MA 575 – Qualifying Exam

Spring 2010

1. Given the dataset $\{\mathbf{x}_i, y_i\}_{i=1}^n$, we wish to obtain a *bootstrapped* least-squares estimate for β under the model

$$\mathbf{y} = X\beta + \mathbf{e}, \quad \mathrm{E}[\mathbf{e}|X] = 0, \quad \mathrm{Var}[\mathbf{e}|X] = \sigma^2 I_n, \tag{1}$$

with $X = [\mathbf{x}_1^T \ \cdots \ \mathbf{x}_n^T]^T$.

If $\mathbf{u}_i \doteq (\delta_{ij})_{j=1}^n$ is a $1 \times n$ vector with zero entries apart from $u_{ii} = 1$, then we can "select" the i-th row of X by $\mathbf{x}_i = \mathbf{u}_i X$, or the i-th entry in $\mathbf{y}$ by $y_i = \mathbf{u}_i \mathbf{y}$. Thus, if $\tilde{X}$ and $\tilde{\mathbf{y}}$ are the bootstrap samples of X and $\mathbf{y}$ respectively, we can represent the bootstrap sample by a matrix B with rows that are similar to $\mathbf{u}_i$, but with i being sampled, and such that $\tilde{X} = BX$ and $\tilde{\mathbf{y}} = By$.

(a) Show that the bootstrapped LSE for β, say $\tilde{\beta}(B)$, is a *weighted* least-squares estimate. More specifically, show that $\tilde{\beta}(B)$ is the least-squares estimator for β in

$$\mathbf{y} = X\beta + \mathbf{e}, \quad \mathrm{E}[\mathbf{e}|X, B] = 0, \quad \mathrm{Var}[\mathbf{e}|X, B] = \sigma^2 W(B)^{-1},$$

where the weight matrix $W(B)^{-1}$ is a function of the bootstrap sample.

(b) Provide an expression for $w(B) \doteq \mathtt{Diag}(W(B)^{-1})$ and argue that

$$w(B)_i \sim \mathtt{Binomial}(n, 1/n), \quad i = 1, \ldots, n.$$

What does $w(B)_i$ represent?

(c) Since $\tilde{\beta}(B)$ depends on $w(B)$, let us denote $w(B)$ as w and set $\tilde{\beta}(w) \doteq \tilde{\beta}(B)$. Now regard $\tilde{\beta}(w)$ as a function of w and use the delta method with a second order approximation around $\mathrm{E}[w]$ to conclude that $\mathrm{E}_w[\tilde{\beta}(w)] \approx \hat{\beta}$, where $\hat{\beta}$ is the least-squares estimator for β in (1).

(d) Argue that with *positive* probability $\tilde{\beta}(w)$ might not exist. What can go wrong? Be as specific as possible.

2. In *regularized* regression on p predictors we seek an estimator $\hat{\beta}_R(\Lambda)$ in the model

$$\mathbf{y} = X\beta + \mathbf{e}, \quad \mathrm{E}[\mathbf{e}|X] = 0, \quad \mathrm{Var}[\mathbf{e}|X] = \sigma^2 I_n, \tag{2}$$

that minimizes

$$SS_\Lambda(\mathbf{b}) = (\mathbf{y} - X\mathbf{b})^T(\mathbf{y} - X\mathbf{b}) + \mathbf{b}^T\Lambda^T\Lambda\mathbf{b}.$$

(a) Show that the least-squares estimator for β in the *augmented* model (with artificial observations)

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} X \\ \Lambda \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix}$$

is $\hat{\beta}_R(\Lambda) \doteq \arg\min_{\mathbf{b}} SS_\Lambda(\mathbf{b})$ and give an expression for it.

Assume that $X \doteq [\mathbf{x}_1 \;\cdots\; \mathbf{x}_p]$ is *orthogonal* and that we now focus on one regularizing parameter λ and set $\Lambda = \sqrt{\lambda} I_p$. Denote $\hat{\beta}(\lambda) \doteq \hat{\beta}(\Lambda)$.

(b) If $\hat{\beta}$ is the least-squares estimate for model (2), show that $(\hat{\beta}_R(\lambda))_i = s_i(\lambda)\hat{\beta}_i$ for $i = 1, \ldots, p$ and provide an expression for $s_i(\lambda)$, the i-th *shrinkage factor.*

(c) Define the hat matrix $H(\lambda)$ with respect to $\hat{\beta}_R(\lambda)$ in model (2) and show that $H(\lambda)$ is symmetric but *not*, in general, idempotent.

(d) Find the *effective degrees of freedom*, defined as the trace of $H(\lambda)$, $df(\lambda) \doteq \mathtt{tr}(H(\lambda))$. Verify that $df(0) = p$ (no regularization) and describe the behavior of $df(\lambda)$ as $\lambda \to \infty$. (Hint: recall that $\mathtt{tr}(AB) = \mathtt{tr}(BA)$.)

(e) If $\mathbf{r}_i$ is the i-th *row* of X, define

$$C(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \mathbf{r}_i^T\hat{\beta}_R(\lambda)}{1 - H(\lambda)_{ii}}\right)^2,$$

where $H(\lambda)_{ii}$ is the i-th entry in the diagonal of $H(\lambda)$. How would you select λ based on $C(\lambda)$? Explain your answer in light of an interpretation for $C(\lambda)$.