MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Tuesday, July 28, 2015: 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. An urn contains a red and b blue balls. After a ball is drawn, it is returned to the urn if it is red; but if it is blue, it is replaced by a red ball from another urn. Let X_n denote the number of red balls in the urn after the foregoing operation has been repeated n times.

   (a) Find $\mathrm{E}[X_1]$.

   (b) Find the moment generating function of X_1.

   (c) Show that

   $$\mathrm{E}[X_{n+1}] = \left(1 - \frac{1}{a+b}\right)\mathrm{E}[X_n] + 1$$

   (d) Use part (c) to show that

   $$\mathrm{E}[X_n] = a + b - b\left(1 - \frac{1}{a+b}\right)^n$$

   (e) Find the probability that the $(n+1)$st ball drawn is red and show that this probability converges to 1 as $n \to \infty$.

   (f) Show that X_n converges in probability to Y where $P(Y = a + b) = 1$.

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

2. Suppose that patients with a certain cancer who are treated successfully are considered disease-free until they develop a recurrence (according to some well-defined criteria). After recurrence, the disease keeps progressing until the patient dies. Let the random variable X denote time to recurrence, and the random variable Y time to death (both in years). Both times are measured from the end of treatment. That is, end of treatment is time 0, and $0 < X < Y$. Suppose that X is uniformly distributed on $[0, 1]$, and conditional on X, the random variable Y is uniformly distributed on $[X, X + 1]$.

   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 mean and variance of Y.

   (b) Find the correlation between X and Y.

   (c) Find $f_Y(y)$, the marginal pdf of Y. (Hint: Pay close attention to the support of (X, Y). It may help to split the support of Y into two parts. This hint may apply to other parts as well.)

   (d) Find $f_{X|Y}(x|y)$, the conditional pdf of X given Y.

   (e) Find $\mathrm{E}[X|Y]$, the conditional mean of X given Y.

   (f) Find $\mathrm{E}[\frac{X}{Y}]$.

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

3. Let $X_1, \ldots, X_n$ be a random sample drawn from an exponential distribution

$$f_X(x|\theta) = \theta e^{-\theta x}, \qquad x > 0,$$

with $\theta > 0$.

(a) Derive the maximum likelihood estimator, $\hat{\theta}$. Show that $1/\hat{\theta}$ has the uniformly minimum variance among all unbiased estimators of $1/\theta$.

(b) Derive the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ as $n \to \infty$, and comment on whether the asymptotic variance, i.e., the limiting variance of $\sqrt{n}(\hat{\theta} - \theta)$, reaches the Cramer-Rao lower bound times n.

(c) Suppose that in a given study, $X_1, \ldots, X_n$ are not observed but, instead, only dichotomized random variables $Z_1, \ldots, Z_n$, are observed; $Z_i = 1$ if $X_i > \tau$, and $Z_i = 0$ if $X_i \leq \tau$, where $\tau > 0$ is a known constant. Find the maximum likelihood estimator $\tilde{\theta}$ of θ based on $Z_1, \ldots, Z_n$ and derive the large sample distribution of $\tilde{\theta}$ in explicit form.

(d) Taking τ to be $\mathrm{E}[X_1]$, compare the asymptotic variances of $\hat{\theta}$ and $\tilde{\theta}$. Which one is larger? Is that the anticipated result, and why?

(e) The *hazard function* corresponding to the pdf f_X is

$$h_X(x|\theta) = \frac{f_X(x|\theta)}{\int_x^\infty f_X(s|\theta)ds}, \qquad x > 0.$$

Now, suppose that in addition to $X_1, \ldots, X_n$, another independent random sample is available: $Y_1, \ldots, Y_n$, drawn from $f_Y(y|\beta) = \beta e^{-\beta y}, y > 0, \beta > 0$. Since, in this case, $h_X(x|\theta) = \theta, x > 0$, and $h_Y(y|\beta) = \beta, y > 0$, one can compare the hazard functions between the two populations by testing $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$, where $\psi = \theta/\beta$ is the *hazard ratio*. Derive a large sample (likelihood ratio, score, or Wald test) size α test of H_0 versus H_1.

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

4. Radiation doses delivered by a certain radiotherapy machine are claimed to be uniformly distributed between a and θ, where $0 \leq a < \theta < \infty$. The constant a is known, but θ is unknown. A quality control administrator is concerned primarily with the maximum dose parameter, θ, and wants to implement formal hypothesis testing on θ. Suppose that $X_1, \ldots, X_n$ is a random sample collected from the machine.

   (a) Show that the maximum likelihood estimator of θ is the maximum order statistic $X_{(n)}$, and prove that it is a *biased* estimator (of θ) at each finite n but a *consistent* estimator as $n \to \infty$.

   (b) Find an unbiased estimator of θ as a function of $X_{(n)}$. Comment on why the Cramer-Rao lower bound (CRLB) fails in this situation.

   (c) Find the uniformly most powerful (UMP) size α test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ with a clear specification of the cutoff in the rejection region.

   (d) Show that asymptotically, as $n \to \infty$, the rejection region in (c) is independent of α. Give an intuitive interpretation of this phenomenon.

   Points: (a) 8, (b) 5, (c) 7, (d) 5.