UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
Advanced Exam I
Friday, January 19, 2018

Please work all problems. Seventy points are required to pass.

1. Transformations and multivariate distributions.

   (a) Let X and Y be random variables. Construct two joint densities for the random vector (X, Y) so that the marginal distributions of X and Y are the same but the joint densities are different.

   (b) Suppose $X_1 \sim Unif(0,1)$ and $X_2 \sim Unif(0,1)$. Let X_1 and X_2 be independent. Find the density of $Z_1 = X_1/X_2$ and its mean.

   (c) Suppose $X|Y \sim N(0, 1/Y)$ and $Y \sim gamma(\alpha, \beta)$. The gamma density is $f(y; \alpha, \beta) \propto y^{\alpha-1}e^{-y/\beta}, y \geq 0$. Find the marginal distribution of X.

2. Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$ where $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p]$. Suppose you want to test $H_0 : \beta_1 = \ldots = \beta_p = 0$. A likelihood ratio statistics is

$$\Lambda(\mathbf{y}) = \frac{max\,\{\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}), \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\}}{max\,\{\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}}$$

   where $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$ is the linear model likelihood, $\boldsymbol{\Theta}_0$ is the parameter space under H_0, and $\boldsymbol{\Theta}$ is the unrestricted parameter space. Let $\mathbf{P}$ be the perpendicular projection matrix for the full model and $\mathbf{P}_0$ be the perpendicular projection matrix for the reduced model.

   (a) Show that $\Lambda(\mathbf{y})$ is a ratio of estimated variances.

   (b) The rejection region is $\{\Lambda(\mathbf{y}) < c\}$ for some c. Describe in general how you find c to create a level α test.

   (c) Given a c, show that there exists an f so that the rejection regions $\{\Lambda(\mathbf{y}) < c\}$ and $\{F(\mathbf{y}) < f\}$ are equivalent where $F(\mathbf{y}) = \frac{\mathbf{y}'(\mathbf{P}-\mathbf{P}_0)\mathbf{y}/p}{\mathbf{y}'(\mathbf{I}-\mathbf{P})\mathbf{y}/(n-p)}$.

3. Consider a segmented simple linear regression problem in one variable, x. In particular, suppose that $n = 6$ values of a response y are related to values $x = 0, 1, 2, 3, 4, 5$ by a Gauss-Markov normal linear model $Y = X\beta + \epsilon$ for

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \\ 1 & 4 & 2 \\ 1 & 5 & 3 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad \text{and } \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix}$$

   Values of x are in the second column of the model matrix. This model allows the linear from $y \approx \beta_0 + \beta_1 x$ for $x \leq 2$ and the linear from $y \approx \beta_0 + 2\beta_1 + (\beta_1 + \beta_2)(x - 2)$ for $x \geq 2$. Notice that there is continuity of these forms at $x = 2$.

(a) This is a full rank model. Argue carefully that this is the case.

Here $(X'X)^{-1} = \begin{pmatrix} .825 & -.474 & .526 \\ -.474 & .421 & -.579 \\ .526 & -.579 & .921 \end{pmatrix}$ and for $Y' = (0, 2, 4, 3, 1, 0)$, $(X'X)^{-1}X'Y = \begin{pmatrix} -.018 \\ 2.053 \\ -3.447 \end{pmatrix}$ and $SSE = .202$.

(b) Is there definitive evidence that a simpler model $y = \beta_0 + \beta_1 x \quad \forall x$ is inadequate here? Explain.

(c) Tomorrow a total of 3 new observations are to be drawn from this model at, respectively, $x = 1, 2$, and 3. Call these y_1^*, y_2^*, and y_3^*. The quantity $(y_3^* - y_2^*) - (y_2^* - y_1^*) = y_3^* - 2y_2^* + y_1^*$ is an empirical measure of change in slope of mean y as a function of x at $x = 2$ based on these new observations. Provide 95% two-sided prediction limits for this quantity. (Plug in completely, but you need not do arithmetic.)

(d) Find the value and degrees of freedom for a t statistic for testing $H_0 : \mu_{y|x=1} = \mu_{y|x=5}$ (the hypothesis that the mean responses are the same for $x = 1$ and $x = 5$).

(e) Write out (plug in completely so that your implied answer is numerical, but you need not do the arithmetic) a test statistic that you could use to test the hypothesis that $H_0 : \mu_{y|x=1} = \mu_{y|x=5} = 1$. Say exactly what null distribution you would use.

(f) It is possible to compute $X(X'X)^{-1}X'$ both for the full model specified at the beginning of this problem and for a model with X matrix consisting of only the first two columns of the original one. The diagonal entries of these two matrices are in the table below. Compare the two patterns above and say why (in the

i	1	2	3	4	5	6
x_i	0	1	2	3	4	5
diagonal entry of $X(X'X)^{-1}X'$ for the original X	.825	.298	.614	.272	.298	.693
diagonal entry of $X(X'X)^{-1}X'$ for the reduced X	.524	.295	.181	.181	.295	.524

context provided at the beginning of this problem) they "make sense".