Topics & Sample Problems for Stat 700 In-Class Test, Fall 2008

(I) [*Multivariate normal and transformations*]

In the theory of simple linear regression, data $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ satisfy

$$Y_i = a + bX_i + \epsilon_i \quad , \qquad \epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

where X_i are treated as known constants and $\vartheta = (a, b, \sigma^2) \in \mathbf{R}^2 \times \mathbf{R}^+$. The standard least-squares estimators of (a, b) and method-of-moments estimator $\hat{\sigma^2}$ can be shown to have the form under $\vartheta = \vartheta_0 = (a_0, 0, \sigma_0^2)$:

$$\hat{b} \; = \; \frac{1}{s_X \sqrt{n-1}} \sum_{j=1}^n c_j \, \epsilon_j, \quad c_j = \frac{X_j - \bar{X}}{(\sum_{k=1}^n (X_k - \bar{X})^2)^{1/2}} \;\; , \quad \sum_{j=1}^n c_j^2 \; = \; 1$$

$$\hat{\sigma^2} \; = \; \frac{1}{n-2} \; \epsilon' \Big(I \, - \, \frac{1}{n} \mathbf{1}\mathbf{1}' \, - \, \mathbf{c}\mathbf{c}' \Big) \, \epsilon$$

where I is the $n \times n$ identity matrix, $\mathbf{1}$ the n-vector with all entries 1, and

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad s_X^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 \;\; , \quad \epsilon \; = \; \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \qquad \mathbf{c} \; = \; \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

(a) Show that $\mathbf{M} = I - \frac{1}{n} \mathbf{1}\mathbf{1}' - \mathbf{c}\mathbf{c}'$ is a projection matrix of rank $n - 2$.

(b) Prove that $(n-2) \, \hat{\sigma}^2 / \sigma_0^2 \sim \chi^2_{n-2}$.

(c) Prove that $\hat{b}$ is independent of $\hat{\sigma}^2$.

(II) [*Identifiability & sufficiency*]

Suppose that a data vector $\mathbf{X} = (X_1, \ldots, X_n)$ is observed, with components *iid* with $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta = (0,1)^2$ and marginal density

$$f(x, \vartheta) \; = \; \frac{\vartheta_2}{\vartheta_1} \, I_{[0 < x \le \vartheta_1]} \, + \, \frac{1 - \vartheta_2}{1 - \vartheta_1} \, I_{[\vartheta_1 < x < 1]}$$

(i) Show that the model is **not** identifiable on Θ, but **is** identifiable on $\Theta^* \; = \; \{\vartheta \in (0,1)^2 \; : \; \vartheta_2 > \vartheta_1\}$.

(ii) Show that the minimal sufficient statistic for $\vartheta \in \Theta^*$ is the set of order statistics $(X_{(1)}, \ldots, X_{(n)})$.

(III) [*Bayes & Decision Theory*]

Consider the model with parameter $\vartheta \in \Theta = (0,1)$, data sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the density

$$f(x, \vartheta) \;=\; \vartheta \, I_{[0<x\leq 1]} \;+\; (1-\vartheta) \, I_{[1<x\leq 2]}$$

and prior density $\pi(\theta) \;=\; 3\theta^2$ on $(0,1)$.

(i) Find the posterior density of θ given $\mathbf{X}$. *Hint: use the sufficient statistic !*

(ii) Under the loss function $L(\vartheta, a) \;=\; (\vartheta - a)^2/\vartheta$, find the Bayes rule (i.e., the π-Bayes optimal estimator $d(\mathbf{X})$).

(IV) [*UMVUE's, exponential families*]

Explain why there is a UMVUE for $\mu\sigma^2$ based on a $\mathcal{N}(\mu, \sigma^2)$ data sample $\mathbf{X} = (X_1, \ldots, X_n)$ with $n \geq 3$, and show how to determine it as a ratio of multiple integrals, but you need not find the explicit form of the estimator. *Hint: find* $E(X_1(X_2 - X_3)^2)$.