PhD Qualifying Examination
Theory Part
Summer 2024

1. (10 points) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from a normal distribution $N(\theta_1, \theta_2)$, where $\theta_1, \theta_2 \in (0, \infty)$. Note that the regularity conditions for the consistency and asymptotic normality (CAN) of the maximum likelihood estimator (MLE) are satisfied by this statistical model and $E(X_i^2) < \infty$.

(a). (2 points) **Find** the joint sufficient statistics of (θ_1, θ_2).

(b). (4 points) Suppose $\theta_1 = \theta_2 = \theta$, where $\theta \in (0, \infty)$. Under the regularity conditions, we know that MLE is consistent, but we can prove its consistency using the theorems we learned. **Find** the MLE of θ (denoted by $\widehat{\theta}_n$) and **show that** $\widehat{\theta}_n$ is a (weakly) consistent estimator of θ.

(c). (4 points) Suppose $\theta_1 = \theta_2 = \theta$, where $\theta \in (0, \infty)$. **Show that** $g\left(n^{\frac{1}{3}}(\bar{X}_n - \theta)\right) \longrightarrow_p \dfrac{1}{2}$ as $n \to \infty$, where the function $g(a) = \dfrac{1}{1 + e^{-a}}$, the sample mean $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$, and $\longrightarrow_p$ denotes the convergence in probability as $n \to \infty$.

2. (10 points) Consider the sequence of random variables $X_1, X_2, X_3, \ldots$ defined by

$$X_n = \begin{cases} n & \text{if } \ Z \geq n, \\ 0 & \text{if } \ Z < n, \end{cases}$$

where Z is a random variable having an Exponential distribution with parameter 1 and thus the PDF of Z is $f(z) = e^{-z}$ for $z \geq 0$. Note that $e = 2.718282...$ and n indicates a sample size so $n \geq 1$.

(a). (4 points) **Show** whether X_n converges to 0 almost surely ($\longrightarrow_{a.s.}$ as $n \to \infty$) or not. **Do NOT** use the results of (b) or (c).

(b). (3 points) **Show** whether X_n converges to 0 in probability ($\longrightarrow_{p}$ as $n \to \infty$) or not. **Do NOT** use the results of (a) or (c).

(c). (3 points) **Show** whether X_n converges to 0 in 2nd moment ($\longrightarrow_{L^2}$ as $n \to \infty$) or not. **Do NOT** use the results of (a) or (b).

3. (10 points) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from $N(\mu, 1)$, where $\mu \in R$ is unknown. Let $\theta = Pr(X_1 \leq c)$, where c is a known constant.

(a). (5 points) Find the MLE of θ (denoted by $\widehat{\theta}$) and the asymptotic variance of $\sqrt{n}(\widehat{\theta} - \theta)$. No need to verify regularity conditions.

(b). (5 points) Find the asymptotic relative efficiency of the estimator $\tilde{\theta} = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq c\}$ with respect to the MLE $\widehat{\theta}$.

4. (10 points) Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables from some continuous distribution with nonzero density at its population median, that is, $f(\theta_1) > 0$ where $f(t)$ is the PDF, $\theta_1 = F^{-1}(0.5)$, and $F(t)$ is the CDF.
Let $\widehat{\theta}_1$ denote the sample median:

$$\widehat{\theta}_1 = F_n^{-1}(0.5) = inf\{\eta : F_n(\eta) \geq 0.5\}, \text{ where } F_n(t) \text{ is the empirical distribution function.}$$

Let $\widehat{\theta}_2$ denote an estimator of $\theta_2 = E[(Y_1 - \theta_1)^2]$:

$$\widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\theta}_1)^2.$$

(a). (4 points) Consider $\widehat{\theta}_2$ as a partial M-estimator and identify the estimating functions $\psi_1(Y_i, \theta)$ and $\psi_2(Y_i, \theta)$ such that $\theta = (\theta_1, \theta_2)^T$ are the corresponding parameters of interest. Show that these estimating functions satisfy (i) and (ii):
(i) $E_\theta[\psi_1(Y_1, \theta)] = 0$ and $E_\theta[\psi_2(Y_1, \theta)] = 0$.
(ii) $\sum_{i=1}^n \psi_1(Y_i, \widehat{\theta}) = o_p(\sqrt{n})$ and $\sum_{i=1}^n \psi_2(Y_i, \widehat{\theta}) = o_p(\sqrt{n})$, where $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)^T$.
Hint: the sample median satisfies $|F_n(\widehat{\theta}_1) - 0.5| < \frac{1}{n}$.

(b). (2 points) Show that the following equation,

$$\begin{cases} E[\psi_1(Y_1, \theta)] = 0 \\ E[\psi_2(Y_1, \theta)] = 0 \end{cases}$$

has a unique solution in the neighborhood of the truth $\theta_0 = (\theta_{01}, \theta_{02})^T$. No need to verify other regularity conditions.

(c). (2 points) Derive the asymptotic variance-covariance matrix of $\sqrt{n}(\widehat{\theta} - \theta_0)$. Provide the matrices $A(\theta_0)$ and $B(\theta_0)$ needed for the calculation. In the calculation, you can assume that terms such as $E[(Y_1 - \theta_{01})^j]$ are known, $j = 1, 2, 3, 4, \ldots$.

(d). (2 points) Derive the influence function for $\widehat{\theta}_2$.
Hint: the influence function for the sample median $\widehat{\theta}_1$ is:

$$IC(x, F, \theta_1) = \frac{0.5 - I\{F(x) \leq 0.5)\}}{f(\theta_1)}.$$

5. (10 points) Consider a vaccine clinical trial to identify the antibody level after applying one of three vaccines (V0, V1, or V2). After two weeks of treatment, we measure the antibody level as the response to the vaccine. We suspect that females (G1) and males (G2) may have different responses. So we recruit 300 males and 300 females and randomly assign 100 males and 100 females into V0, V1 and V2 groups. Let $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{20}, \gamma_{21}, \gamma_{22})^T$. We fit the following model.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad i = 1, 2, \quad j = 0, 1, 2, \quad k = 1, ..., 100,$$

where $\epsilon_{ijk} \sim^{iid} N(0, \sigma^2)$.

Note that the least square estimator of $E(y_{ijk})$ is $\bar{y}_{ij.}$.

(a). (3 points) Show that $\theta = \alpha_1 - \alpha_2 + \gamma_{1.} - \gamma_{2.}$ is estimable, where $\gamma_{i.} = \sum_j \gamma_{ij}/3$.

(b). (4 points) An investigator plans to fit the following reduced model

$$y_{ijk} = \mu + \beta_j + \epsilon_{ijk}, \quad i = 1, 2, \quad j = 0, 1, 2, \quad k = 1, ..., 100,$$

where $\epsilon_{ijk} \sim^{iid} N(0, \sigma^2)$. Please write down the corresponding null hypothesis and conduct a test at level of 0.05.

(c). (3 points) Construct the 95% simultaneous confidence intervals for $\theta_{ij} = \alpha_i + \beta_j + \gamma_{ij}, \quad i = 1, 2, \quad j = 0, 1, 2$.

6. (10 points)

(a). (5 points) Consider the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$. Please show that if $\theta = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable, then $\lambda \in \mathcal{C}(\boldsymbol{X}^T)$.

(b). (5 points) Let $\boldsymbol{A}$ be an idempotent matrix. Show that $(\boldsymbol{A} + \boldsymbol{A}^T)/2$ is idempotent if and only if $\boldsymbol{A}$ is symmetric.

7. (10 points) Consider a matched design to estimate treatment effect. We fit the following generalized Linear Mixed Model

$$y_{ij}|u_i \sim \text{Poisson}(\lambda_{ij})$$

$$\log(\lambda_{ij}) = \mu + u_i + x_j\beta, \quad i = 1, \ldots, m, \quad j = 1, 2$$

where $u_i \sim_{\text{iid}} \mathcal{N}(0, \sigma^2)$, $x_j = j - 1$, so x_j is 0 or 1.

We are interested in estimating β. There is no closed form solution if we use the likelihood method, so we apply the conditional likelihood method. The log conditional likelihood is:

$$\ell = \sum_i \sum_j y_{ij}(\mu + u_i) + \sum_i \sum_j y_{ij} x_j \beta - \sum_i \sum_j [\log(y_{ij}!) + e^{\mu + u_i + x_j\beta}]$$

(a). (3 points) Let $Y_i \sim Poisson(\lambda_i), i = 1, 2$, and Y_1 and Y_2 be independent. Show that

$$Y_1 | Y_1 + Y_2 = y \sim \text{Binomial}(y, \lambda_1/(\lambda_1 + \lambda_2)).$$

(b). (7 points) Estimate β using the conditional likelihood method.