BASIC PHD WRITTEN EXAMINATION

THEORY, SECTION 2

(9:00 AM–1:00 PM, July 29, 2021)

INSTRUCTIONS:

(a) This is a **CLOSED-BOOK** examination.

(b) The time limit for this examination is four hours.

(c) Answer both questions that follow.

(d) Put the answers to different questions on separate sets of paper.

(e) Put your exam code, **NOT YOUR NAME**, on each page. The same code is used for Section 1 and Section 2 of the PhD Theory Exam. Please keep the code confidential and do not share this information with any students or faculty. Sharing your code with either students or faculty is viewed as a violation of the UNC honor code.

(f) Return the examination with a signed statement of the UNC honor pledge, separately from your answers. The pledge statement is given on the last page of the exam handout.

(g) In the questions to follow, you are required to answer only what is asked, and not to tell all you know about the topics involved.

1. (25 points) Consider two linear models

$$Y = X_1\gamma_1 + X_2\gamma_2 + X_3\gamma_3 + \epsilon_1,$$

and

$$Y = X_1\beta_1 + X_3\beta_3 + \epsilon_2,$$

where Y is $n \times 1$, X_3 is a single column but X_1 and X_2 may have multiple columns, $E(\epsilon_k) = 0$, $\mathrm{Cov}(\epsilon_k) = \sigma^2 I$, ϵ_k are *iid* for $k = 1, 2$, and all parameters are unknown. Further, let J_n denote the $n \times 1$ vector of ones, and assume $J_n \in C(X_1)$, where $C(X_1)$ denotes the column space of X_1. Further, let M_1 denote the orthogonal projection operator onto $C(X_1)$, and M_{12} denotes the orthogonal projection operator onto $C(X_1, X_2)$.

(a) (6 points) Researchers are interested in knowing when the least squares estimates have $\mathrm{sign}(\hat{\beta}_3) \neq \mathrm{sign}(\hat{\gamma}_3)$. Show that $\mathrm{sign}(\hat{\beta}_3) \neq \mathrm{sign}(\hat{\gamma}_3)$ if and only if $\mathrm{sign}[X_3'(I - M_1)Y] \neq \mathrm{sign}[X_3'(I - M_{12})Y]$.

(b) (3 points)]Assuming $\hat{\beta}_3$ is positive, show that $\mathrm{sign}(\hat{\beta}_3) \neq \mathrm{sign}(\hat{\gamma}_3)$ if and only if $X_3'(I - M_1)Y < X_3'M_3Y$.

(c) Let M_3 denote the orthogonal projection operator onto the orthogonal complement of X_1 with respect to $C(X_1, X_2)$.

   (i) (5 points) Express M_3 *only* in terms of M_1 and M_{12}.

   (ii) (3 points) Define $r = \frac{X_3'(I-M_1)M_3(I-M_1)Y}{\sqrt{X_3'(I-M_1)M_3(I-M_1)X_3}\sqrt{Y'(I-M_1)M_3(I-M_1)Y}}$. Show that $r = \frac{X_3'M_3Y}{\sqrt{X_3'M_3X_3}\sqrt{Y'M_3Y}}$.

(d) Now consider the more general case in which $Y \sim N_n(\mu, \sigma^2 I_n)$, where I_n is the $n \times n$ identity matrix and Y is $n \times 1$. Under the null model, $\mu = J_n\alpha$, and under the alternative model, $\mu = J_n\alpha + X\beta$, where X is $n \times p$ of rank $p < n$ that has been centered so that $X'J_n = 0_p$. Furthermore, if a random variable, $W \sim \chi^2(a)$, then the density of W is given by $f(w) = \frac{w^{a/2-1}\exp(-w/2)}{\Gamma(a/2)2^{a/2}}, \quad w > 0$, and if $Z \sim \mathrm{Beta}(b, c)$, then $f(z) = \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} z^{b-1}(1 - z)^{c-1}, \quad 0 < z < 1$.

   (i) (5 points) Define $R^2 = \frac{Y'X(X'X)^{-1}X'Y}{\|Y - \frac{1}{n}J_nJ_n'Y\|^2}$, where $\|a\|^2 = a'a$ for any vector a. Show that under the null model, R^2 has a beta distribution and find its expected value.

   (ii) (3 points) Under the setup of part (i), if p is increasing with n such that $\frac{p}{n-1-p} = \lambda$, with $0 < \lambda < \infty$, find the limit of R^2 as both $n, p \to \infty$ under the null model.

Points: (a) 6; (b) 3; (c) (i) 5; (ii) 3; (d) (i) 5; (ii) 3.

2. The random variable Y_1 is distributed as binomial(n_1, θ) and independently, Y_2 is distributed as binomial$(n_2, \theta\lambda)$. Here, $\theta, \theta\lambda \in (0, 1)$.

Define any new symbols you introduce. For full credit: Simplify answers to the fullest extent and show all your work, not just the final answer.

(a) (2 points) Express the above model in a way that shows it clearly to be a generalized linear regression model. Identify the model components.

(b) (5 points) Suppose that $n_1 = n_2 = n$ and $\theta = \lambda$. Under these conditions, consider Y_1/n and $\sqrt{Y_2/n}$ as two estimators of θ. Derive their asymptotic variances (as $n \to \infty$) and investigate under what condition one is smaller than the other.

(c) (5 points) Under the same conditions as (b), develop a linear combination of the two estimators that would be a consistent estimator of θ with the smallest asymptotic variance among all such linear combinations. Find the asymptotic variance of that estimator. If forming the linear combination requires knowing θ, suggest a way to implement the estimator.

(d) (5 points) Obtain the Fisher information matrix for the parameter vector $(\theta, \lambda)^\top$. Note: Allow $n_1 \neq n_2$ and $\theta \neq \lambda$.

(e) (3 points) Suppose that $n_1 = n_2 = n$. Derive the score test statistic for the hypothesis $H_0 : \theta = \lambda$. Use the notation $\hat{\theta}_0$ for the MLE under H_0. You do not need to derive an explicit form for $\hat{\theta}_0$.

(f) (5 points) Suppose that $n_1 = n_2 = n$ and $\theta = \lambda$. Consider the residuals $R_1 = Y_1 - n\theta$ and $R_2 = Y_2 - n\theta^2$. Find the optimal linear combination (estimating function, quasi-score) of R_1 and R_2 for estimating θ. Denote the resulting estimator by $\tilde{\theta}_n$. Calculate the asymptotic (as $n \to \infty$) variance of $\sqrt{n}(\tilde{\theta}_n - \theta)$.

Points: (a) 2; (b) 5; (c) 5; (d) 5; (e) 3; (f) 5.

2021 PhD Theory Exam, Section 2

Statement of the UNC honor pledge:

"In recognition of and in the spirit of the honor code, I certify that I have neither given nor received aid on this examination and that I will report all Honor Code violations observed by me."

(Signed) ______________________________
NAME

(Printed) ______________________________
NAME