MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

Tuesday, August 1, 2017, 9:00 AM - 3:00 PM

1. Suppose that X is distributed as uniform on $(0, 1)$ and, independently, Y is distributed as exponential with mean 1. Define $T := X + Y$.

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

   (a) Given a constant $t \in (0, 1)$ derive an explicit expression for $P(T \leq t)$.

   (b) Given a constant $t \in (1, \infty)$ derive an explicit expression for $P(T \leq t)$.

   (c) Find $\mathrm{E}[T]$, $\mathrm{Var}(T)$ and $\mathrm{Corr}(X, T)$.

   (d) Define $W = 13X - T$. Find $\mathrm{Cov}(T, W)$. Are T and W independent? Justify.

   (e) Find constants a and b such that $\mathrm{E}[a + bT - X] = 0$ and $\mathrm{Var}(a + bT - X)$ is minimized.

   (f) An urn contains 6 balls; 3 red and 3 blue. A "step" is defined as drawing a ball at random from the urn, and replacing it by a ball of the other color (taken from another urn). That is, if the ball drawn is red, it is replaced with a blue ball; if the ball drawn is blue, it is replaced with a red ball. The number of balls in the urn remains equal to 6 after each step. Let the random variable Z_n denote the number of red balls in the urn after n steps (the initial number is $Z_0 = 3$). Prove that $\mathrm{E}[Z_n] = 3$ for all $n \geq 1$. Hint: $\mathrm{E}[Z_{n+1}|Z_n]$.

   Points: (a) 5, (b) 5, (c) 2.5, (d) 2.5, (e) 5, (f) 5.

2. Suppose that the time interval between neuron firings in a certain experiment in the neurophysiology lab follows a Gamma distribution with probability density function

$$f(y|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} \exp(-y/\beta), \quad y > 0, \quad \alpha > 0, \quad \beta > 0.$$

Let $Y_1, \ldots, Y_n$ be an independent sample from that distribution.

(a) Assuming α is known, derive the maximum likelihood estimator (MLE) $\hat{\beta}$ of β, and show that $\hat{\beta}$ is an unbiased estimator of β.

(b) Researchers are interested in estimating the survivor function $S(t) = P(Y > t)$, the probability of no neuron firing before time t. Derive the maximum likelihood estimator (MLE) of $S(t)$, given that $\alpha = 1$.

(c) Let

$$V_1 = \begin{cases} 1, & Y_1 > t \\ 0, & \text{otherwise.} \end{cases}$$

Show that V_1 is an unbiased estimator of $S(t)$.

(d) In this and the remaining parts α is fixed at $\alpha = 1$. Show that the conditional probability density function of Y_1 given $U = \sum_{i=1}^n Y_i$ is

$$f_{Y_1|U}(y_1|u) = \begin{cases} \frac{n-1}{u^{n-1}}(u - y_1)^{n-2}, & 0 < y_1 < u \\ 0, & \text{otherwise.} \end{cases}$$

(e) Show that

$$E(V_1|U) = \left(1 - \frac{t}{U}\right)^{n-1} I(U > t)$$

and that $E(V_1|U)$ is an unbiased estimator of $S(t)$ that has the smallest variance among unbiased estimators, that is, $E(V_1|U)$ is the uniformly minimum variance unbiased estimator (UMVUE). The notation $I(A)$ is for the indicator function; 1 if A is true and 0 if A is false.

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

3. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with cumulative distribution function F. Let $F_n(x)$ be the empirical distribution function defined by

$$F_n(x) = \frac{\text{number of } X_i \leq x}{n}$$

(a) Let $Y_i = I(X_i \leq x)$ with $I(\cdot)$ being an indicator function. The empirical distribution function can be written as $F_n(x) = n^{-1} \sum_{i=1}^{n} Y_i$. Show that $F_n(x)$ is a consistent estimator of $F(x)$ by showing that

$$\lim_{n\to\infty} \mathrm{E}\{F_n(x)\} = F(x),$$

and

$$\lim_{n\to\infty} \mathrm{Var}\{F_n(x)\} = 0,$$

for every real x.

(b) Given a specific $x \in A := \{t : 0 < F(t) < 1\}$, describe the asymptotic distribution of $F_n(x)$ when $n \to \infty$, and derive an approximate 95% confidence interval for $F(x)$ when the sample size n is large.

(c) The median tolerance limit (MTL) is defined as the concentration of a particular toxic substance at which half of the animals survive. To test whether a concentration x is the MTL, one can test the null hypothesis $H_0 : F(x) = 0.5$ versus the alternative hypothesis $H_1 : F(x) \neq 0.5$. Find the likelihood ratio test (LRT) statistic, and find its distribution under H_0 (either its asymptotic distribution or its exact distribution, whichever you find easier).

(d) To test whether the sample is from a null distribution $G(x)$ or alternative distribution $H(x)$, the Neyman-Pearson lemma can be used to find the uniformly most powerful (UMP) test. Assuming that the null distribution $G(x)$ corresponds to the probability density function

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad -\infty < x < \infty,$$

and the alternative distribution $H(x)$ corresponds to the probability density function

$$h(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty,$$

derive the explicit critical region of the UMP test, given a *single observation* X_1 and type-I error probability α.

(e) Derive the statistical power of the decision rule of the UMP test in (d) under the alternative hypothesis. Use the fact that, under the alternative hypothesis, X_1 follows the t distribution with 1 degree of freedom (t_1), and X_1^2 follows the F distribution with degrees of freedom 1 and 1 ($F_{1,1}$).

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