\{\beta_1, \beta_2, \beta_3\}$.

(6a) (4 pts) Is β_3 always estimable? Provide your argument. Is $\beta_1 x_1^2 + \beta_2 x_1 + \beta_3$ estimable? If so, give a confidence interval with coverage probability equal to $1 - \alpha$.

(6b) (5 pts) For any given $x \in \mathbb{R}$, the kernel method outputs the prediction $\hat{f}_h(x)$ given by

$$\hat{f}_h(x) = \frac{\sum_{i=1}^n e^{-\frac{(x-x_i)^2}{h^2}} y_i}{\sum_{i=1}^n e^{-\frac{(x-x_i)^2}{h^2}}},$$

where $h > 0$ is the bandwidth that controls the bias-variance tradeoff. First, compute the expected prediction error at a single point x_0 and describe how the bias and variance are affected as h varies. Then, give an unbiased estimator for the expected in-sample prediction error.

7. A data measures the total body bone mineral density for teenagers. We denote the measurement of the i^{th} teenager at time t_{ij} by y_{ij}, and assume the following model,

$$y_{ij} = \alpha_i t_{ij} + \epsilon_{ij}, \quad i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, n_i, \tag{1}$$

where all the ϵ_{ij}'s are independently sampled from $\mathcal{N}(0, \sigma^2)$. To model the dependence of the bone density gain on calcium intake, we further assume

$$\alpha_i = c \cdot d_i + \xi_i, \quad i = 1, \ldots, N, \tag{2}$$

where $d_i \in \mathbb{R}$ denotes the daily calcium supplement of the i^{th} teenager, and all the ξ_i's are independently sampled from $\mathcal{N}(0, \tau^2)$. The observed data consists of $\{(y_{i1}, \ldots, y_{in_i}, t_{i1}, \ldots, t_{in_i}, d_i)\}_{i=1}^N$, and the unknown parameters are $\{c, \tau^2, \sigma^2\}$.

(7a) (5 pts) Plug (2) into (1) and show that the resulting model is an example of linear mixed models. Provide conditions to ensure its identifiability.

(7b) (5 pts) Suppose the sums of squared times $\{\sum_{j=1}^{n_i} t_{ij}^2\}_{i=1}^N$ are all equal. Derive the closed-form solution for the MLE of c (Hint: the Woodbury matrix identity is $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, where $A \in \mathbb{R}^{n\times n}, U \in \mathbb{R}^{n\times k}, C \in \mathbb{R}^{k\times k}, V \in \mathbb{R}^{k\times n}$).

(7c) (6 pts) Suppose the condition in (7b) is not necessarily satisfied. Derive the EM algorithm for computing the MLE of $\{c, \tau^2, \sigma^2\}$.