BASIC PHD WRITTEN EXAMINATION IN BIOSTATISTICS

THEORY, SECTION 1

(9:00 AM- 1:00 PM
Tuesday, August 7, 2012)

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. Let N be a Poisson random variable with parameter μ, and let $X_1, X_2, \ldots,$ be a sequence of i.i.d. Poisson random variables with parameter λ, where $0 < \mu, \lambda < \infty$. Define

$$U = 1(N > 0) \sum_{i=1}^{N} X_i,$$

where $1(A)$ is the indicator of the event A. Do the following:

(a) Show that $E(U) = \mu\lambda$ and $\text{var}(U) = \mu\lambda(1 + \lambda)$.

(b) In this part, we add a subscript k to the Poisson parameters μ and λ defined above to denote dependence on an integer $k \geq 1$. Specifically let $\mu = \mu_k = k$ and $\lambda = \lambda_k = h/k$, where $0 < h < \infty$ is a fixed scalar. We want to study what happens to U as $k \to \infty$. Let $D_i = 1(X_i = 1)$, for all $i \geq 1$, and define

$$T = 1(N > 0) \sum_{i=1}^{N} D_i.$$

Do the following:

(i) Derive the limits of $E(U)$ and $\text{var}(U)$ as $k \to \infty$.

(ii) Show that $\text{pr}(X_i \neq D_i) = \lambda_k^2\{1 + o(\lambda_k)\}$ as $k \to \infty$.

(iii) Show that

$$1(U \neq T) \leq 1(N > 0) \sum_{i=1}^{N} 1(X_i \neq D_i);$$

and thus $U - T \to 0$, in probability, as $k \to \infty$.

(iv) Show that $T - \sum_{i=1}^{k} D_i \to 0$, in probability, as $k \to \infty$.

(v) Show that U converges in distribution to a Poisson random variable with parameter h, as $k \to \infty$.

(c) We now modify the setting in (b) so that $\mu = \mu_k = h/k$ and $\lambda = \lambda_k = k$. Do the following:

(i) Derive the limits of $E(U)$ and $\text{var}(U)$ as $k \to \infty$.

(ii) Show that $U \to 0$ in distribution as $k \to \infty$.

2. Let $Y_1, ..., Y_n$ be i.i.d random variables from a distribution with mean μ and finite variance. Due to non-response, we may not be able to observe all the Y_i's for these n subjects. Let $R_1, ..., R_n$ denote indicator of response, i.e., $R_i = 1$ means that Y_i observed and $R_i = 0$ otherwise. Suppose that we also collect additional information $X_1, ..., X_n$, which are i.i.d random variables, from these n subjects. Assume that R_i and Y_i are independent given X_i and that the random vectors (Y_i, R_i, X_i) are i.i.d. for $i = 1, \ldots, n$. Define $\pi(x) = P(R_i = 1|X_i = x)$ and assume $\pi(x)$ is known and bounded by a positive constant from below for any x in the support of X_i.

   (a) A simple estimator for μ is the average of the observed Y_i's:
   $$\hat{\mu}_1 = \sum_{i=1}^{n} R_i Y_i / \sum_{i=1}^{n} R_i.$$
   Derive the asymptotic limit of $\hat{\mu}_1$, denoted by μ^*, and give the asymptotic distribution of $\sqrt{n}(\hat{\mu}_1 - \mu^*)$. Leave expressions in the result.

   (b) A Horwitz-Thompson estimator for μ is given as
   $$\hat{\mu}_2 = n^{-1} \sum_{i=1}^{n} R_i Y_i / \pi(X_i).$$
   Show that $\hat{\mu}_2$ is a consistent estimator for μ and derive the asymptotic distribution of $\sqrt{n}(\hat{\mu}_2 - \mu)$. Leave expressions in the result.

   (c) For any measurable function $g(X_i)$ with finite second moment, we define
   $$\hat{\mu}_g = n^{-1} \left\{ \sum_{i=1}^{n} R_i Y_i / \pi(X_i) + \sum_{i=1}^{n} (1 - R_i/\pi_i(X_i)) g(X_i) \right\}.$$
   Show that $\hat{\mu}_g$ is a consistent estimator for μ and derive the asymptotic distribution of $\sqrt{n}(\hat{\mu}_g - \mu)$. Leave expressions in the result.

   (d) Determine a function g which minimizes the asymptotic variance of $\hat{\mu}_g$. Denote this function by $g^*(x)$.

   (e) Suppose that X_i is a discrete random variable with K categories. Suggest a consistent estimator for $g^*(x)$, denoted by $\hat{g}$. Justify your answer.

   (f) Following (e), derive the asymptotic distribution of $\sqrt{n}(\hat{\mu}_{\hat{g}} - \mu)$.

   (g) How would you estimate g^* if X_i is a continuous variable?

3. (a) In this part, let T_0 be an unbiased estimator of an unknown parameter θ and consider the properties of T_0 under squared error loss.

   (i) Show that T_0+c is not a minimax estimator under squared error loss, where $c \neq 0$ is a known constant.

   (ii) Show that the estimator cT_0 is not minimax under squared error loss unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T of θ, where $c \in (0,1)$ is a known constant and $R_T(\theta)$ is the frequentist risk function for T.

   (b) In this part, let $X = 1$ or 0 with probabilities p and q respectively, and consider the estimation of p with loss function $L(p,a)$ equal to 1 when $|a-p| \geq 0.25$ and equal to 0 otherwise. The most general randomized estimator is $T_0 = U$ when $X = 0$ and $T_0 = V$ when $X = 1$, where U and V are two random variables with known distributions.

   (i) Evaluate the risk function and the maximum risk of T_0 when U and V are uniform on $(0, 0.5)$ and $(0.5, 1)$, respectively.

   (ii) Is T_0 is minimax? Justify your answer rigorously.

   (c) In this part, one has a sample of n iid normal random variables with mean θ and variance σ^2, $X_1, \ldots, X_n$.

   (i) Assume $0 < \sigma^2 < K$ is known, where K is a finite positive constant. Is the sample mean $\overline{X}$ minimax with respect to the loss function $L(\theta, a) = \{\theta - a\}^2/\sigma^2$? Justify your answer rigorously.

   (ii) Redo part (i) without assuming σ^2 is known.

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