DEPARTMENT OF MATHEMATICS AND STATISTICS
UMASS - AMHERST
ADVANCED EXAM - STATISTICS (I)
January 11, 2016

Work all problems. Show all work. Explain your answers. State the theorems used whenever possible. 70 points are required to pass, including at least 35 points from questions 1-3 and 20 points from question 4.

1. Let $F(x, y)$ represent the joint cdf of some two-dimensional random vector (X, Y), and let $F_1(x)$ and $F_2(y)$, respectively, be the marginal cdfs of X and Y. Define functions:

$$U(x, y) = min\{F_1(x), F_2(y)\}$$
$$L(x, y) = max\{F_1(x) + F_2(y) - 1, 0\}.$$

   (a) (7 points) Prove that $L(x, y)$ and $U(x, y)$ are each distribution functions and that their marginal distributions are the same as those of $F(x, y)$.

   (b) (7 points) Prove that $L(x, y) \leq F(x, y) \leq U(x, y)$.

   (c) (7 points) Suppose that X and Y are $N(0, 1)$ random variables, but their joint distribution function is otherwise unknown. Graph (as a function of x) the bounds $L(x, y_0)$ and $U(x, y_0)$ on $F(x, y_0)$ as in (b). Plot this separately for $y_0 = 0$, $y_0 = -1$, and $y_0 = 1$.

2. Consider the problem of conducting a standard hypothesis test concerning the mean of a normal distribution. The data are therefore $X_1, \ldots X_n \sim N(\mu, \sigma^2)$. The null hypothesis is $\mu = \mu_0$.

   (a) (7 points) There are three standard alternative hypotheses often treated in this context. List them here.

   (b) (7 points) For each of the three alternatives in the previous part, give the rejection region for an exact level-α test, in terms of quantiles of known distributions.

   (c) (7 points) On a single plot, sketch the typical power curve for each of the tests in the previous part, plotted against all values of μ on the real line. Be sure to label the three lines, and make clear the relative levels of the three lines for all values of μ. Also be sure to label the type-I error rate.

3. Consider $X_1, X_2, \ldots X_n$ a random sample of size n from the pdf $f(x|\beta) = \frac{2}{\beta^2}(\beta - x)\mathbb{I}(0 < x < \beta)$, where $\mathbb{I}(k)$ is the indicator function on k.

   (a) (7 points) Show that $Z_i = (\beta - X_i)/\beta$ has a distribution that is independent of the parameter β.

   (b) (7 points) Find the MLE of β.

   (c) (7 points) Let $Y_1 \leq Y_2 \cdots \leq Y_n$ be the order statistics corresponding to $Z_1, Z_2, \ldots, Z_n$. Use the information in the previous parts to find an approximate $100(1 - \alpha)\%$ lower confidence bound for β.

4. Suppose $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i, i = 1, \ldots, n, n > p$ and $e_i \sim_{iid} N(0, \sigma^2)$. Let $Y = (Y_1, \ldots, Y_n)'$, X be the n by $p + 1$ design matrix (of rank k), and $\beta = (\beta_0, \ldots, \beta_p)'$. Let y be the observed Y. Let 1_n be a vector of length n with all 1s. Let $\hat{\beta}$ be a least squares estimate of β.

   (a) (7 points) Assume $3\beta_1 - \beta_0$ is estimable. Describe a method to test they hypothesis $H_0 : 3\beta_1 - \beta_0 = 2$ versus the alternative $H_a : 3\beta_1 - \beta_0 \neq 2$

   (b) (7 points) Prove that $X'(y - X\hat{\beta}) = 0$.

   (c) (7 points) Prove that

$$(\overline{y}1_n - y)'(\overline{y}1_n - y) = (\overline{y}1_n - X\hat{\beta})'(\overline{y}1_n - X\hat{\beta}) + (y - X\hat{\beta})'(y - X\hat{\beta}).$$

   (d) (8 points) Prove or disprove the following: $R^2 = 1$ if and only if y is in the column space of X.

   (e) (8 points) Suppose that $k < (p + 1)$. Construct a full rank $\tilde{X}$ and vector θ so that $X\beta = \tilde{X}\theta$.