

BASIC PHD WRITTEN EXAMINATION

THEORY, SECTION 2

(9:00 AM–1:00 PM, July 26, 2018)

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

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon, \quad (1)$$

where Y is $n \times 1$, X_1 is $n \times p_1$ of rank p_1 , β_1 is $p_1 \times 1$, X_2 is $n \times p_2$ of rank p_2 , (X_1, X_2) is full rank, β_2 is $p_2 \times 1$, and ϵ is $n \times 1$, with $E(\epsilon) = 0$ and $\text{Cov}(\epsilon) = \sigma^2 I$ where σ^2 is unknown. Suppose in reality that $\beta_2 = 0$, and thus the model used by the data analyst is an overfitted model given by (1) but the true model is

$$Y = X_1\beta_1 + \epsilon. \quad (2)$$

Let $\hat{\sigma}_{\text{overfit}}^2$ denote the usual estimator of σ^2 based on the overfitted model, where (β_1, β_2) are unknown.

- (a) (3 points) Derive the expectation of $\hat{\sigma}_{\text{overfit}}^2$ assuming the model in (2) is true. Express your answer in the simplest possible form.
- (b) (5 points) Assuming all relevant inverses exist, Derive an explicit expression for the least squares estimate of β_1 based on the model in (1), and derive the expectation of the least squares estimate of β_1 from the model in (1) assuming the model in (2) is true. Express your answers in the simplest possible form.
- (c) (6 points) Assume normality of ϵ and the model in (2) is true. Derive an exact $(1 - \alpha) \times 100\%$ confidence interval for σ^2 based on the model in (1) and give explicit conditions as to when the expected length of the confidence interval for σ^2 based on the model in (2) is smaller than that of the model in (1).
- (d) Now consider the special case in which $\gamma_1, \gamma_2, \gamma_3$ are interior angles of a triangle so that $\gamma_1 + \gamma_2 + \gamma_3 = 180$ degrees. Suppose we have available measurements Y_1, Y_2, Y_3 of $\gamma_1, \gamma_2, \gamma_3$, respectively. We assume that $Y_i \sim N(\gamma_i, \sigma^2)$, $i = 1, 2, 3$, σ^2 is unknown and the Y_i 's are independent.
 - (i) (7 points) Derive the F test for testing the null hypothesis that the triangle is equilateral, and state the distribution of the F statistic under the null and alternative hypotheses. Express your test statistic in the simplest possible form.
 - (ii) (4 points) Derive a 95% joint confidence region for (γ_1, γ_2) .

2. (25 points) Suppose that the pair of random variables (X, Y) has the form $X = U + T_1$ and $Y = U + T_2$, where U, T_1 and T_2 have independent Poisson distributions with means $E[U] = \psi, E[T_1] = \lambda_1, E[T_2] = \lambda_2$. Note that the marginal distributions of X and Y are Poisson and that X and Y are independent if and only if $\psi = 0$. The objective of this problem is to develop large-sample tests of independence of X and Y .

- (a) (4 points) Show that the distribution of (X, Y) is given by

$$P(X = x, Y = y) = e^{-(\psi + \lambda_1 + \lambda_2)} \lambda_1^x \lambda_2^y \sum_{u=0}^{\min(x, y)} \left(\frac{\psi}{\lambda_1 \lambda_2} \right)^u \frac{1}{u!(x-u)!(y-u)!},$$

for nonnegative integers x and y .

- (b) (7 points) The hypothesis $H_0 : \psi = 0$ is of interest. Obtain an explicit expression for the log-likelihood, $l(\psi, \lambda_1, \lambda_2; x, y)$, based on a single pair (x, y) . Then show that the score vector evaluated at $\psi = 0$ is

$$\begin{pmatrix} \frac{xy}{\lambda_1 \lambda_2} - 1 \\ \frac{x}{\lambda_1} - 1 \\ \frac{y}{\lambda_2} - 1 \end{pmatrix}.$$

Further, obtain the expected information matrix (3×3) , also evaluated at $\psi = 0$. Important: The general form of the score vector and expected information matrix may be complicated. Only the form evaluated at $\psi = 0$ is required in this problem.

- (c) (5 points) Now consider testing the hypothesis $H_0 : \psi = 0$ using independent pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ from the same distribution as (X, Y) above. Derive an explicit expression for the score test statistic, and identify its asymptotic (as $n \rightarrow \infty$) distribution under H_0 .
- (d) Since $\text{cov}(X, Y) = \psi$ it is natural to develop a test statistic based on the “sample covariance”. The sample covariance is S/n where $S := \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$, $\bar{X} = \sum_{i=1}^n X_i/n$ and $\bar{Y} = \sum_{i=1}^n Y_i/n$. Since (\bar{X}, \bar{Y}) is a minimal sufficient statistic under H_0 , a test can be based on the conditional distribution of S given (\bar{X}, \bar{Y}) under H_0 .
- (i) (7points) Compute the **exact** conditional variance of S given $\bar{X} = \bar{x}$ and $\bar{Y} = \bar{y}$ under H_0 .
- (ii) (2 points) By standardizing S , obtain a test statistic that is asymptotically standard normal under H_0 .

2018 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