MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Wednesday, August 8, 2018, 9:00 AM - 3:00 PM

Room: BCBS Auditorium (MHRC 0001)

INSTRUCTIONS:

- This is a **CLOSED BOOK** examination.
- Answer all questions.
- Put the answers to different questions on **separate sets of paper**. Write on **one side** of the sheet only, **inside the marked box**.
- 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.
- Aim for clarity in your writing. Define any new notation you use; even common notation such as $\bar{x}$ should be defined.

1. Suppose that the conditional distribution of Y given X is normal with mean X and variance X and that the marginal distribution of X is Uniform(0,1). In what follows, show all your work and derivations, not just the final answer.

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

   (b) Find $\mathrm{Cov}(Y - X, X)$. Are $Y - X$ and X independent? Justify your answer.

   (c) Find the numerical value of $P(X > Y)$.

   (d) Suppose that $Z_1, Z_2, \cdots$ is a sequence of iid random variables each distributed as standard normal (mean 0, variance 1). For $n = 1, 2, \cdots$, derive a general expression for

   $$P\left(Z_{n+1} < Z_n | Z_n = \max(Z_1, \cdots, Z_n)\right).$$

   (e) Find the numerical value of $\mathrm{E}[|Z_1 - Z_2|]$.

   Points: 5 for each part.

2. Suppose that the random variables Y_1 and Y_2 are independent and identically distributed as exponential with mean $\mu > 0$. Define the order statistics as follows: $X_1 = \min(Y_1, Y_2)$, $X_2 = \max(Y_1, Y_2)$. Also define the range, $R = X_2 - X_1 = |Y_1 - Y_2|$, and the sample mean $\bar{Y} = (Y_1 + Y_2)/2$. In what follows, find explicit expressions and simplify them as much as possible. Show all your work and derivations, not just the final answer.

   (a) Prove that $\bar{Y}$ is the unique uniformly minimum-variance unbiased estimator of μ.

   (b) Prove rigorously and without using the formula for the pdf of order statistics, that X_1 is distributed as exponential with mean $\mu/2$.

   (c) Show that $\mathrm{E}[R] = \mu$ and find $\mathrm{Var}(R)$.

   (d) Find $\mathrm{E}[X_2]$, $\mathrm{Var}(X_2)$ and $\mathrm{Corr}(X_1, X_2)$.

   (e) Find $\mathrm{E}[R|\bar{Y}]$.

   Points: 5 for each part.

3. Human-to-mosquito transmission of *P. falciparum* malaria is mediated by sexual stage parasites called gametocytes. Mosquito feeding assay is a method to assess whether the gametocytes in a patient are transmissible to mosquitos. Through the detection of oocysts in the mosquito midgut by microscopy, researchers can confirm such transmissibility. A random sample of n subjects were assayed, and the outcomes are denoted $Y_1, \ldots, Y_n$, where $Y_i = 1$ if oocysts were detected in the i-th subject and $Y_i = 0$ otherwise.

   In what follows, show all your work and derivations, not just the final answer.

   (a) As a simple initial model, suppose that the probability of transmission is $p = \mathrm{E}[Y_i], 1 \le i \le n$. Derive the maximum likelihood estimator (MLE) of p and its large sample distribution.

   (b) For a given $\alpha \in (0, 1)$, find either an exact or an approximate $(1 - \alpha)$ confidence interval for p.

   (c) Researchers hypothesize that transmissibility depends on gametocyte density in the patient's blood. Let x_i denote the gametocyte density in the i-th subject, $1 \le i \le n$. Researchers propose the following model for transmissibility,

   $$\mathrm{E}[Y_i] = p_i = \frac{1}{1 + \exp(\beta_0 - \beta_1 x_i)},$$

   where p_i is the subject-specific transmission probability. Assuming β_0 is known, show that the maximum likelihood estimator (MLE) of β_1 satisfies the equation

   $$\sum_{i=1}^{n} x_i(y_i - p_i) = 0,$$

   where y_i is the observed value of Y_i.

   (d) To test the association between transmissibility and gametocyte density, the researchers propose the null hypothesis $H_0 : \beta_1 = 0$ versus the alternative hypothesis $H_1 : \beta_1 \neq 0$. Find an asymptotic test with type-I error probability α.

   (e) The study aims to estimate the gametocyte density threshold θ, defined to be the density with transmission probability equal to 0.95. One biostatistician suggested that, using the MLE in (c), a feasible estimator of θ is $\hat{\theta} = k/\hat{\beta}_1$. Find the constant k such that $\hat{\theta}$ is the MLE of θ. Further, find either an exact or an approximate 95% confidence interval for θ.

   Points: 5 for each part.