2016 MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART I

July 26, 2016, 9 am – 3 pm

1. A biased coin is being flipped repeatedly, with probability of heads (success) p, $0 < p < 1$, and probability of tails (failure) $q = 1 - p$. Let the random variable X denote the number of failures before the second success. Hence, the pmf of X is

$$P(X = x) = (x + 1)p^2 q^x \qquad x = 0, 1, 2, \cdots .$$

Conditional on $X = x$, a random variable Y has the uniform distribution on the interval $(0, x + 1)$. In the following questions, derive explicit expressions, show your work and simplify the answers as much as possible.

(a) What is the joint density (pdf/pmf) of X and Y? For densities, don't forget to specify the domain (support).

(b) Find $f_Y(y)$, the density of Y.

(c) Find $\mathrm{E}[Y]$.

(d) Find $\mathrm{Cov}(X, Y)$.

(e) Define $T = 2Y - X$. Find $\mathrm{Cov}(T, X)$.

(f) Are T and X independent? Justify your answer.

Points: 25/6 each.

2. Coronary artery calcification can be considered as an early indicator of coronary heart disease, which can lead to heart failure and arrhythmias. Such calcification can be assessed via cardiac computed tomography (CT) scanning with calcium score calculated using a modified Agatston method, with a range from 0 to a very large positive number. Suppose that this score can be transformed into a new random variable X that is distributed as normal with *unknown* mean μ and variance 1. Using $X_1, \ldots, X_n$, a random sample of size n drawn from this distribution, researchers are interested in estimating how probable it is that X excceds 0, that is, in estimating $P(X > 0)$.

You may define $\theta = P(X > 0)$ if you wish. If needed, use the notation $\Phi(t)$ to denote the cumulative distribution function of the standard normal distribution evaluated at t.

(a) Express $P(X > 0)$ as a function of μ.

(b) Find an unbiased estimator of $P(X > 0)$.

(c) Find the maximum likelihood estimator of $P(X > 0)$.

(d) Find the Cramer-Rao lower bound on the variance of unbiased estimators of $P(X > 0)$.

(e) Find the uniformly minimum variance unbiased estimator (UMVUE) of $P(X > 0)$.

Points: 5 each.

3. Suppose that the discrete random variable X takes four possible values, a_1, a_2, a_3, and a_4. However, under different values of parameter θ, the probabilities of observing certain values, denoted by $p(x|\theta)$, are not necessarily identical. The following table lists the distributions $p(x|\theta)$, one per row, under different values of θ, where $\theta = \theta_j$, $j = 1, 2, 3$. Note that $\sum_{i=1}^{4} p(a_i|\theta_j) = 1$ for each $j \in \{1, 2, 3\}$.

	a_1	a_2	a_3	a_4
θ_1	0.3	0.4	0.1	0.2
θ_2	0.4	0.1	0.2	0.3
θ_3	0.2	0.1	0.5	0.2

Suppose that one observation, X, is available (You can think of this as a sample of size 1).

(a) Find the maximum likelihood estimator (MLE) of θ under different values of X (that is, as a function of X).

(b) Derive the critical region of the likelihood ratio test for the null hypothesis $H_0 : \theta = \theta_1$ against $H_1 : \theta \neq \theta_1$, with type-I error probability $\alpha = 0.1$, assuming the parameter space is $\Theta = \{\theta_1, \theta_2, \theta_3\}$.

(c) Give the test function of the likelihood ratio test in (b) in explicit form. Explain explicitly how one would apply the testing procedure using the single observation X.

(d) Find the uniformly most powerful (UMP) test for testing the null hypothesis $H_0 : \theta = \theta_1$ against the alternative hypothesis $H_1 : \theta = \theta_2$, with type-I error probability $\alpha = 0.1$,

(e) Comment on whether the UMP test for the hypothesis in (d) is also the UMP test for the hypothesis in (b). If you think it is, please provide the rationale. If you think it is not, derive the UMP test for the hypothesis in (b).

Points: 5 each.