

BASIC PHD WRITTEN EXAMINATION

THEORY, SECTION 2

(9:00 AM–1:00 PM, July 31, 2020)

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 the linear model

$$Y = X\beta + Z\gamma + \epsilon, \quad (1)$$

where Y is $n \times 1$, X is $n \times p$ of rank $r \leq p$, Z is $n \times q$ of rank $s \leq q$, β is $p \times 1$, and γ is $q \times 1$. Further, assume that ϵ is $n \times 1$, where $E(\epsilon) = 0$ and $\text{Cov}(\epsilon) = \sigma^2 I$, where I denotes the $n \times n$ identity matrix and $(\beta, \gamma, \sigma^2)$ are all unknown parameters. Furthermore, let M denote the orthogonal projection operator onto $C(X)$, where $C(X)$ denotes the column space of X .

(a) Show that $C(X, Z) = C(X, (I - M)Z)$.

(b) Let SSE denote the error sum of squares from the model in (1). Show that

$$\text{SSE} = Y'(I - M)Y - Y'(I - M)Z[Z'(I - M)Z]^{-1}Z'(I - M)Y,$$

where G^{-} denotes the generalized inverse of an arbitrary matrix G .

(c) Suppose $Z'(I - M)Z$ is non-singular. Show that the least squares estimate of γ is given by

$$\hat{\gamma} = [Z'(I - M)Z]^{-1}Z'(I - M)Y.$$

(d) Show that

(i) $\lambda'\gamma$ is estimable if and only if $\lambda' = \rho'(I - M)Z$ for some vector ρ in R^n , where λ is $q \times 1$.

(ii) if $Z'(I - M)Z$ is singular, neither γ nor $X\beta$ are estimable.

(e) Suppose γ is estimable, and let $(\hat{\beta}, \hat{\gamma})$ denote the least squares estimates of (β, γ) , and assume $Z'(I - M)Z$ is non-singular. Derive the simplest possible expression for the $(n + q) \times (n + q)$ covariance matrix of $(X\hat{\beta}, \hat{\gamma})$.

(f) Now suppose $\epsilon \sim N_n(0, \sigma^2 V)$ for the model in (1), where V is positive definite and known. Further, suppose that γ is estimable, $X'V^{-1}Z = 0$, and we wish to test $H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$. Under H_0 , explicitly find projection matrices A_1 and A_2 in the simplest possible form to show that the statistic

$$B = \frac{Y'A_1Y}{Y'A_2Y}$$

has a beta distribution. Also find the distribution of B under the alternative hypothesis.

Points: (a) 3; (b) 4; (c) 4; (d) 4; (e) 5; (f) 5.

2. (25 points) Suppose that Y_1, \dots, Y_n are independent random variables and each Y_i is distributed as normal with mean $\mu > 0$ and variance μ . Define $\bar{Y}_n = (1/n) \sum_{i=1}^n Y_i$ and let $\hat{\mu}$ denote the maximum-likelihood estimator of μ . Note: Asymptotics in this problem are in terms of $n \rightarrow \infty$. In what follows, obtain explicit expressions and simplify them as much as possible. Define any new notation you introduce.
- (a) Show that the distribution of Y_1 belong to the exponential family of distributions. Identify the canonical parameter, the canonical link function, and the cumulant function (a function of the canonical parameter).
 - (b) Find the asymptotic variance of $\hat{\mu}$, that is, the limiting variance of $\sqrt{n}(\hat{\mu} - \mu)$ as $n \rightarrow \infty$. Sketch the asymptotic relative efficiency of \bar{Y} relative to $\hat{\mu}$. Hint: Obtaining an explicit form for $\hat{\mu}$ is not necessary.
 - (c) Now suppose Y_i is normally distributed with mean $\mu_i > 0$ and variance μ_i , and $\log(\mu_i + \mu_i^2) = x_i^\top \beta = \eta_i$, where x_i is a $p \times 1$ vector of covariates, $i = 1, \dots, n$, and β is a vector of unknown parameters. Obtain expressions for the score function and the expected (Fisher) information matrix.
 - (d) Develop a quasi-score equation that is a linear combination of the residuals $Y_i - \mu_i$ for estimating β in the regression model in part (c). Let $\tilde{\beta}$ denote the resulting estimator. Obtain an expression for the asymptotic covariance matrix of $\sqrt{n}(\tilde{\beta} - \beta)$. Develop an expression for the “sandwich” (robust, empirical) covariance (covariance of $\tilde{\beta}$) estimator.
 - (e) In the context of the regression model part in (c), suppose that $x_i^\top = (1, 0)$ for $i = 1, \dots, n_0$ and $x_i^\top = (1, 1)$ for $i = 1 + n_0, \dots, n$, and $n_0/n \rightarrow \gamma$ as $n \rightarrow \infty$, where $\gamma \in (0, 1)$ is a known constant. The parameter vector is $(\beta_1, \beta_2)^\top$. Derive the score test for testing the hypothesis $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$.

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

2020 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