Statistics comprehensive exam. January 2017

Instructions: *The exam has 6 multi-part problems. All of the problems will be graded. Write your code words on each of your answer sheets. Do not put your name or UNM ID on any of the sheets. Be clear, concise, and complete. All solutions should be rigorously explained.*

1. Skyler selects a random sample of zartylblats and tests each for strength by dropping it on the floor to see if it breaks. Denote by θ the probability that a dropped zartylblat will break; and assume independence for the sample. In the sample of size $n = 10$, Skyler observed $X = 8$ broken and 2 unbroken zartylblats.

   (a) What is the likelihood function for θ in this problem?

   (b) Skyler announces that her posterior distribution for θ has density function

   $$\pi(\theta|X = 8) = c\ \theta^{10}(1-\theta)^6, 0 < \theta < 1$$

   where $c = \Gamma(18)/(\Gamma(11)\Gamma(7)) = 136136$. What was Skyler's prior distribution? Give the answer as either a density function (correctly normalized if possible) or give the name and the value(s) of any parameter(s).

   (c) Blake has a different prior distribution-she believes that θ can take only one of two different values, 1/3 and 2/3, each with prior probability 1/2. With the same data, what is Blake's posterior distribution for θ?

   (d) Alex wants to find the (frequentist) P-value for a test of the hypothesis $H_0 : \theta = 1/2$ against the one-sided alternative $H_1 : \theta < 1/2$, with the same data. Write an expression, in terms of probability, to compute the P-value in this case and with the same data.

   (e) One of the following is the correct numerical value of the P-value (P). Report the correct value and tell how you knew (or how you eliminated the others). If you are unsure, explain. You should not need a calculator or table.

   $a)P = 0.000 \quad b)P = 0.011 \quad c)P = 0.500 \quad d)P = 0.800 \quad e)P = 0.989 \quad f)P = 1.000$

2. (a) Two random scalar quantities x,z have a joint distribution with complete conditionals, $(x|z) \sim N(x|\phi z, v)$ and $(z|x) \sim N(x|\phi x, v)$ for some known positive parameter $\phi \in (-1, 1)$ and where $v = s(1 - \phi^2)$ for some $s > 0$

   i. What are the marginal distribution $p(x)$ and $p(z)$? Show your reasoning.

   ii. What is the joint distribution of (x, z)?

   iii. What is the precision (inverse covariance) matrix of the joint distribution of (x, z)?

3. Suppose we want to study the effect of eduction (in years) X on people's income Y and collect data (X_i, Y_i) from a simple random sample of size n from general population. It is known that the distribution of income is usually heavily right-skewed, thus we assume the following linear model (assume $Y > 0$ and both X and Y are centered):

$$logY_i = X_i\beta + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2).$$

   (a) Assuming σ is unknown, what are the MLEs for β and σ^2? Are they unbiased and why?

   (b) Assuming σ is known, is the MLE of β the uniformly minimum variance unbiased estimator for β? Why or why not?

   (c) Assuming σ is known, give an approximate 95% confidence interval for β when n is large. Describe any theorem that you use to produce your confidence interval.

4. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a $Be(\mu, 1)$ and let $Y_1, ..., Y_m$ be an i.i.d. sample from a $Be(\nu, 1)$ pdf. Assume these samples are independent.

   (a) Find likelihood ratio test for $H_0 : \mu = \nu$ versus $H_A : \mu \neq \nu$.

   (b) Show that the test depends upon the statistic

$$T = \frac{\sum_{i=1}^{n} logX_i}{\sum_{i=1}^{n} logX_i + \sum_{i=1}^{m} logY_i}$$

   (c) Find the distribution of T under the null, and use this to determine a test of size $\alpha = 0.1$.

   (d) Use the asymptotic distribution of a function of the likelihood ratio under the null, to determine a test of size $\alpha = 0.1$

5. Let $\boldsymbol{Y}$ be a $n \times 1$ vector that is normally distributed with mean vector $\boldsymbol{\mu}$ and covariance $\sigma^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Under the null model, $\boldsymbol{\mu} = \mathbf{1}_n\alpha$, where $\mathbf{1}_n$ is a vector of ones of length n. The alternative model has mean vector $\boldsymbol{\mu} = \mathbf{1}_n\alpha + \boldsymbol{X}\beta$ where $\boldsymbol{X}$ is an $n \times p$ matrix of rank $p < n$ that has been centered so that $\boldsymbol{X}^T\mathbf{1}_n = \mathbf{0}_p$.

   (a) If the null model is true, show that

$$R^2 \equiv \frac{\boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}}{||\boldsymbol{Y} - \mathbf{1}_n\bar{\boldsymbol{Y}}||^2}$$

   has a beta distribution and find its expected value ($||\boldsymbol{Y}||^2 = \boldsymbol{Y}^T\boldsymbol{Y}$ and $\bar{\boldsymbol{Y}}$ is the mean of $\boldsymbol{Y}$).

   (b) If p is increasing with n, such that $p/(n-1-p) = r$, with $0 < r < \infty$, what happens to R^2 as $n, p \to \infty$, assuming that the null model is true? Show it.

6. (a) Show that if $\xi \sim N(0,1)$ then $P(|\xi| > z) \leq exp(-z^2/2)$ for any $z > 0$

(b) Let $\{\xi_n\} \sim N(0,1)$ iid be independent standard normal random variables. For what values of $c > 0$ (if any) is it true that only finitely-many of the events

$$A_n = \{|\xi_n| > c\sqrt{log\ n}\}$$

will occur?