MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Friday, August 12, 2011: 9:00 AM - 3:00PM

Room: MG 1305

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.
- 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 certain village is infested with a colony of rats and we are keenly interested in studying how quickly the population of rats grows. Suppose we know that the rat colony contains n female rats and that in a particular year, the number of (female) offspring that each female rat produces follows a distribution with probability mass function given by

$$f(x) = \begin{cases} \frac{\alpha}{x!} & x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

for some fixed constant α. Assume that all offspring are able to breed in the next year and that all rats behave independently.

(a) For what value of α is the above a legitimate distribution?

(b) What is the expected number of female offspring that each female rat will produce in a year? What is the variance?

(c) Let Y_1 denote the total number of female offspring produced in one year from the n parents. Find the PMF, expectation, and variance of Y_1 in terms of n.

(d) Assume that no female rats ever die. Then after a single year, there are $Y_1 + n$ female rats which can each produce offspring as before. Let Y_k denote the number of female *offspring* produced in the k^{th} year. What is the distribution of $Y_k | Y_{k-1}, Y_{k-2}, \ldots, Y_1$?

(e) Inductively show that $E[Y_k] = 2^{k-1}n$.

(f) Compute the covariance between Y_1 and $Y_2 - Y_1$. Are Y_1 and $Y_2 - Y_1$ independent?

(g) An exterminator is hired each year to expose all female rats to a chemical. In any given year, all rats are exposed and exposure to the chemical causes a female rat to die with probability $1 - p$. Female rats that die are removed from the population and do not produce offspring in the particular year that they are exposed/die. Survival in a given year is independent of age. Offspring respond to exposure independently of the mother. In terms of p, compute the expected number of female rats after the first year (both surviving members of the original colony and new offspring).

(h) For what values of p does the expected number of female rats decrease from one year to the next?

Points: (a) 2, (b) 2, (c) 3, (d) 3, (e) 4, (f) 4 , (g) 4, (h) 3.

2. Let $Y_1, Y_2, \ldots, Y_n$ be *i.i.d.* Unif(0,1) random variables. The notation $Y_{(k)}$ will be used to denote the k-th order-statistic.

   (a) Find the moment generating function (MGF) and characteristic function (CF) of Y_1.

   (b) Show that $Y_{(1)} \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

   (c) Compute the density function of the median of $Y_1, \ldots, Y_n$ when n is <u>even</u>, say $n = 2m$, and the median is $T = (Y_{(m)} + Y_{(m+1)})/2$.

   (d) Find the MGF of $-\ln Y_{(k)}$ and show that $-\ln Y_{(k)}$ has the same distribution as $\sum_{i=1}^{n-k+1} X_i$, where $X_1, \ldots, X_{n-k+1}$ are independent, exponentially distributed random variables.

   Points: (a) 6, (b) 6, (c) 8, (d) 5.

3. Suppose X is distributed on $(0,1)$ with probability density

$$f(x|\theta) = (1-\theta) + \frac{\theta}{2\sqrt{x}} \qquad 0 < x < 1, 0 \leq \theta \leq 1.$$

Note: There is only 1 observation in this problem, sample size is 1.

(a) Find $E_\theta[X]$. Note: The subscript in E_θ is to emphasize that the expectation depends on θ.

(b) Consider the estimator

$$\delta(X) = \begin{cases} a(X^{-1/2} + b) & \text{if } c < X < 1 \\ 0 & \text{if } 0 < X \leq c \end{cases}$$

where a, b and c are given constants.

Find $E_\theta[\delta(X)]$. Further, verify that $E_\theta[\delta(X)]$ can be expressed as

$$E_\theta[\delta(X)] = (1-\theta)E_0[\delta(X)] + \theta E_1[\delta(X)].$$

Reminder: E_0 is the expectation when $\theta = 0$ and E_1 is the expectation when $\theta = 1$.

(c) Show that for any choice of $c \in (0,1)$, there exist a and b such that $\delta(X)$ is an unbiased estimator of θ. Derive explicit expressions for such a and b (in terms of c). Show that $a \to 0$ as $c \to 0$.

(d) Assuming that a and b have been chosen so that $\delta(X)$ is unbiased for θ, show that $\text{Var}_0(\delta(X)) = 2a$. ($\text{Var}_0(\delta(X))$ is the variance of $\delta(X)$ when $\theta = 0$).

(e) Using the last two parts, show that an estimator, $\delta(X)$, that is unbiased for θ, can be found such that $\text{Var}_0(\delta(X))$ is arbitrarily close to 0. What does this say about the existence of a uniformly minimum-variance unbiased estimator (UMVUE) of θ?

Points: (a) 5, (b) 5, (c) 5, (d) 5, (e) 5.

4. Let $X_1, \cdots, X_n$ be iid random variables from the pdf

$$f(x|\theta) = (1-\theta) + \frac{\theta}{2\sqrt{x}} \qquad 0 < x < 1, 0 \leq \theta \leq 1.$$

That is, we have a random sample of size n from the population f. The parameter θ is unknown.

(a) Derive the uniformly most powerful level α test ($0 < \alpha < 1$) for $H_0 : \theta = 0$ against $H_1 : \theta = 1$. Specifiy the critical region as concisely and as explicitly as possible. Justify your answers. Do not use any approximations.

(b) For the special case of $n = 5$ and $\alpha = 0.01$, find the critical region (exactly). Also, find the (exact) power of the test. You can not compute the power without a computer, so simply express it in terms of cdf's of standard distributions (e.g. standard normal, t, F, etc).

(c) An investigator wants to design a study in which the test derived above will be applied. The investigator desires a Type I Error probability of 0.01 and a Type II Error probability of 0.01. Find the minimum required sample size n (exact, or approximate, whichever is easier).

(d) This part pertains to the special case of $n = 1$ (sample size $= 1$). Find $\hat{\theta}$, the maximum-likelihood estimator (MLE) of θ. Show that the MLE is biased. Then find constants a and b such that $T(X_1) = a + b\hat{\theta}$ is unbiased for θ. Do you see any potential problems with $T(X_1)$ as an estimator of θ?

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