MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Friday, August 9, 2013: 9:00 AM - 3:00PM

Room: BCBS Auditorium

INSTRUCTIONS:

- This is a **CLOSED BOOK** examination.
- Submit answers to **exactly** 3 out of 4 questions. If you submit answers to more than 3 questions, then only questions 1-3 will be counted.
- Put the answers to different questions on **separate sets of paper**. Write on **one side** of the sheet only.
- Put your code letter, **not your name**, on each page, in the upper right corner.
- Return the examination with a **signed honor pledge form**, separate from your answers.
- You are required to answer **only what is asked** in the questions and not to tell all you know about the topics.

1. A study collected data on ear infections and respiratory infections contracted by a random sample of children between the ages of 1 and 5. Here we consider a simple model that might be used in the analysis.

   Let the random variable X denote the number of respiratory infections and the random variable Y denote the number of ear infections developed by a given child. Suppose that there is an unobservable random variable U that represents the child's propensity to develop infections; children with larger values of U tend to develop more infections. Suppose further that, U is distributed as exponential with mean $\mu > 0$, and that given U, the random variables X and Y are conditionally independent and identically distributed as Poisson random variables with mean U.

   In what follows, derive explicit expressions and simplify them as much as possible. Show *all* your derivations, not just the final answer. Hint: Conditioning.

   (a) Find $\mathrm{E}[X]$ and $\mathrm{Var}(X)$. Does X have a Poisson distribution?

   (b) Find $\mathrm{Corr}(X, Y)$ and $\mathrm{Corr}(X, U)$. Which correlation is larger?

   (c) Is X an unbiased predictor of U? Compute its prediction variance. Note: If a random variable W is viewed as a predictor of U, we say that W is an *unbiased predictor* of U if $\mathrm{E}[W - U] = 0$. We call $\mathrm{Var}(W - U)$ the *prediction variance.*

   (d) Find constants a and b such that the random variable $a + bX$ is an unbiased predictor of U and such that its prediction variance is as small as possible; that is $\mathrm{E}[U - (a + bX)] = 0$ and $\mathrm{Var}(U - (a + bX))$ is minimized.

   (e) We are interested in the probability of no ear infections; overall and in children with no respiratory infections. Develop expressions for $P(Y = 0)$ and $P(Y = 0|X = 0)$. Which probability is larger? Hint: Consider the ratio $P(Y = 0|X = 0)/P(Y = 0)$.

   Points: (a) 1, (b) 3, (c) 3, (d) 9, (e) 9.

2. Let $X_1, X_2, \cdots, X_n$ be iid random variables from the discrete distribution with pmf

$$f(j) = \theta_j \qquad j \in \{1, 2, 3\},$$

where $\theta_j > 0, j = 1, 2, 3$ and $\theta_1 + \theta_2 + \theta_3 = 1$. Define T_n to be the sample minimum, $T_n = X_{(1)}$, and define M to be the sample median.

Note: The usual formulae for the pdf of order statistics assume continuous random variables, which is not the case here.

(a) If $n \geq 3$, derive an expression for the probability that each of the three possible values will be observed at least once in the sample.

(b) Derive the pmf of T_n.

(c) Prove whether or not T_n converges in probability to a constant as $n \to \infty$. If it does, identify that constant.

(d) Now consider the specific case of $n = 3$ and $\theta_1 = \theta_2 = \theta_3 = \frac{1}{3}$. Compute the pmf of the sample median. Compute the mean and the variance of the sample median.

Points: (a) 6, (b) 6, (c) 6, (d) 7.

3. Let $X_1, \cdots, X_n$ be a random sample from the probability density function

$$f(x|\theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}}, \qquad x > 0, \theta > 0.$$

In what follows, derive explicit expressions and simplify them as much as possible. Show *all* your derivations, not just the final answer.

(a) Find a minimal sufficient statistic for θ.

(b) Find the maximum likelihood estimator of θ (show that it is a maximum).

(c) Find the expected value of the estimator in (b).

(d) Is the estimator in (b) UMVUE of θ? Explain.

(e) Find the Cramer-Rao lower bound on the variance of unbiased estimators of θ.

(f) Find the Cramer-Rao lower bound on the variance of unbiased estimators of θ^2.

(g) Let $Y_1, \cdots, Y_m$ be a random sample from the pdf

$$g(y|\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, \qquad y > 0, \theta > 0.$$

Note that $f(x|\theta)$ and $g(y|\theta)$ involve the same (unknown) parameter θ. Let $\hat{\theta}_1$ be the MLE of θ in (b) above based on $n = 32$ observation, and $\hat{\theta}_2$ be the MLE of θ based on the sample $Y_1, \cdots, Y_m$ of $m = 44$ observations. Which estimator of θ ($\hat{\theta}_1$ or $\hat{\theta}_2$) is preferable? Justify your answer.

Points: (a) 4, (b) 5, (c) 2, (d) 2, (e) 5, (f) 2, (g) 5.

4. Let $X_1, \cdots, X_n$ be a random sample from the pmf

$$P(X_i = j) = P_j, \qquad j = 1, \cdots, 4.$$

The vector parameter $\theta = (P_1, P_2, P_3, P_4)^\top$ satisfies $\sum_{j=1}^4 P_j = 1$ and $P_j > 0, j = 1, \cdots, 4$. Let

$$y_j = \#\{x_i | x_i = j, i = 1, \cdots, n\}, \qquad j = 1, \cdots, 4.$$

Consider testing the hypothesis $H_0 : P_1 = P_2$ and $P_3 = P_4$ versus $H_a : H_0$ is not true.

(a) Derive the maximum likelihood estimator of θ under H_0.

(b) Derive the maximun likelihood estimator of θ under H_a.

(c) Develop the maximum likelihood ratio test for testing H_0 versus H_a.

(d) Identify a distribution that may be used in constructing the rejection region for the test in (c). You may assume that n is large.

Points: (a) 7, (b) 7, (c) 7, (d) 4