STT 872, 867-868 Spring Preliminary Examination
Wednesday, August 20, 2025
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 $X_1, \dots, X_n$ be iid rvs from the density

$$f_\theta(x) = 2\theta x e^{-\theta x^2}, x > 0, \theta > 0.$$

   (1a) (3 pts) Find the MLE of $1/\theta$.

   (1b) (3 pts) Let $g(\theta) = 1/\theta$. Let the loss function be $(\theta\delta(X_1, \dots, X_n) - 1)^2$. Use the exponential prior for θ, that is the density is $\pi(\theta) = \tau e^{-\tau\theta}, \theta > 0, \tau > 0$. For $n > 1$ find the Bayes estimator of $g(\theta)$.

   Hint: Gamma distribution has density

$$f(y) = \frac{y^{m-1} e^{-\frac{y}{b}}}{b^m \Gamma(m)}, b > 0, y > 0 \tag{1}$$

   and moments $\mathbb{E}(Y^r) = (\Gamma(m+r)/\Gamma(m))b^r, r > -m$.

   (1c) (4 pts) Let $\tau \to 0$. Obtain a minimax estimator of $g(\theta)$ under the loss function in (1b). Explain why it is minimax.

   (1d) (4 pts) In full details, derive a UMP unbiased test of size $\alpha \in (0, 1)$ for

$$H_0 : \theta = 1 \quad \text{vs} \quad H_1 : \theta \neq 1$$

2. Let X take on values $1, -1, 0$ with probabilities p, $2p$, $1 - 3p$, respectively.

   (2a) (3 pts) Do we have MLR in the statistic $T(X) = |X|$?

   (2b) (4 pts) Is $T(X) = |X|$ admissible for its expectation under the squared error loss?

   (2c) (3 pts) In full details, construct a UMP test at level $\alpha \in (0, 1)$ for

$$H_0 : p = 0.2 \quad \text{vs} \quad H_1 : p > 0.2$$

3. Let $X_1, \cdots, X_n$ be i.i.d. from the log normal distribution with the parameter $\lambda > 0$. The density is

$$f(x; \lambda) = \frac{\exp\left\{-0.5(\ln x)^2/\lambda^2\right\}}{x\lambda\sqrt{2\pi}}, x > 0.$$

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

   (3b) (5 pts) Let $g(\lambda) = \lambda^3$. Formulate the two methods of finding UMVUE of a parameter and use one of them to find it for $g(\lambda)$.

   Hint: To get moments of chi-square distribution with n degrees of freedom ($\chi^2_{(n)}$) you may use Gamma distribution in (1) with $m = n/2$ and $b = 2$.

4. Let $X_1, \cdots, X_n$ be iid rvs from exponential distribution with density

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

   (4a) (4 pts) Find the MRE estimator of ξ under the loss $|\delta(X_1, \cdots, X_n) - \xi|$.

   (4b) (4 pts) Construct the uniformly most accurate 95% upper confidence bound for ξ.

   Hint: Think of the corresponding uniformly most powerful test.

5. The linear models considered in this problem are all of full column rank. The following three questions are separate and not related to each other.

   (5a) (5 pts) Consider two simple linear regression models,

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

   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 $\mathcal{N}(0, \sigma^2)$. Construct simultaneous confidence intervals for all the linear functions $\alpha_k + \beta_k x_k, \forall x_k \in \mathbb{R}, k = 1, 2$, with joint coverage probability greater than or equal to $1 - \alpha$.

   (5b) (5 pts) To test if the first observation is an outlier, instead of $y_i = x_i'\beta + \epsilon_i, i = 1, 2, \ldots, n$, we consider the following outlier shift model,

   $$y_1 = x_1'\beta + \gamma + \epsilon_1 \text{ and } y_i = x_i'\beta + \epsilon_i, i = 2, \ldots, n,$$

   where $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ and $\beta \in \mathbb{R}^p, \gamma \in \mathbb{R}$. Consider the hypothesis testing problem $H_0 : \gamma = 0, H_1 : \gamma \neq 0$. We may view the first observation as an outlier, if we reject H_0. Provide the test statistic and α-level rejection region.

   (5c) (5 pts) For a pressure gauge, suppose that the pressure and the corresponding gauge reading follow a linear model,

   $$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n,$$

   where $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, and $y_i \in \mathbb{R}$ is the gauge reading for the pressure $x_i \in \mathbb{R}$. Now suppose we have m repeated gauge readings $\{y_{0j}\}_{j=1}^m$ (with independent errors) for an unknown pressure x_0. Use the data $\{(x_i, y_i)\}_{i=1}^n \cup \{y_{0j}\}_{j=1}^m$ to construct a $1 - \alpha$ level confidence region for x_0.

6. Consider the model,

   $$y_i = f(x_i) + \epsilon_i, \ i = 1, 2, \ldots, n,$$

   where $f(\cdot)$ can be a nonlinear function, $x_i = \frac{i}{n}$, and $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

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

   $$\hat{f}_m(x) = \sum_{j=1}^m \hat{\beta}_j \phi_j(x), \ \hat{\beta}_j = \frac{1}{n} \sum_{i=1}^n \phi_j(x_i) y_i,$$

   where $\{\phi_j(x)\}_{j=1}^\infty$ are the trigonometric basis. Namely, the method predicts $f(x)$ as a linear combination of the first m trigonometric basis functions. The explicit expression of trigonometric basis is not needed here, but the following property is important:

   $$\frac{1}{n} \sum_{s=1}^n \phi_j(s/n)\phi_k(s/n) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k, \end{cases}$$

   for all $1 \leq j, k \leq n - 1$. First, compute the expected in-sample prediction error of the projection method, and describe how the bias and variance are affected as m varies (assume $m \leq n - 1$). Then, give an unbiased estimator for the expected in-sample prediction error.

(6b) (5 pts) Suppose we would like to select among two model candidates: (1) $f(x_i) = \sum_{j=1}^{5} \beta_j \phi_j(x_i)$; (2) $f(x_i) = \sum_{j=1}^{6} \beta_j \phi_j(x_i)$. Prove that when the first model is the true model, Akaike Information Criterion (AIC) has non-vanishing probability of selecting the wrong model, no matter how large the sample n is.

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

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

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

$$\alpha_i = c_0 + c_1 \cdot d_i + \xi_i, \quad i = 1, \ldots, N, \tag{3}$$

where $d_i \in \mathbb{R}$ denotes the daily calcium supplement of the ith teenager, and all the ξ_i's are independently sampled from $\mathcal{N}(0, \tau^2)$. The observed data consists of $\{(y_{i1}, \ldots, y_{in}, t_{i1}, \ldots, t_{in}, d_i)\}_{i=1}^{N}$, and the unknown parameters are $\{c_0, c_1, \tau^2, \sigma^2\}$.

(7a) (5 pts) Plug (3) into (2) and show that the resulting model is an example of linear mixed models. Provide conditions to ensure its identifiability.

(7b) (5 pts) Assume the condition $t_{ij} = t_j$ for all $i = 1, \ldots, N, j = 1, \ldots, n$. Derive the closed-form solution for the MLE of (c_0, c_1). Note that (τ^2, σ^2) are also unknown parameters. (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) (5 pts) Suppose the condition in (7b) is not necessarily satisfied. Derive the EM algorithm for computing the MLE of $\{c_0, c_1, \tau^2, \sigma^2\}$.