BASIC PHD WRITTEN EXAMINATION IN BIOSTATISTICS

THEORY, SECTION 1

(9:00 AM–1:00 PM, July 25, 2016)

INSTRUCTIONS:

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

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

(c) Answer any TWO (2) (BUT ONLY TWO) of the THREE (3) 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 will be 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) Let $X_1, ..., X_n$ be i.i.d from the following distribution

$$\begin{cases} 0 & \text{with probability } p, \\ \text{Uniform}[0,\theta] & \text{with probability } 1-p. \end{cases}$$

First, we assume that p is a known constant in $(0,1)$ and that $\theta > 0$ is the only parameter of interest.

(a) (5 points) Based on only one observation X_1, find all the unbiased estimators for θ and calculate their variances. Does UMVUE exist for θ? Justify your answer.

(b) (3 points) Based on n observations $X_1, ..., X_n$, let $X_{(n)} = \max\{X_1, ..., X_n\}$ be the maximal observation. Show that $(X_{(n)}, \sum_{i=1}^n I(X_i > 0))$ is a sufficient statistic for θ. Furthermore, show that $\widehat{\theta} = X_{(n)}$ maximizes the observed likelihood function.

(c) (5 points) What is the exact distribution of $\widehat{\theta}$? Compute $E[\widehat{\theta}]$ and $Var(\widehat{\theta})$ and show that $\widehat{\theta}$ is consistent for θ.

(d) (6 points) Derive the asymptotic distribution of $n(\widehat{\theta} - \theta)$.

Now assume that both p and θ are unknown.

(e) (6 points) Calculate the maximum likelihood estimator for p to obtain the maximum likelihood estimator for $E[X_1]$. Derive the asymptotic distribution for the latter after proper normalization.

2. (25 points) Suppose that $y_1, \ldots, y_n$ are independent binary random variables, where

$$P(y_i = 1|\beta_0, \beta_1, x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)},$$

where $x_1, \ldots, x_n$ are fixed covariates and they are not all equal.

(a) (6 points) Suppose that (β_0, β_1) are both unknown and suppose we wish to test

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0.$$

Derive the Uniformly Most Powerful Unbiased (UMPU) α level test for this hypothesis and express the rejection region and critical value in the simplest possible form. Please note that there need not be a closed form for the distribution of the test statistic.

(b) (5 points) Using the UMPU conditional test from part (a), compute an explicit closed form for its conditional mean and conditional variance under the null hypothesis to find an explicit form for an asymptotically correct approximation to the UMPU test. You are allowed to assume that the conditional test statistic is asymptotically normal.

(c) (4 points) Derive the score test for the hypothesis in part (a), and compare its form to the approximate UMPU test derived in part (b).

(d) (6 points) Now consider the more general problem in which we have p covariates, and

$$P(y_i = 1|\boldsymbol{\beta}, \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})},$$

where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ is a $p \times 1$ vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients. Suppose we wish to test

$$H_0 : \ell'\boldsymbol{\beta} = \theta_0 \quad \text{versus} \quad H_1 : \ell'\boldsymbol{\beta} \neq \theta_0,$$

where θ_0 is a specified scalar and ℓ is a specified and non-zero $p \times 1$ vector. Derive the UMPU size α test for this hypothesis and express the rejection region and critical value in the simplest possible form.

(e) (4 points) Describe in detail a non-parametric bootstrap algorithm for computing the exact p-value based on the UMPU test of part (d).

3. (25 points) Suppose $S \sim \text{Binomial}(n, p)$ and conditional on $S = s$, let $X_1, \ldots, X_{s+1}$ be iid from a $N(\mu, 1)$ distribution. The value of n is known whereas (p, μ) are both unknown, $0 < p < 1$, and $-\infty < \mu < \infty$. We observe $(S, X_1, \ldots, X_{S+1})$ and we wish to test $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ at level α.

   (a) (5 points) Write out the joint density of $(S, X_1, \ldots, X_{S+1})$ and show that it belongs to a full-rank exponential family, and find the two dimensional complete sufficient statistic. Do the same thing for the special case that $\mu = 0$.

   (b) (3 points) Derive the joint MLE's of (p, μ), denoted by $(\hat{p}, \hat{\mu})$.

   (c) (5 points) Assuming that standard MLE theory applies, derive the joint asymptotic distribution of $(\hat{p}, \hat{\mu})$, properly normalized.

   (d) (6 points) Let $\phi(S, X_1, \ldots, X_{S+1})$ be *any* unbiased level α test of H_0 versus H_1. Write out what unbiasedness means for the power function $\beta(p, \mu)$ of such a test, and explain in detail why unbiasedness implies that $\beta(p, 0) = \alpha$ for all p.

   (e) (6 points) Find the complete form of the UMPU test of H_0 versus H_1, including specification of the rejection region in terms of the sample mean of the X_i's and the $1 - \alpha$ quantile of a well known distribution.

2016 PhD Theory Exam, Section 1

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