STA244 **3/05/2003**

Midterm Exam

Please provide concise neatly written solutions. Cross out any part of a solution that you do not want to be graded.

1. Assume that we have a sample of size n, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ from the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{1}$$

   with $\epsilon \sim N(0, I_n/\phi)$ and $\mathbf{X}$ is $n \times p$ of rank p. Find the distribution of $\mathbf{e}$, where $\mathbf{e}$ is the vector of ordinary residuals, $\mathbf{Y} - \hat{\mathbf{Y}}$, and $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\beta}}$ is the usual maximum likelihood (MLE)/ordinary least squares (OLS) estimate of $\boldsymbol{\beta}$.

2. After updating our prior distributions for $\boldsymbol{\beta}, \phi$ based on the model/data in equation (1), assume we have the following posterior distributions

$$\boldsymbol{\beta} | \phi, Y \sim N(\hat{\boldsymbol{\beta}}, (X'X)^{-1}/\phi) \tag{2}$$

$$\phi | Y \sim \text{Gamma}(\frac{n-p}{2}, \frac{\hat{\sigma}^2(n-p)}{2}) \tag{3}$$

   where $\hat{\sigma}^2 = MSE$ is the usual estimate of σ^2 from OLS regression. Find the posterior distribution of $\epsilon | Y$ assuming (1) is correct. Contrast with results in (1).

3. Suppose we have data on length increments I_{ij} for the ith fish at age j, and that increments are normally distributed

$$I_{ij} \sim N(\mu_j, \sigma^2)$$

   for $j = 1, 2, \ldots a_i$, $i = 1, \ldots, n$. For the fish i at age a_i when captured, let $L_i = \sum_{j=1}^{a_i} I_{ij}$ be the total length at age a_i (age of capture).

   (a) Find the distribution of $L_i | a_i$.

   (b) Suppose that to a reasonable approximation, we can model

$$\sum_{j=1}^{a_i} \mu_{ij} = \beta_0 + \beta_1 a_i + \beta_2 a_i^2$$

   Given this assumption, what is the MLE of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$? Will these be the same as the ordinary least squares estimates? (explain)

4. Suppose that the conditional distribution of $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ given $\mathbf{X}_1$ and $\mathbf{X}_2$ has mean

$$\mathsf{E}[\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2] = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 \tag{4}$$

and constant variance,

$$\mathsf{Cov}[\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2] = \sigma^2 I_n \tag{5}$$

where I_n is a $n \times n$ identity matrix, $\mathbf{X}_1$ is a full rank $n \times p$ matrix and $\mathbf{X}_2$ is a full rank $n \times q$ matrix of predictor variables, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p and q dimensional vectors of regression coefficients, respectively.

(a) Suppose the true mean function is given by (4), but we want to fit the mean as a function of $\mathbf{X}_1$ alone. Find $\mathsf{E}[\mathbf{Y}|\mathbf{X}_1]$ assuming that $\mathbf{X}_2|\mathbf{X}_1$ has a distribution with finite mean, $\mathsf{E}[\mathbf{X}_2|\mathbf{X}_1] = g(\mathbf{X}_1)$, and variance, $\mathsf{V}[\mathbf{X}_2|\mathbf{X}_1] = v(\mathbf{X}_1)$. (hint: recall iterated expectation)

(b) Find $\mathsf{Cov}[\mathbf{Y}|\mathbf{X}_1]$. Give sufficient conditions for this to be constant for all values of $\mathbf{X}_1$.

(c) If we fit a regression of $\mathbf{Y}$ on $\mathbf{X}_1$ of the form $\mathbf{X}_1\boldsymbol{\alpha}$ using ordinary least squares (OLS), what is the expected value of $\hat{\boldsymbol{\alpha}}$ (the OLS estimator), when the mean is actually given by $\mathsf{E}[\mathbf{Y}|\mathbf{X}_1]$?

(d) If the distibution of $\mathbf{Y}$ given $\mathbf{X}_1$ and $\mathbf{X}_2$ has mean and variance given by (4) and (5), respectively, is the OLS estimator $\hat{\boldsymbol{\alpha}}$ from above the best linear unbiased estimator of $\boldsymbol{\beta}_1$? Explain.