BASIC PHD WRITTEN EXAMINATION IN BIOSTATISTICS

THEORY, SECTION 1

(9:00 AM- 1:00 PM
Wednesday, August 12, 2009)

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 code letter, **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.

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 A and B be two different events in a probability space related to a random experiment. Suppose that n independent and identical trials of the experiment are carried out and that we observe the frequencies of occurrence of the events $A \cap B$, $A \cap B^c$, $A^c \cap B$, and $A^c \cap B^c$. The results can be summarized in the following 2×2 contingency table:

	A	A^c	Total
B	X_{11}	X_{12}	n_1
B^c	X_{21}	X_{22}	n_2
Total	m_1	m_2	n

(a) Let $p_{ij} = E(X_{ij})/n$, $i = 1, 2$, $j = 1, 2$, where $\sum_{ij} p_{ij} = 1$. The distribution of $X = (X_{11}, X_{12}, X_{21}, X_{22})$ is multinomial, wih probability function given by

$$f(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{n!}{\prod_i \prod_j x_{ij}!} \prod_i \prod_j p_{ij}^{x_{ij}}.$$

Verify that this distribution is in the exponential family of distributions, and write the distribution in its canonical form.

(b) Show that A and B are independent if and only if $\log(\frac{p_{11}}{p_{22}}) = \log(\frac{p_{12}}{p_{22}}) + \log(\frac{p_{21}}{p_{22}})$.

(c) Let $\theta = a_0 \log(\frac{p_{11}}{p_{22}}) + a_1 \log(\frac{p_{12}}{p_{22}}) + a_2 \log(\frac{p_{21}}{p_{22}})$, where (a_0, a_1, a_2) are given constants. Assuming that $a_0 = 1$ and $a_1 = a_2 = -1$, derive a UMPU size α test for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, and derive the conditional power function of the test. (Hint: Use a theorem for multiparameter exponential families to construct the UMPU test).

(d) Derive a UMPU size α test for testing $H_0 : P(A) \geq P(B)$ versus $H_1 : P(A) < P(B)$. (Hint: Use the techniques of part (c) in setting up the hypothesis in terms of θ and then constructing the test).

(e) Derive the likelihood ratio statistic, denoted by Λ_n, for the hypothesis in part (c) and show that it is asymptotically equivalent to the Pearson chi-square statistic. Specifically,

(i) show that

$$-2\log(\Lambda_n) = \sum_{j=1}^{2} \sum_{i=1}^{2} \frac{(X_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}} + o_p(1),$$

where $\hat{p}_{ij}$ denotes the maximum likelihood estimate of p_{ij} under H_0.

(ii) find the asymptotic distribution of $-2\log(\Lambda_n)$ under H_0 and H_1.

Scoring: (a) (2 points); (b) (3 points); (c) (7 points); (d) (6 points); (e)(i)(5 points), (ii) (2 points).

2. Let λ have exponential density $\theta e^{-\theta\lambda}$, for $0 < \theta < \infty$. Conditional on λ, let (X, Y) be a pair of independent Poisson random variables with respective p.m.f.'s $\lambda^x e^{-\lambda}/x!, x = 0, 1, \ldots,$ and $(\beta\lambda)^y e^{-\beta\lambda}/y!, y = 0, 1, \ldots,$ for $0 < \beta < \infty$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d. sample where (X_1, Y_1) has the same unconditional joint distribution as (X, Y). Do the following:

(a) Determine the following properties of the unconditional distribution of (X, Y):

(i) Show that $EX = \theta^{-1}$, $EY = \beta\theta^{-1}$, $\mathrm{var}(X) = \theta^{-1} + \theta^{-2}$, $\mathrm{var}(Y) = \beta\theta^{-1} + \beta^2\theta^{-2}$, and $\mathrm{cov}(X, Y) = \beta\theta^{-2}$.

(ii) Show that the unconditional joint density of (X, Y) is

$$\left(\frac{\theta}{\theta + \beta + 1}\right) \frac{(x + y)!}{x!y!} \left(\frac{1}{\theta + \beta + 1}\right)^x \left(\frac{\beta}{\theta + \beta + 1}\right)^y.$$

(b) Show that the maximum likelihood estimator based on a sample of size n for θ is $\hat{\theta}_n = \overline{X}_n^{-1}$ and for β is $\hat{\beta}_n = \overline{Y}_n/\overline{X}_n$, where $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$.

(c) Show that

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta \\ \hat{\beta}_n - \beta \end{pmatrix} \to_d N\left(0, \begin{bmatrix} \theta^2(\theta + 1) & \beta\theta^2 \\ \beta\theta^2 & \theta\beta(\beta + 1) \end{bmatrix}\right).$$

(d) Let $T_1 = \sqrt{n\overline{X}_n/2}\left(\overline{Y}_n/\overline{X}_n - 1\right)$ and $T_2 = \sqrt{n\overline{X}_n/2}\,\ln\left(\overline{Y}_n/\overline{X}_n\right)$, and show that under the null hypothesis $H_0 : \beta = 1$, that

(i) $T_1 \to_d N(0, 1)$,

(ii) $T_1 - T_2 \to_p 0$, and

(iii) $T_2 \to_d N(0, 1)$.

(e) Suppose $\beta = 1$ and we wish to make inference on $\tau = \theta/(\theta + 2)$. Using result (ii) of part (a), derive the Bayes estimator of τ under squared error loss with prior density

$$\pi(\tau) \propto \tau^{a_0 - 1}(1 - \tau)^{b_0 - 1},$$

where the scalars $a_0 > 0$ and $b_0 > 0$ are specified hyperparameters. Show that this Bayes estimator is admissible.

Scoring:

(a)(i)(3 points), (ii) 3 points; (b) (4 points); (c) (5 points); (d)(i)(2 points), (ii) (2 points), (iii) (1 point); (e) (5 points).

3. Suppose that there is a random variable A having discrete probability distribution with support on the non-negative integers. Assume that $P(A = j) = p_j, j = 0, 1, 2, \ldots$, such that $p_j > 0$ and $\sum_j p_j = 1$.

Define the random variables $B_n, n = 1, 2, \ldots$ recursively:

$$B_{n+1} = \sum_{k=1}^{B_n} A_{nk},$$

where $A_{nk}, n = 1, 2, \ldots, k = 1, \ldots, B_n$ are iid with the same distribution as A. Assume that B_0 is a known positive integer and let $P_{ij} = P(B_{n+1} = j | B_n = i) = P(\sum_{k=1}^{i} A_{nk} = j)$.

(a) Show that, in general, $E(B_{n+1}) = E(B_n)E(A_{nk})$ and $\text{Var}(B_{n+1}) = E(B_n)\text{Var}(A_{nk}) + \text{Var}(B_n)\{E(A_{nk})\}^2$, $n \geq 1$.

In the sequel, suppose that $B_0 = 1$ and that $E(A_{nk}) = \mu < \infty$ and $\text{Var}(A_{nk}) = \sigma^2 < \infty, n = 1, 2, \ldots, k = 1, \ldots, B_n$

(b) Show that for $n \geq 1$, $E(B_n) = \mu^n$ and $\text{Var}(B_n) = \sigma^2\mu^{n-1}(1 - \mu^n)\{1 - \mu\}^{-1}$ if $\mu \neq 1$ and $n\sigma^2$ if $\mu = 1$.

(c) Under the conditions in (b), show that if $\mu < 1$ then $P(B_n = 0) \to 1$ as $n \to \infty$.

(d) Define $P_{1j}^{(n)} = P(B_n = j | B_0 = 1)$. Show that $E(z^{B_n})$, denoted by $\phi_n(z)$, may be expressed as $\sum_{j=0}^{\infty} P_{1j}^{(n)} z^j$, for a scalar z such that $|z| \leq 1$. Demonstrate that $\phi_n(0) = P(B_n = 0 | B_0 = 1)$.

(e) Define $\phi(z) = E(z^{A_{nk}}) = \sum_{j=0}^{\infty} p_j z^j$, for a scalar z, with $|z| \leq 1$, for $k = 1, \ldots, B_n$. Show that $\sum_{j=0}^{\infty} P(A_{n1} + A_{n2} + \ldots A_{nk} = j | B_n = k) z^j = \{\phi(z)\}^k$.

(f) Establish that the following recursive relationship holds:

$$\phi_n(z) = \phi_{n-1}\{\phi(z)\}, n \geq 1.$$

(Hint: condition on B_{n-1}).

Scoring:

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