MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Friday, August 10, 2012: 9:00 AM - 3:00PM

Room: 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, 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. Suppose that the random variable X is distributed as Poisson with mean $\mu > 0$, and that given $X = x$, the random variable Y is distributed as Poisson with mean x.

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

   Hint: Conditioning is very helpful in several parts of this problem.

   (a) Find $\mathrm{E}[Y]$ and $\mathrm{Var}(Y)$.

   (b) Find $\mathrm{Corr}(X, Y)$.

   (c) Find the conditional probability mass function of X given $Y = 0$. Does it follow a recognizable form? If so, identify it.

   (d) Find $\mathrm{E}[X|Y \geq 1]$. Hint: Make use of the result in (c) and of what is known about X.

   (e) Find $\mathrm{E}[\frac{Y}{X+1}]$. Verify whether this expectation is less than 1 for all μ. Note: Evaluating an infinite sum is needed to get the final answer.

   (f) Define $T = wX + (1 - w)Y$, where w is a non-random constant, $w \in [0, 1]$. If T is to be used as an estimator of μ, what is the best choice of w? Justify. Why does the answer make sense?

   Points: (a) 3, (b) 3, (c) 5, (d) 5, (e) 5, (f) 4.

2. The glucose tolerance level in patients with diabetes is normally distributed with mean 6 and variance 9. The glucose tolerance level in people without diabetes has a normal distribution with mean 4 and variance 4. When glucose tolerance is used as a test for diabetes, individuals with glucose tolerance level of 5 and greater are classified as having diabetes (that is, the test is considered positive), while individuals with glucose tolerance level below 5 are classified as non-diabetic (that is, the test is considered negative). Let p denote the probability that a randomly chosen person actually has diabetes, $0 < p < 1$.

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

   (a) What is the probability that a person who has diabetes gets correctly classified as having diabetes according to the result of the glucose tolerance test?

   (b) What is the probability that a person who does not have diabetes gets correctly classified as diabetes free by the glucose tolerance test?

   (c) What is the probability of a positive test result for a randomly chosen person? The answer is a function of p.

   (d) A randomly chosen subject has tested positive. What is the probability that he/she actually has diabetes? The answer is a function of p.

   (e) A randomly chosen subject has a glucose tolerance level of 6. What is the probability that that subject has diabetes? The answer is a function of p.

   (f) Suppose that two glucose tolerance measurements, taken one month apart, on a random diabetic individual, are distributed as bivariate normal, and their correlation is 0.5 (and means and variances are as given above). A randomly chosen diabetic subject tested today has a glucose tolerance value of 6. If the same subject is tested again a month from now, what is the probability of a positive test? (Compute the numerical answer)

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

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

$$\frac{3}{\theta} x^2 e^{-\frac{x^3}{\theta}}, \qquad\qquad \theta > 0,\ x > 0.$$

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 sufficient statistic for θ.

(b) Is the statistic in (a) minimal? If it is, provide a proof. If it is not, find a minimal sufficient statistic.

(c) Find the maximum likelihood estimator of θ.

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

(e) Is the estimator in (c) UMVUE of θ? Justify.

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

(g) Derive the rejection region for the most powerful size α test of $H_0 : \theta = \theta_0$ against $H_a : \theta = \theta_1$, where $\infty > \theta_1 > \theta_0 > 0$. Is the given test uniformaly most powerful of its size against the alternative $H_1 : \theta > \theta_0$? Justify.

Points: (a) 2 , (b) 3 , (c) 4 , (d) 4 , (e) 2 , (f) 5 , (g) 5.

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

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

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

(a) Find the maximum likelihood estimators for θ and β.

(b) Derive the likelihood ratio test for testing $H_0 : \theta = \theta_0$ against $H_a : \theta > \theta_0$, with β unknown, and θ_0 a given positive constant. Identify a distribution (approximate or exact, whichever is easier) that may be used in constructing the rejection region for the test.

(c) Derive the likelihood ratio test for testing $H_0 : \beta = \beta_0$ against $H_a : \beta > \beta_0$, with θ unknown, and β_0 a given positive constant. Identify a distribution (approximate or exact, whichever is easier) that may be used in constructing the rejection region for the test.

Points: (a) 7 , (b) 9 , (c) 9.