STT 872, 867-868 Fall Preliminary Examination
Wednesday, August 25, 2021
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. (5 pts) Let $X_1, \cdots, X_n$ be iid from the Poisson distribution with

$$P\left(X_i = k\right) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Find the UMVU estimators of λ^k for any $k > 0$ and $e^{-\lambda}$.

2. Let X and Y be independent variables such that

$$X \sim \text{Binomial}(n, \theta) \text{ and } Y \sim \text{Binomial}(n, \theta^2),$$

where $\theta \in (0,1)$ is an unknown parameter and

$$P(X = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

(a) (3 pts) Find the minimal sufficient statistic for θ.
(b) (3 pts) Is the above minimal sufficient statistic complete?

3. Let $\theta = (\alpha, \lambda)$ and let P_θ denote the gamma distribution with shape parameter α and scale $1/\lambda$. So P_θ has density

$$p_\theta(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-x\lambda}}{\Gamma(\alpha)}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) (3 pts) Find the Fisher information matrix $I(\theta)$, expressed using the "psi" function $\psi = \Gamma'/\Gamma$ and its derivatives.
(b) (3 pts) What is the Cramer–Rao lower bound for the variance of an unbiased estimator of $\alpha + \lambda$?

4. Suppose $X \sim \text{Binomial}(\theta)$. Let the prior for θ be given by $\text{Beta}(\alpha, \beta)$ with density

$$\lambda(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad \theta \in (0,1)$$

(a) (3 pts) Under the squared error loss, find the Bayes estimator of θ^2.
(b) (3 pts) Find the mean and variance of the marginal density of X.
(c) (6 pts) Under squared error loss, is X/n minimax estimator of θ? Is it admissible?

5. (4 pts) Suppose $X_1, \cdots, X_n$ are i.i.d. from $N(0, \sigma^2)$. Find the uniformly most accurate confidence bound using the uniformly most powerful test for $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma > \sigma_0$. Note: The density of $N(\mu, \sigma^2)$ is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

6. Suppose we observe a single observation X from $N(\theta, \theta^2)$.
(a) (3 pts) Do the densities for X have monotone likelihood ratios?

(b) (4 pts) Let ϕ^* be the most powerful level α test of $H_0 : \theta = 1$ versus $H_1 : \theta = 2$. Explain how one would obtain this ϕ^*. Is ϕ^* also the most powerful level α test of $H_0 : \theta = 1$ versus $H_1 : \theta = 4$? Note: Explicit computation of constants in the test not needed.

7. Consider the full-rank linear regression model:

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

where $y_i \in \mathbb{R}, x_i \in \mathbb{R}^p, \beta \in \mathbb{R}^p$, and $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

(a) (3 pts) Suppose that part of the response data $\{y_i\}_{i=m+1}^n$ is missing. One data imputation strategy is to obtain OLS (denoted by $\hat{\beta}$) based on the observed data $\{(y_i, x_i)\}_{i=1}^m$, and then replace missing data y_i with the estimated value $\hat{y}_i = x_i'\hat{\beta}$ for $m + 1 \leq i \leq n$. Is the MSE based on the filled-up data $\{(y_i, x_i)_{1 \leq i \leq m}, (\hat{y}_i, x_i)_{m+1 \leq i \leq n}\}$ unbiased for σ^2? Provide your argument.

(b) (3 pts) Suppose that the response is censored: $y_i^* = y_i$ if $y_i > 0$ and $y_i^* = 0$ if $y_i \leq 0$. The observed data is $\{(y_i^*, x_i)\}_{i=1}^n$ instead of $\{(y_i, x_i)\}_{i=1}^n$. Consider the following two estimators for β: (1) OLS based on the whole data $\{(y_i^*, x_i)\}_{i=1}^n$; (2) OLS based on the data with positive response $\{(y_i^*, x_i)\}_{i:y_i^*>0}$. Prove whether they are unbiased or not.

(c) (4 pts) Suppose σ^2 is known. Consider two models: (1) linear model including the first q predictors (2) linear model having all the p predictors. How do you select between the two models using AIC? Assume $\frac{1}{n}\sum_{i=1}^n x_i x_i' \to \Sigma \succ 0$ as $n \to \infty$. Prove that as $n \to \infty$, when the first model is the true model, AIC has non-vanishing probability of selecting the wrong model; when the second model is correct, the probability of selecting the wrong model converges to zero.

8. Consider two full-rank linear regression models (with one-dimensional predictor):

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

where $y_i \in \mathbb{R}, z_i \in \mathbb{R}, x_i \in \mathbb{R}$. 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 a bivariate normal distribution $\mathcal{N}(0, \Sigma)$ with the unknown covariance matrix $\Sigma \in \mathbb{R}^{2\times 2}$.

(a) (4 pts) Show that the maximum likelihood estimators for β, γ, Σ are as follows:

$$(\hat{\beta}, \hat{\gamma}) = (X'X)^{-1}X'W, \quad \hat{\Sigma} = \frac{1}{n}W'(I - X(X'X)^{-1}X')W,$$

where $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^{n\times 1}, W = \begin{pmatrix} y_1 & z_1 \\ \vdots & \vdots \\ y_n & z_n \end{pmatrix} \in \mathbb{R}^{n\times 2}$.

(b) (6 pts) Construct a $100(1-\alpha)\%$ confidence region for (β, γ). Further develop simultaneous confidence intervals for all the linear functions of the form $\{t_1\beta + t_2\gamma; t_1, t_2 \in \mathbb{R}\}$ with joint coverage probability equal to $1 - \alpha$. Hint: You may want to use the following fact without proof. Let $T^2 = A'S^{-1}A$ where $A \in \mathbb{R}^p \sim \mathcal{N}(0, \Omega)$ and $S \in \mathbb{R}^{p\times p}$ is independently distributed as $\sum_{i=1}^m B_i B_i'$ with $B_1, \ldots, B_m \overset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$. Then $p^{-1}(m - p + 1)T^2 \sim F_{p,m-p+1}$.

9. Consider the linear mixed model,

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

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.

(a) (3 pts) Suppose $b_i \sim \mathcal{N}(0, \sigma^2 D), \epsilon_i \sim \mathcal{N}(0, \sigma^2 I_{n_i}), X_i = Z_iC_i, C_i = I_k \otimes a_i'$, where $a_i \in \mathbb{R}^q$. The parameters under this model are $\{\beta, \sigma^2, D\}$. Give conditions for $\{a_i, Z_i\}_{i=1}^N$ to ensure model identifiability.

(b) (3 pts) Suppose $k = 1$, and $n_i = 1, Z_i = 1, b_i \sim \mathcal{N}(0, \sigma_b^2), \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i = 1, 2, \ldots, N$, where the variances $\{\sigma_i^2\}_{i=1}^N$ are assumed known. The parameters under this model are $\{\beta, \sigma_b^2\}$. Derive the confidence region for β based on the Wald method.

(c) (4 pts) Suppose $k = 1, m = 1$, and $X_i = Z_i = (1, 1, \ldots, 1)', b_i \sim \mathcal{N}(0, \sigma_b^2), \epsilon_i \sim \mathcal{N}(0, I_{n_i})$ for $i = 1, 2, \ldots, N$. The parameters under this model are $\{\beta, \sigma_b^2\}$. Derive the EM algorithm for computing the maximum likelihood estimators of β, σ_b^2.