MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Friday, August 13, 2010: 9:00 AM - 3:00PM

Room: MH 0001, Blue Cross/Blue Shield Auditorium

INSTRUCTIONS:

a. This is a **CLOSED BOOK** examination.

b. Answer 3 out of 4 questions.

c. Put the answers to different questions on separate sets of paper; staple them separately.

d. Put your code letter, **not your name**, on each page.

e. Return the examination with a signed statement of the honor pledge on a page separate from your answers.

f. 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 V_0, V_1, V_2 are i.i.d. Bernoulli(θ) random variables, and, independently, Z is distributed as Bernoulli(γ), $0 \le \theta \le 1, 0 \le \gamma < 1$. Define

$$X = ZV_0 + (1 - Z)V_1,$$

$$Y = ZV_0 + (1 - Z)V_2.$$

In what follows, show your work, derive simple explicit expressions and justify your answers rigorously.

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

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

(c) Is Y independent of Z?

(d) Find the conditional distribution of Y given X. Show that $\mathrm{E}[Y|X]$ can be expressed as

$$\mathrm{E}[Y|X] = \alpha + \beta X,$$

and find explicit expressions for α and β.

Now suppose that the pairs $(X_1, Y_1), \cdots, (X_n, Y_n)$ are i.i.d. from the same distribution as (X, Y) above.

(e) Define the sample means $\bar{X} = (X_1 + \cdots + X_n)/n$ and $\bar{Y} = (Y_1 + \cdots + Y_n)/n$. Of the two estimators of θ:

$$T_1 = \bar{X}, \qquad T_2 = \frac{1}{2}(\bar{X} + \bar{Y}),$$

which one is better and why?

(f) If θ is known, give an unbiased estimator of γ based on the sample $(X_1, Y_1), \cdots, (X_n, Y_n)$.

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

2. Suppose that the pair (X, Y) is distributed according to the pdf

$$f_{XY}(x,y) = \frac{12xy}{\theta^6}, \quad 0 < x < \theta,\, 0 < y < x^2,\, \theta > 0.$$

In what follows, show your work, derive simple explicit expressions and justify your answers rigorously.

(a) Find the marginal pdf of Y.

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

(c) Find the conditional pdf of Y given X.

(d) Find $\mathrm{E}[Y|X]$.

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

(f) Find $\mathrm{Cov}(X, Y - \frac{2}{3}X^2)$.

(g) Find $P(Y < X)$.

Points: $\frac{25}{7}$ each part.

3. Suppose that $(X_1, Y_1), \cdots, (X_n, Y_n)$ is a random sample from a distribution with pdf

$$f_{XY}(x, y) = \frac{12xy}{\theta^6}, \qquad 0 < x < \theta,\ 0 < y < x^2,$$

where $\theta > 0$ is an unknown parameter.

In what follows, show your work, derive simple explicit expressions and justify your answers rigorously.

(a) Find a minimal sufficient statistic for θ.

(b) Show that the marginal pdf of $X_i, 1 \le i \le n$, is

$$f_{X_i}(x) = \frac{6x^5}{\theta^6}, \qquad 0 < x < \theta.$$

Explain why this density belongs to a scale family.

(c) Let $\bar{X}$ denote the mean of the X-sample, $\bar{X} = (X_1 + \cdots + X_n)/n$. Find a constant k_1 such that $\hat{\theta}_1 := k_1 \bar{X}$ is unbiased for θ.

(d) Let $X_{(n)}$ be the maximum of the X-sample. Find a constant k_2 such that $\hat{\theta}_2 := k_2 X_{(n)}$ is unbiased for θ.

(e) Of the two unbiased estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, which one is preferable? Explain rigorously why.

(f) Show that $X_{(n)}/\theta$ is a pivotal quantity, and use it to construct the shortest $1 - \alpha$ pivotal confidence interval $(0 < \alpha < 1)$. Justify rigorously why the interval is shortest.

(g) Compute the numerical values of $\hat{\theta}_1$ and $\hat{\theta}_2$ based on the 3 observations:

$$(x_1 = 5, y_1 = 20), \quad (x_2 = 13, y_2 = 151), \quad (x_3 = 18, y_3 = 222).$$

What is a weakness of $\hat{\theta}_1$ illustrated by these data?

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

4. Suppose that $Y_1, \cdots, Y_n, n > 1$, is a random sample from a distribution with pmf

$$f(y|\theta) = \theta^y(1-\theta), \qquad y = 0, 1, 2, \cdots,$$

where $\theta \in [0, 1]$ is an unknown parameter.

In what follows, show your work, derive simple explicit expressions and justify your answers rigorously.

(a) Find a complete sufficient statistic for θ.

(b) Compute the Cramer-Rao lower bound on the variance of unbiased estimators of θ.

(c) Find the UMVUE of θ. (Hint: $f(0) =?$) Does it achieve the Cramer-Rao lower bound?

(d) Derive an explicit expression for $\hat{\theta}_n$, the maximum-likelihood estimator of θ. Is $\hat{\theta}_n$ an unbiased estimator of θ? If it is biased, what is the direction of the bias?

(e) The estimator $\hat{\theta}_n$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathrm{n}(0, \nu) \quad \text{as } n \to \infty.$$

Give an explicit expression for ν.

(f) If $n = 25$ and the observed $\hat{\theta}_n$ is 0.6, compute an approximate 95% confidence interval for θ. (An exact interval is acceptable).

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