Statistics 600 Midterm Exam

November 15, 2007

1. Suppose the data generating model is

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

where the usual assumptions $E(\epsilon|X) = 0$ and $\mathrm{cov}(\epsilon|X) \propto I$ hold. Additionally, assume the X_i are iid standard normal.

Let $Z_i = 1$ if $X_i > 0$ and $Z_i = -1$ if $X_i \le 0$, and suppose we fit a working model

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} Z_i$$

using ordinary least squares. What are the limiting values of $\hat{\beta}$ and $\hat{\alpha}$?

Solution:

$$\begin{aligned}
\hat{\beta} &= \frac{n^{-1}\sum Y_i(Z_i - \bar{Z})}{n^{-1}\sum(Z_i - \bar{Z})^2} \\
&= \beta\frac{n^{-1}\sum X_i(Z_i - \bar{Z})}{n^{-1}\sum(Z_i - \bar{Z})^2} + \frac{n^{-1}\sum \epsilon_i(Z_i - \bar{Z})}{n^{-1}\sum(Z_i - \bar{Z})^2} \\
&\to \beta E|X| \\
&= \beta\sqrt{2/\pi}\int_0^\infty x\exp(-x^2/2)dx \\
&= \beta\sqrt{2/\pi}
\end{aligned}$$

$$\begin{aligned}
\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{Z} \\
&\to \alpha
\end{aligned}$$

2. Suppose the data generating model is

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

where the usual assumptions $E(\epsilon|X) = 0$ and $\mathrm{cov}(\epsilon|X) \propto I$ hold. Let D be the design matrix, and assume that

$$n^{-1}D'D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & r \\ 0 & r & 1 \end{pmatrix}.$$

Suppose we aim to estimate the quantity

$$A = c_1\beta_1 + c_2\beta_2$$

where c_1 and c_2 are constants such that $c_1^2 + c_2^2 = 1$. We will use $\hat{A} = c_1\hat{\beta}_1 + c_2\hat{\beta}_2$ as our estimator.

When $r \neq 0$, determine which values of c_1, c_2 give the largest, and the smallest values for $\mathrm{var}\,\hat{A}$. Also state what happens when $r = 0$.

Solution: Letting $\hat{\beta}$ be the OLS estimate of β, we get

$$\mathrm{cov}\,\hat{\beta} = n^{-1}\sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/(1-r^2) & -r/(1-r^2) \\ 0 & -r/(1-r^2) & 1/(1-r^2) \end{pmatrix}.$$

Let $(c_1, c_2) = (\cos(\theta), \sin(\theta))$, so that

$$\mathrm{var}\,\hat{A} = \frac{\sigma^2}{n(1-r^2)}\left(1 - 2r\cos(\theta)\sin(\theta)\right).$$

When $r = 0$, the variance of $\hat{A}$ does not depend on the values of c_1 and c_2. When $r \neq 0$, the extreme values of the variance occur at the extreme values of $r\cos(\theta)\sin(\theta)$. These occur at $\theta = \pi/4, 3\pi/4, 5\pi/4$, and $7\pi/4$, or, equivalently, when (c_1, c_2) are $(1/\sqrt{2}, 1/\sqrt{2})$, $(-1/\sqrt{2}, 1/\sqrt{2})$, $(-1/\sqrt{2}, -1/\sqrt{2})$, and $(1/\sqrt{2}, -1/\sqrt{2})$.

The second derivative of $\mathrm{var}\,\hat{A}$ has the same sign as $r\cos(\theta)\sin(\theta)$.

When $r > 0$ the variance is minimized at $(1/\sqrt{2}, 1/\sqrt{2})$ and $(-1/\sqrt{2}, -1/\sqrt{2})$ and the variance is maximized at $(1/\sqrt{2}, -1/\sqrt{2})$ and $(-1/\sqrt{2}, 1/\sqrt{2})$.

When $r < 0$ the variance is minimized at $(1/\sqrt{2}, -1/\sqrt{2})$ and $(-1/\sqrt{2}, 1/\sqrt{2})$ and the variance is maximized at $(1/\sqrt{2}, 1/\sqrt{2})$ and $(-1/\sqrt{2}, -1/\sqrt{2})$.

3. Suppose M is a symmetric $n \times n$ matrix such that $M \cdot 1_n = 1_n$, where 1_n is an n-dimensional column vector of 1's.

We observe data from a linear model

$$Y = X\beta + \epsilon,$$

where X is a $n \times p + 1$ design matrix with first column identically 1, and $E(\epsilon|X) = 0$. Let $\hat{Y}_M = MY$, and let $R_M = Y - \hat{Y}_M$ be the residuals. Suppose we are interested in the following two properties of $\hat{Y}_M$ as an estimator of EY.

(a) $\hat{Y}_M$ is unbiased for EY.

(b) R_M has zero sample correlation with each column of X.

Show that (b) implies (a), and discuss why (a) does not imply (b).

Solution:

For $\hat{Y}_M$ to be unbiased we need that

$$EMY = MX\beta = X\beta,$$

so $(M - I)X\beta = 0$.

For R_M to be uncorrelated with each column of X, note that orthogonality is sufficient since $1_n' R_M = 0$. Thus we need that

$$X' R_M = X'(I - M)Y \equiv 0.$$

Since Y is a random vector that can take on any value, the sample correlation will only be identically zero if $X'(I - M) = 0$, or equivalently $(M' - I)X = 0$. Since M is symmetric, this implies that $(M - I)X = 0$, and hence $(M - I)X\beta = 0$, so $\hat{Y}_M$ is unbiased.

On the other hand, if we randomly select half our data and use OLS to estimate EY, the result is still unbiased. However it will not be true in that case that the residuals are uncorrelated with the covariates.

4. Suppose we observe data from the model

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i,$$

where $E(\epsilon|X) = 0$ and $\mathrm{var}(\epsilon|X) \propto I$. Using OLS, we fit the model

$$\hat{Y}_i = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j Z_{ij},$$

where $Z_{ij} = c_j X_{ij}$. Derive expressions for the mean and covariance matrix of the $\hat{\gamma}_j$. How do the standardized coefficient estimates (each coefficient estimate divided by its estimated standard deviation) compare to the standardized coefficient estimates for the OLS fit

$$\hat{Y}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j X_{ij}?$$

Solution:

Let $F = \mathrm{diag}(1, c_1, \ldots, c_p)$, i.e. a $p+1 \times p+1$ diagonal matrix with $1, c_1, \ldots, c_p$ on the diagonal. Let Z and X denote the $n \times p+1$ design matrices containing the Z_j and the X_j variables, respectively. Note that $Z = XF$. Therefore,

$$\begin{aligned}
\hat{\gamma} &= (Z'Z)^{-1}Z'Y \\
&= F^{-1}(X'X)^{-1}X'Y \\
&= F^{-1}\hat{\beta}.
\end{aligned}$$

It follows that

$$\mathrm{cov}\ \hat{\gamma} = \sigma^2 F^{-1}(X'X)^{-1}F^{-T},$$

and in particular,

$$\mathrm{var}\ \hat{\gamma}_{jj} = \sigma^2 (X'X)^{-1}_{jj} / F^2_{jj}.$$

By a direct calculation, it follows that the standardized coefficient estimates are identical:

$$\hat{\gamma}_j / \sqrt{\sigma^2 (Z'Z)^{-1}_{jj}} = \hat{\beta}_j / \sqrt{\sigma^2 (X'X)^{-1}_{jj}}.$$

5. Suppose we observe data from a population in which

$$\begin{array}{lcr} E(Y|X=-1) & = & -1 \\ E(Y|X=1) & = & 1 \\ E(Y|X=0) & = & \theta \end{array}$$

We observe $n/2$ cases where $X=-1$ and $n/2$ cases where $X=1$ (no cases where $X=0$ are available). Assume that $\mathrm{var}(Y|X)=\sigma^2$ for $X=-1,0,1$, and the Y_i are uncorrelated given X.

Based on the observed data of sample size n, we fit a working model

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X,$$

and based on this fitted model we construct a 95% prediction interval at $X=0$. Derive an approximate expression for the actual coverage probability of this (nominal) 95% prediction interval.

Solution: Let Y_0^* be an observation at $X=0$ that is independent of the training data, and let $\hat{\alpha},\hat{\beta}$ be the coefficient estimates based on the training data.

Based on the estimate

$$\widehat{\mathrm{SD}}(Y_0^* - \hat{\alpha}) \approx \sigma\sqrt{(n+1)/n},$$

we get the following nominal interval

$$\begin{array}{rcl} 0.95 & = & P(-Q \le (Y_0^* - \hat{\alpha})/\widehat{\mathrm{SD}}(Y_0^* - \hat{\alpha}) \le Q) \\ & = & P\left(\hat{\alpha} - Q\hat{\sigma}\sqrt{(n+1)/n} \le Y_0^* \le \hat{\alpha} + Q\hat{\sigma}\sqrt{(n+1)/n}\right). \end{array}$$

The actual coverage of this interval is not 0.95, since $E(Y_0^* - \hat{\alpha}) \ne 0$. To approximate the actual coverage, recenter to get

$$P\left(-Q - \theta/\widehat{\mathrm{SD}}(Y_0^* - \hat{\alpha}) \le (Y_0^* - \theta - \hat{\alpha})/\widehat{\mathrm{SD}}(Y_0^* - \hat{\alpha}) \le Q - \theta/\widehat{\mathrm{SD}}(Y_0^* - \hat{\alpha})\right).$$

Thus the actual coverage probability is approximately

$$F(Q - \theta/\sigma\sqrt{(n+1)/n}) - F(-Q - \theta/\sigma\sqrt{(n+1)/n}),$$

where F is the standard normal CDF.

6. Recall that the Box-Cox transform is

$$Y_i^{(\lambda)} \equiv \frac{Y_i^\lambda - 1}{\lambda},$$

and the procedure is to identify the value of λ that maximizes

$$-\frac{n}{2}\log\hat{\sigma}_\lambda^2 + (\lambda - 1)\sum_i \log Y_i,$$

where

$$\hat{\sigma}_\lambda^2 = \min_\beta n^{-1}\|Y^{(\lambda)} - X\beta\|^2.$$

Let $Z_i = cY_i$, where $c > 0$ is a constant. Determine how the optimal Box-Cox parameter for the Y_i is related to the optimal Box-Cox parameter for the Z_i.

Solution:

$$Z_i^{(\lambda)} = \frac{c^\lambda Y_i^\lambda - 1}{\lambda}.$$

Since the additive constant $-1/\lambda$ doesn't affect the residuals,

$$\hat{\sigma}_{z,\lambda}^2 = c^{2\lambda}\hat{\sigma}_\lambda^2.$$

Therefore the Box-Cox criterion for the Z_i is

$$\begin{aligned} -\frac{n}{2}\log\hat{\sigma}_{\lambda,z}^2 + (\lambda - 1)\sum_i \log Z_i &= -\frac{n}{2}\log(c^{2\lambda}) - \frac{n}{2}\log\hat{\sigma}_\lambda^2 + n(\lambda - 1)\log c + (\lambda - 1)\sum_i \log Y_i \\ &= -\frac{n}{2}\log\hat{\sigma}_\lambda^2 - n\log(c) + (\lambda - 1)\sum_i \log Y_i. \end{aligned}$$

Since this differs from the Box-Cox criterion for the Y_i by a value that doesn't depend on λ, we can conclude the the optimal Box-Cox parameters are the same.