BASIC PHD WRITTEN EXAMINATION IN BIOSTATISTICS

THEORY, SECTION 1

(9:00 AM–1:00 PM, July 27, 2015)

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 $X_1, \ldots, X_n$ be an i.i.d. sample from the density

$$f(x) = \alpha(x-\mu)^{\alpha-1}1\{\mu \le x \le \mu+1\},\ 0<\alpha<\infty,\ -\infty<\mu<\infty,$$

where $1\{A\}$ is the indicator of A. Let $X_{(1)}$ and $X_{(n)}$ be, respectively, the smallest and largest values of the sample. Do the following:

(a) (3 points) Compute $E(X_1-\mu)^{-r}$ and show it is bounded for any $r<\alpha$.

(b) (6 points) Assume that μ is known. Show that the MLE of α is $\tilde{\alpha}_n = \left[-n^{-1}\sum_{i=1}^n \log(X_i-\mu)\right]^{-1}$ and that $\sqrt{n}(\tilde{\alpha}_n-\alpha) \to_d N(0,\alpha^2)$.

For the remainder of the problem, assume that both μ and α are unknown.

(c) (6 points) Define $\tilde{\mu}_n = X_{(1)}$, $\hat{\mu}_n = X_{(n)}-1$, $Y_n = n^{1/\alpha}(\tilde{\mu}_n-\mu)$ and $Z_n = n(\mu-\hat{\mu}_n)$, and show that, for all $0 \le y, z < \infty$, $\Pr(Y_n > y, Z_n > z) \to e^{-y^\alpha-\alpha z}$, as $n\to\infty$, and that $Y_n, Z_n \ge 0$ almost surely for all $n \ge 1$.

(d) (6 points) Let $\hat{\alpha}_n = \left[-n^{-1}\sum_{i=1}^n \log(X_i-\hat{\mu}_n)\right]^{-1}$, and show that for any $0<s<\alpha$, $\hat{\alpha}_n - \tilde{\alpha}_n = O_P(n^{-1\wedge s})$, where $a\wedge b$ denotes the minimum of a and b. Hint: you may use without proof the fact that for any $0<r\le 1$, there exists a constant $0<C_r<\infty$ such that $\log(1+\Delta) \le C_r\Delta^r$, for all $0\le\Delta<\infty$. Using this fact, show that

$$0 \le \frac{1}{\tilde{\alpha}_n} - \frac{1}{\hat{\alpha}_n} \le C_{s\wedge 1}|\hat{\mu}_n-\mu|^{s\wedge 1} n^{-1}\sum_{i=1}^n (X_i-\mu)^{-s\wedge 1},$$

and then complete the proof.

(e) (4 points) Show that for any $1/2<\alpha<\infty$, $\sqrt{n}(\hat{\alpha}_n-\alpha) \to_d N(0,\alpha^2)$. Hint: use part (d) above.

2. (25 points) Suppose that the distribution of a discrete random variable X is given below:

X	-2	-1	0	1	2
$p(x)$	$\theta_1(1-\theta_2)$	$\left(\frac{1}{2}-\alpha\right)\left(\frac{1-\theta_1}{1-\alpha}\right)$	$\alpha\left(\frac{1-\theta_1}{1-\alpha}\right)$	$\left(\frac{1}{2}-\alpha\right)\left(\frac{1-\theta_1}{1-\alpha}\right)$	$\theta_1\theta_2$

where $0 < \theta_1 \leq \alpha < 1/2$, α is *known*, $0 < \theta_2 < 1$, and both (θ_1, θ_2) are unknown. Suppose we wish to test the hypothesis

$$H_0 : \theta_1 = \alpha, \ \theta_2 = 1/2$$
$$H_1 : \theta_1 < \alpha, \ \theta_2 \neq 1/2$$

at level α (the same value as the known parameter in the distribution of X) based on one observation X.

(a) (6 points) Derive the α level likelihood ratio test (LRT) for H_0 versus H_1 and obtain its power function.

(b) (4 points) Consider the transformation $g(x) = -x$. Show that distribution of X under H_0, the hypotheses, and the LRT are invariant under this transformation.

(c) (6 points) Derive a uniformly most powerful (UMP) α level invariant test for the hypothesis above and compare its power function with that of the LRT.

(d) (5 points) Is there a unique non-randomized UMP α level invariant test? If so, provide it and justify your answer in either case.

(e) (4 points) Derive the least powerful α level non-randomized invariant test.

3. (25 points) Let $(X_1, Y_1)^{\mathrm{T}}, ..., (X_n, Y_n)^{\mathrm{T}}$ be n i.i.d random vectors with

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim MN\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \Sigma\right), \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} > 0.$$

Our goal is to estimate $(\mu_X, \mu_Y)^{\mathrm{T}}$ and Σ. However, the actual data collected may have Y missing. If we let R_i indicate whether Y_i is observed, then the observed data from these n subjects consist of $\left\{R_i, R_i(X_i, Y_i)^{\mathrm{T}} + (1 - R_i)(X_i, 0)^{\mathrm{T}}, i = 1, ..., n\right\}$, where missing values are replaced by 0's but are meaningless. We assume that $R_1, ..., R_n$ are i.i.d and that they are independent of $(X_1, Y_1)^{\mathrm{T}}, ..., (X_n, Y_n)^{\mathrm{T}}$. Moreover, $\pi = P(R_1 = 1)$ is a known positive constant. The joint likelihood of the observed data is

$$\prod_{i=1}^{n} \left[f(X_i, Y_i)^{R_i} \left\{ \int f(X_i, y) dy \right\}^{1-R_i} \pi^{R_i}(1-\pi)^{1-R_i} \right],$$

where $f(x, y)$ is the joint density of (X_1, Y_1).

(a) (3 points) Show that all model parameters are identifiable.

(b) (6 points) Write down the detail of the EM algorithm for calculating the maximum likelihood estimators for the parameters.

(c) (4 points) Give the asymptotic joint distribution for the maximum likelihood estimator for $(\mu_X, \mu_Y)^{\mathrm{T}}$ in terms of the true parameters.

To estimate μ_Y, we can impute missing Y_i's value as follows. We fit a linear regression model: $Y = \alpha + \beta X + \epsilon$, using only the complete data ($R_i = 1$) and assuming ϵ and X to be independent. For subject i with missing Y_i, we then impute Y_i as $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i$, where $(\widehat{\alpha}, \widehat{\beta})^{\mathrm{T}}$ is obtained from the maximum likelihood estimation using only the complete data under this linear model. Finally, an estimator for μ_Y is

$$\widehat{\mu}_Y = n^{-1} \sum_{i=1}^{n} \left\{ R_i Y_i + (1 - R_i) \widehat{Y}_i \right\}.$$

Please answer the following questions.

(d) (3 points) Identify the true values for α and β in terms of μ's and Σ. What is the distribution of ϵ?

(e) (4 points) Give the asymptotic distribution of the maximum likelihood estimator $(\widehat{\alpha}, \widehat{\beta})^{\mathrm{T}}$.

(f) (5 points) Describe how you would obtain the asymptotic distribution of $\widehat{\mu}_Y$. You don't need to provide the explicit expression for the distribution.

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