STT 872, 867-868 Fall Preliminary Examination
Tuesday, August 15, 2023
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 $X_1, ..., X_n$ be i.i.d. from the log normal distribution with the parameter $\lambda > 0$.

   Hint: The density is
$$f(x; \lambda) = \frac{1}{x\lambda\sqrt{2\pi}} \exp\left(-\frac{(\ln x)^2}{2\lambda^2}\right), x > 0.$$

   (1a) (3 pts) Find the complete and sufficient statistic for λ based on $X_1, ..., X_n$.

   (1b) (6 pts) Let $g(\lambda) = (\lambda^2 - 1)^2$. Formulate the two methods of finding UMVUE of $g(\lambda)$ and use one of them to find it.

   Hint: Think about the moments of normal distribution.

   (1c) (4 pts) Compute the Fisher infomation about λ containing in X_1. Give the lower bound for the variance of the unbiased estimator of λ.

   (1d) (3 pts) Consider the following inverse gamma prior for λ^2 given by
$$\pi(\lambda^2) = \frac{\beta^\alpha}{\Gamma(\alpha)(\lambda^2)^{\alpha+1}} \exp\left(-\frac{\beta}{\lambda^2}\right), \alpha > 1, \beta > 0.$$

   Find the Bayes estimate of λ^2 with the loss function $(\delta(X_1, \cdots, X_n) - \lambda^2)^2$.

   Hint: If X follows inverse gamma distribution with parameters α and β, then
$$E(X) = \frac{\beta}{\alpha - 1} \qquad \text{Var}(X) = \frac{\beta^2}{(\alpha-1)(\alpha-2)}$$

   (1e) (6 pts) Compute the Bayes risk of the estimator in (1d). Is the estimator admissible? (Do not forget the proof).

   (1f) (3 pts) Is the complete sufficient statistic in (1a) admissible? (Do not forget the justification).

   (1g) (5 pts) Construct the UMP unbiased test for $H : \lambda = 1$ vs $K : \lambda \neq 1$ at the significance level $\alpha \in (0, 1)$.

2. Let $X_1, ..., X_n$ be i.i.d. from a distribution with the density
$$f_a(x) = 0.5\sqrt{a}x^{-3/2} I(x > a),\ a > 0.$$

   (2a) (3 pts) Find the minimal sufficient statistic for a based on $X_1, ..., X_n$. (Do not forget the proof).

   (2b) (3 pts) Is the statistic in (2a) complete? (Do not forget the justification).

   (2c) (4 pts) Construct the UMP test for $H : 0 < a \leq a_0$ vs $K : a > a_0$ at the significance level $\alpha \in (0, 1)$.

3. Let $\boldsymbol{y}$ be $N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, β is a $k + 1 \times 1$ vector of regression coefficients.

   (3a) (5 pts) Suppose $k = 4$ and $n = 32$. Derive a test procedure to test the hypothesis $H_0 : \beta_1 = \beta_2 = 12\beta_3 = 12\beta_4$.

   (3b) (5 pts) For some reason, we transform the data by $\boldsymbol{z} = c\boldsymbol{y}$ and $\boldsymbol{W} = \boldsymbol{XK}$, where $\boldsymbol{K}$ is nonsingular. Can you use the same test statistic that you developed in part (3a). Show your work.

   (3c) (5 pts) Suppose $\boldsymbol{C}$ is a $q \times (k + 1)$ matrix of rank $q \leq k + 1$ and we are interested in testing $H_0 : \boldsymbol{C\beta} = 0$. Derive an estimator for $\boldsymbol{\beta}$ under this reduced model. Find the expected values and covariance of your estimator.

4. Consider the social network regression model:

$$y_{ij} = x_{ij}^T\beta + a_i + b_j + \epsilon_{ij}, \quad 1 \le i \ne j \le n,$$

where $y_{ij} \in \mathbb{R}$ denotes the value of the relationship in the direction from node i to node j, $x_{ij} \in \mathbb{R}^p$ is the vector of fixed covariates for the ordered pair (i, j), and (a_i, b_i) represent nodal heterogeneity of i as "sender" and "receiver" respectively. To quantify the sender-receiver correlations, it is assumed that $\{(a_i, b_i)\}_{i=1}^n \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$, $\{(\epsilon_{ij}, \epsilon_{ji})\}_{1\le i<j\le n} \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Omega)$, with $\Sigma = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$, $\Omega = \begin{pmatrix} \sigma^2 & \tau \\ \tau & \sigma^2 \end{pmatrix}$, and ϵ_{ij}'s are independent from the a_i's, b_j's. The parameters in the model are β, Σ, Ω. The observations are $\{(y_{ij}, x_{ij})\}_{1\le i\ne j\le n}$.

(4a) (5 pts) Provide conditions to ensure model identifiability.

(4b) (5 pts) Consider a simplified setting where $x_{ij} = 1, 1 \le i \ne j \le n$, and σ_{ab}, Ω are known. The unknown parameters in the current model are $\beta, \sigma_a^2, \sigma_b^2$. Use the following two sums of squares to construct unbiased estimators of σ_a^2 and σ_b^2:

$$\mathbf{S}_r = \sum_{i=1}^n (Y_{i\cdot} - Y_{\cdot\cdot})^2, \quad \mathbf{S}_c = \sum_{j=1}^n (Y_{\cdot j} - Y_{\cdot\cdot})^2,$$

where $Y_{i\cdot} = \frac{1}{n-1}\sum_{k\ne i} y_{ik}, Y_{\cdot j} = \frac{1}{n-1}\sum_{k\ne j} y_{kj}, Y_{\cdot\cdot} = \frac{1}{n(n-1)}\sum_{i\ne j} y_{ij}$.

5. In the 2010 General Social Survey, subjects were asked who they voted for in the 2004 and 2008 Presidential elections. Let (y_{i1}, y_{i2}) denote the pair of observations for subject i in the sample, where $y_{i1} \in \{0, 1\}$ denotes the outcome in 2004 and $y_{i2} \in \{0, 1\}$ in 2008. To study such matched-pair data, consider the following model,

$$\mathbb{P}(y_{i1} = 1) = \frac{e^{\alpha_i}}{1 + e^{\alpha_i}}, \quad \mathbb{P}(y_{i2} = 1) = \frac{e^{\alpha_i+\beta}}{1 + e^{\alpha_i+\beta}}, \ \ 1 \le i \le n. \tag{1}$$

(5a) (5 pts) For model (1), assume independence of responses for different subjects and for the two observations on the same subject. Provide an interpretation of the parameters α_i and β. Show that the model is a generalized linear model.

(5b) (5 pts) [continued from (5a)] A popular approach to estimate β is based on conditional likelihood instead of likelihood. Obtain the conditional distribution $\mathbb{P}((y_{i1}, y_{i2}) \mid y_{i1} + y_{i2} = 1)$ and use it to derive the closed-form expression of maximum conditional likelihood estimator for β.

(5c) (5 pts) For model (1), assume independence of responses for different subjects, and $\alpha_1, \ldots, \alpha_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$[1]. The parameters in the current model are β, μ, σ^2. First show that the random effects induce positive correlations: $\mathrm{Cov}(y_{i1}, y_{i2}) \ge 0$. Then describe the EM algorithm for computing the MLEs.

[1]Conditionally on the random effect α_i, treat y_{i1} and y_{i2} as independent and they follow distributions in (1).