ST790 HW2

Steven Xu

September 17, 2019

1.

(i)

The 95% confidence band is created by connecting the point-wise confidence interval

$$f(\mathbf{x}_0^T\hat{\boldsymbol{\beta}}) \pm z_{0.025}\sqrt{\mathbf{x}_0\text{Var}(\hat{\boldsymbol{\beta}})\mathbf{x}_0^T}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, $\text{Var}(\hat{\beta}) = \hat{\sigma}(\mathbf{X}^T\mathbf{X})^{-1}$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2}{n-p-1}$.

(ii)

Algorithm 1: Point-wise confidence interval for $\mathbf{x}_0^T\boldsymbol{\beta}$

```
B = ∅;
b = 0;
for i = 1:M do
    β*_i ~Normal(β̂, σ̂(X^T X)^{-1});
    if (β̂ − β*_i)^T (X^T X)(β̂ − β*_i) ≤ σ̂_2 χ²_{p+1,0.95} then
        B ⟵ {B, f(x_0^T β*_i)};
        b ⟵ b + 1;
    else
        continue;
    end
end
```

The point-wise confidence interval for $\mathbf{x}_0^T\boldsymbol{\beta}$ can then be approximated using endpoints $\mathbf{B}_{(1)}$ and $\mathbf{B}_{(b)}$ where $\mathbf{B}_{(i)}$ is the $i-$th order statistic of $\mathbf{B}$. The 95% confidence band is again created by connecting the lower/upper end points for the point-wise confidence interval.

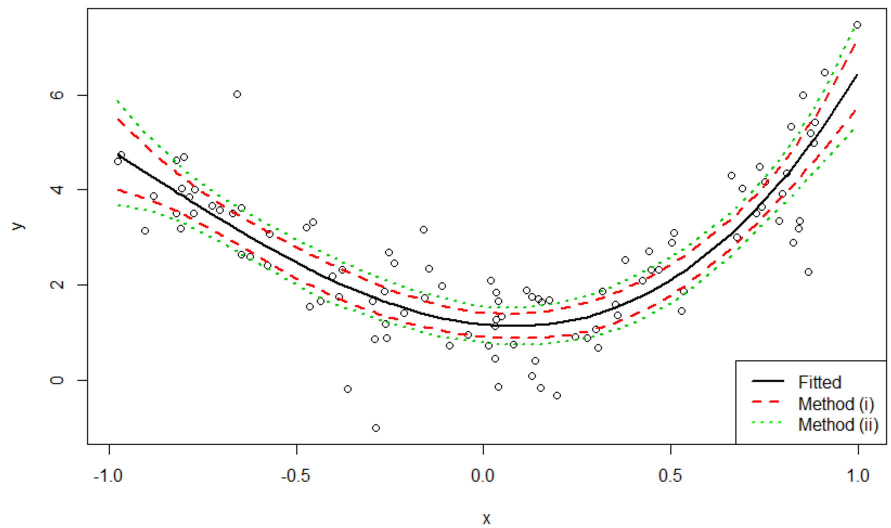

Figure 1: 95% confidence band for $f(\mathbf{x}^T\boldsymbol{\beta})$

Based on the result the second method seems to provide a wider band.

2.

(i)

We first prove that $h(x) \perp [A - p(x)]g(x)$, assuming $x \sim f(x)$.

$$\begin{aligned}
\mathbb{E}\{h(x)[A-p(x)]g(x)\} &= \int h(x)[A-p(x)]g(x)f(x)dx \\
\text{Let } u(x) &= h(x)g(x) \\
&= \int u(x) \int [A-p(x)]f(x)dx\ dx - \int u'(x) \int [A-p(x)]f(x)dx\ dx \\
\because\ p(x) &= P(A=1|x) = \mathbb{E}(A|x) \\
\therefore\ \int [A-p(x)]f(x)dx &= \mathbb{E}[A-p(x)] = \mathbb{E}(A) - \mathbb{E}(A) = 0 \\
\therefore\ \mathbb{E}\{h(x)[A-p(x)]g(x)\} &= 0
\end{aligned}$$

Therefore $h(x) \perp [A-p(x)]g(x)$. Notice that this result does not depend on the form of $h(\cdot)$ or $g(\cdot)$. Additionally, we have

$$\mathbb{E}\left\{\mathbf{H}(\mathbf{X})\text{diag}(\mathbf{A}-p(\mathbf{X}))\mathbf{G}(\mathbf{X})^T\right\} = \mathbb{E}\left\{\sum_{i=1}^n [A_i - p(X_i)]\mathbf{h}(\mathbf{X})_i \mathbf{g}(\mathbf{X})_i^T\right\} = \mathbf{0}$$

for any mapping $\mathbf{H}(\cdot)$, $\mathbf{G}(\cdot)$.

$$\begin{aligned}
\text{Let }\ \mathcal{S}(\boldsymbol{\gamma},\boldsymbol{\beta}) &= \sum_{i=1}^n \left[Y_i - \tilde{\mathbf{X}}_i^T\boldsymbol{\gamma} - \{A_i - p(\mathbf{X}_i)\}(\tilde{\mathbf{X}}_i^T\boldsymbol{\beta})\right]^2 \\
\frac{\partial \mathcal{S}(\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma}} &= -2\sum_{i=1}^n \left[Y_i - \tilde{\mathbf{X}}_i^T\boldsymbol{\gamma} - \{A_i - p(\mathbf{X}_i)\}(\tilde{\mathbf{X}}_i^T\boldsymbol{\beta})\right]\tilde{\mathbf{X}}_i \\
&= -2\left[\tilde{\mathbf{X}}^T\mathbf{Y} - \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\boldsymbol{\gamma} - \tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right] \overset{set}{=} 0
\end{aligned}$$

$$
\begin{aligned}
&\text{Let us assume } \tilde{\mathbf{X}}^T\tilde{\mathbf{X}} \succ 0\\
\because\ &\frac{\partial^2 \mathcal{S}(\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma}^2} = 2\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} \succ 0\\
\therefore\ &\hat{\boldsymbol{\gamma}} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\left[\tilde{\mathbf{X}}^T\mathbf{Y} - \tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right]\\
\mathcal{S}(\hat{\boldsymbol{\gamma}},\boldsymbol{\beta}) &= \left[\mathbf{Y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}} - \text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right]^T\left[\mathbf{Y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}} - \text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right]\\
&= \left[(\mathbf{I}-P_{\tilde{\mathbf{X}}})\mathbf{Y} - (\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right]^T\left[(\mathbf{I}-P_{\tilde{\mathbf{X}}})\mathbf{Y} - (\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right]\\
\frac{\partial \mathcal{S}(\hat{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= -2\left[\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\mathbf{Y} - \tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}\right]\\
&\overset{set}{=} 0\\
&\text{Let us assume } \tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}} \succ 0\\
\because\ &\frac{\partial^2 \mathcal{S}(\hat{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = 2\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}} \succ 0\\
\therefore\ &\hat{\boldsymbol{\beta}} = \left[\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}}\right]^{-1}\left[\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\mathbf{Y}\right]\\
&\text{Let } (\mathbf{I}-P_{\tilde{\mathbf{X}}})\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\tilde{\mathbf{X}} = \bar{\mathbf{X}}\\
\therefore\ &\hat{\boldsymbol{\beta}} = \left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right]^{-1}\left[\bar{\mathbf{X}}^T\mathbf{Y}\right]
\end{aligned}
$$

$$
\begin{aligned}
\bar{\mathbf{X}}^T\mathbf{Y} &= \bar{\mathbf{X}}^T\left[h(\mathbf{X}) + \text{diag}\{\mathbf{A}\}\tilde{\mathbf{X}}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}\right]\\
&= \bar{\mathbf{X}}^T\left[h(\mathbf{X}) + \text{diag}\{\mathbf{A}-p(\mathbf{X})+p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}\right]\\
&= \bar{\mathbf{X}}^T\left[h(\mathbf{X}) + \bar{\mathbf{X}}\boldsymbol{\beta}_0 + \text{diag}\{p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}\right]\\
\therefore\ \hat{\boldsymbol{\beta}} &= \left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right]^{-1}\bar{\mathbf{X}}^T\left[h(\mathbf{X}) + \text{diag}\{p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}\right] + \boldsymbol{\beta}_0\\
&= \left[\frac{1}{n}\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right]^{-1}\frac{1}{n}\bar{\mathbf{X}}^T\left[h(\mathbf{X}) + \text{diag}\{p(\mathbf{X})\}\tilde{\mathbf{X}}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}\right] + \boldsymbol{\beta}_0\\
&\text{Further assume } \frac{1}{n}\bar{\mathbf{X}}^T\bar{\mathbf{X}} \xrightarrow{P} \mathbf{C},\ \text{where } \mathbf{C} \text{ is a positive-definite matrix}\\
\therefore\ &\left[\frac{1}{n}\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right]^{-1} \xrightarrow{P} \mathbf{C}^{-1}\\
&\frac{1}{n}\bar{\mathbf{X}}^T h(\mathbf{X}) = \frac{1}{n}\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})h(\mathbf{X})
\end{aligned}
$$

Although $(\mathbf{I}-P_{\tilde{\mathbf{X}}})h(\mathbf{X})$ is a vector, we can treat it as the first column of a matrix with rest of columns being $\mathbf{0}$. Therefore based on previously proved orthogonality,

$$
\begin{aligned}
\frac{1}{n}\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\left[(\mathbf{I}-P_{\tilde{\mathbf{X}}})h(\mathbf{X}) \quad \mathbf{0} \quad \ldots \quad \mathbf{0}\right] &\xrightarrow{P} \mathbb{E}\left\{\tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}\left[(\mathbf{I}-P_{\tilde{\mathbf{X}}})h(\mathbf{X}) \quad \mathbf{0} \quad \ldots \quad \mathbf{0}\right]\right\}\\
&= \mathbf{0}\\
\therefore\ \frac{1}{n}\bar{\mathbf{X}}^T h(\mathbf{X}) &\xrightarrow{P} \mathbf{0}
\end{aligned}
$$

Using similar strategy, we can show that

$$
\frac{1}{n}\bar{\mathbf{X}}^T\text{diag}\{p(\mathbf{X})\}\tilde{\mathbf{X}}\beta_0 \xrightarrow{P} \mathbf{0}
$$

Since $X_i \perp \epsilon_i\ \mathbb{E}(\epsilon_i) = 0\ \forall\ i$, we have $\mathbb{E}(\tilde{\mathbf{X}}^T\boldsymbol{\epsilon}) = \mathbf{0}$.

$$
\begin{aligned}
\because\ \bar{\mathbf{X}}^T &= \tilde{\mathbf{X}}^T\text{diag}\{\mathbf{A}-p(\mathbf{X})\}(\mathbf{I}-P_{\tilde{\mathbf{X}}})\\
\therefore\ \bar{\mathbf{X}}_{(i)} &\in \mathcal{C}(\tilde{\mathbf{X}})\ \forall\ i
\end{aligned}
$$

where $\bar{\mathbf{X}}_{(i)}$ denotes the $i-$th row of $\bar{\mathbf{X}}$.

$$\therefore\ \frac{1}{n}\bar{\mathbf{X}}^T\boldsymbol{\epsilon} \xrightarrow{P} \mathbb{E}(\bar{\mathbf{X}}^T\boldsymbol{\epsilon}) = \mathbf{0}$$

$$\therefore\ \hat{\boldsymbol{\beta}} \xrightarrow{P} \mathbf{C}^{-1}\mathbf{0} + \boldsymbol{\beta}_0 = \boldsymbol{\beta}_0$$

This shows that $\hat{\boldsymbol{\beta}}$ is indeed a consistent estimator for $\boldsymbol{\beta}_0$.

(ii)

Table 1: Simulation result for $\hat{\boldsymbol{\beta}}$

	N=50				100				200				500			
	S1C1	S1C2	S2C1	S2C2	S1C1	S1C2	S2C1	S2C2	S1C1	S1C2	S2C1	S2C2	S1C1	S1C2	S2C1	S2C2
$\hat{\beta}_0$	0.55 (0.33)	0.52 (0.29)	0.51 (0.30)	0.53 (0.30)	0.52 (0.21)	0.52 (0.18)	0.51 (0.20)	0.51 (0.14)	0.51 (0.15)	0.51 (0.13)	0.50 (0.15)	0.51 (0.13)	0.50 (0.09)	0.50 (0.08)	0.50 (0.09)	0.50 (0.09)
$\hat{\beta}_1$	-1.08 (0.52)	-1.05 (0.42)	-1.03 (0.48)	-1.03 (0.47)	-1.04 (0.33)	-1.03 (0.26)	-1.02 (0.31)	-1.02 (0.32)	-1.01 (0.21)	-1.01 (0.17)	-1.00 (0.22)	-1.01 (0.22)	-1.01 (0.14)	-1.01 (0.11)	-1.00 (0.13)	-1.00 (0.14)
$\hat{\beta}_2$	1.03 (0.54)	1.02 (0.46)	1.02 (0.46)	1.04 (0.48)	1.02 (0.34)	1.00 (0.28)	1.03 (0.30)	1.00 (0.29)	1.00 (0.24)	1.00 (0.19)	1.00 (0.22)	1.00 (0.21)	1.01 (0.14)	1.00 (0.11)	1.00 (0.13)	1.00 (0.13)

Note: S stands for Scenario, and C stands for Case; Values are Mean (SD).