STK3100 Exercises, Week 5

Vinnie Ko, Jonas Moss

September 27, 2018

Exercise 4.4

i)

From p.124 of the book, we have $\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{i,j}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0$. By replacing $x_{i,j}$ with $\frac{\partial \eta_i}{\partial \beta_j}$, we obtain

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{i,j}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0$$

ii)

In generalized least squares, $M = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\text{Var}(y_i)}$ is minimized with respect to model parameters. The solution then should satisfy $\frac{\partial M}{\partial \beta_j} = \frac{\partial \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\text{Var}(y_i)}}{\partial \beta_j} = 0$ for all $j = 1, \cdots, p$. So,

$$\frac{\partial M}{\partial \beta_j} = \sum_{i=1}^{n} \frac{1}{\text{Var}(y_i)} \frac{\partial (y_i - \mu_i)^2}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = -2 \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0.$$

By simplifying the last part, we obtain $\sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0$.

Exercise 4.6

Notice that

$$\text{Var}(Y_i) = \mu_i = \begin{cases} \mu_A = g^{-1}(\eta_A) = \exp(\eta_A) = \exp(\beta_0 + \beta_1) & \text{if} \quad 1 \le i \le n_A \\ \mu_B = g^{-1}(\eta_B) = \exp(\eta_B) = \exp(\beta_0) & \text{if} \quad n_A + 1 \le i \le n_A + n_B \end{cases}.$$

So, we have

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i=1}^{n_A + n_B} \frac{(y_i - \mu_i) \cdot 1}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \\ &= \sum_{i=1}^{n_A} \frac{(y_i - \mu_i)}{e^{\eta_A}} e^{\eta_A} + \sum_{i=n_A+1}^{n_B} \frac{(y_i - \mu_i)}{e^{\eta_B}} e^{\eta_B} \\ &= \sum_{i=1}^{n_A} (y_i - \mu_i) + \sum_{i=n_A+1}^{n_B} (y_i - \mu_i) \\ &= 0 \end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1} &= \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_i}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \\
&= \sum_{i=1}^{n_A} \frac{(y_i - \mu_i) \cdot 1}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \\
&= \sum_{i=1}^{n_A} \frac{(y_i - \mu_i)}{e^{\eta_A}} e^{\eta_A} \\
&= \sum_{i=1}^{n_A} (y_i - \mu_i) \\
&= 0.
\end{aligned}
$$

This implies $\sum_{i=1}^{n_A}(y_i - \mu_i) = \sum_{i=1}^{n_A}(y_i - \mu_A) = 0$ and $\sum_{i=n_A+1}^{n_B}(y_i - \mu_i) = \sum_{i=n_A+1}^{n_B}(y_i - \mu_B) = 0$.

Thus, $\widehat{\mu}_A = \frac{1}{n_A}\sum_{i=1}^{n_A} y_i$ and $\widehat{\mu}_B = \frac{1}{n_B}\sum_{i=n_A+1}^{n_A+n_B} y_i$.

Exercise 4.9

From p.126 of the book, we have that $\boldsymbol{W}$ is a diagonal matrix with diagonal elements $w_i = \frac{1}{\mathrm{Var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$. Under ordinary normal linear model, $g(\mu) = \mu = \eta$. So, $\frac{\partial \mu_i}{\partial \eta_i} = 1$ and $w_i = \frac{1}{\mathrm{Var}(Y_i)}$. Consequently, $\boldsymbol{W} = \sigma^{-2}\boldsymbol{I}$.
By plugging this result into $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}$, we obtain $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}$.

Exercise 5.4

Recall equation (5.5, p. 173), which is true for the logistic model:

$$
\sum_{i=1}^{N} n_i\,(y_i - \pi_i)\,x_{ij} = 0,\ j = 1, \ldots, p.
$$

When this equation holds for β_0, $x_{i,j} = 1$ and we have $\sum_{i=1}^{n} n_i y_i = \sum_{i=1}^{n} n_i \pi_i$. By dividing $N = \sum_{i=1}^{n} n_i$ on both sides, we obtain $\overline{y} = \frac{1}{N}\sum_{i=1}^{n} n_i y_i = \frac{1}{N}\sum_{i=1}^{n} n_i \pi_i = \pi$. So, the estimated (overall) success probability equals overall sample proportion of successes.

Since the equation is specific to logistic regression, the result doesn't hold for binary GLMs where the link function is not logistic link function.

Exercise 5.6

The equation in question is:

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1} = \left(\boldsymbol{X}^{\mathrm{T}}\mathrm{Diag}(n_i\widehat{\pi}_i(1-\widehat{\pi}_i))\boldsymbol{X}\right)^{-1}$$

(a)

Let n_i increase as kn_i with k. Then $\left\{X^T \mathrm{diag}\left[kn_i\widehat{\pi}_i\left(1-\widehat{\pi}_i\right)\right]X\right\}^{-1} = k^{-1}\widehat{\mathrm{var}(\beta)}$.

(b)

Note that in case of ungrouped data $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^{\mathrm{T}}\mathrm{Diag}(\widehat{\pi}_i(1-\widehat{\pi}_i))\boldsymbol{X}\right)^{-1} = \left(\left(\sum_{i=1}^{n} x_{i,h}\, x_{i,j}\, \widehat{\pi}_i(1-\widehat{\pi}_i)\right)_{h,j}\right)^{-1}$.

So, when N increases, the number of terms within the sum increases and $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}})$ therefore decreases.

Additional Exercise 13

a)

We know that (i) $M_X\left(at\right) = M_{aX}\left(t\right)$ and (ii) $M_{\sum_{i=1}^n X_i}\left(t\right) = \prod_{i=1}^{n} M_{X_i}\left(t\right)$ when the variables X_i are independent. Thus $M_{n^{-1}\sum_{i=1}^n X_i}\left(t\right) = M_{\sum_{i=1}^n X_i}\left(n^{-1}t\right)$ so it suffices to show that $\prod_{i=1}^{n} M_{X_i}\left(t\right)$ is an MGF from the exponential dispersion family when X_i are iid from an exponential dispersion family. But this is clear, as $\prod_{i=1}^{n} M_{X_i}\left(t\right) = \prod_{i=1}^{n} \exp\left(\frac{b\left(\theta + ta\left(\phi\right)\right) - b\left(\theta\right)}{a\left(\phi\right)}\right) = \exp\left(\frac{b\left(\theta + ta\left(\phi\right)\right) - b\left(\theta\right)}{a\left(\phi\right)/n}\right)$.

b)

We know that the Bernoulli distribution has MGF $\left[(1-p) + pe^t\right]$, so the MGF of $\sum_{i=1}^{n} X_i$ is $\left[(1-p) + pe^t\right]^n$. This is the MGF of the binomial distribution, so the distribution of $n^{-1}\sum_{i=1}^{n} X_i$ is the same as the distribution of Y/n where Y is Binomial$\left(n, \pi\right)$-distributed.

Additional Exercise 14

Note that we stick to the notation of the book $L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log f(Y_i, \boldsymbol{\beta}) = \log \prod_{i=1}^{n} f(Y_i, \boldsymbol{\beta}) = \log \ell(\boldsymbol{\beta})$, which is the opposite of the usual notation $\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log f(Y_i, \boldsymbol{\beta}) = \log \prod_{i=1}^{n} f(Y_i, \boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$.

i)

Lemma 1

For arbitrary i and h,

$$\begin{aligned}
\mathrm{E}\left[S_{i,h}(\boldsymbol{\beta})\right] &= \int S_{i,h}(\boldsymbol{\beta}) f_{Y_i}\, dy \\
&= \int \frac{\partial \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h} f_{Y_i}\, dy \\
&= \int \frac{1}{f_{Y_i}(y_i, \boldsymbol{\beta})} \frac{\partial f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h} f_{Y_i}\, dy \\
&= \int \frac{\partial f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h}\, dy \\
&= \frac{\partial}{\partial \beta_h} \int f_{Y_i}(y_i, \boldsymbol{\beta})\, dy \\
&= \frac{\partial 1}{\partial \beta_h} \\
&= 0.
\end{aligned}$$

So,

$$\mathrm{E}\left[S_h(\boldsymbol{\beta})\right] = \sum_{i=1}^{n} \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta})\right] = \sum_{i=1}^{n} 0 = 0.$$

$\square$

Lemma 2 (a.k.a information matrix equality)
This is a proof of $\mathcal{J}_{i,h,j}(\boldsymbol{\beta}) = \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta}) S_{i,j}(\boldsymbol{\beta})\right]$, stated at page 126.

First, note that for an arbitrary function $g(x)$, $\dfrac{\partial g(x)}{\partial x} = g(x) \dfrac{\partial \log g(x)}{\partial x}$.
Then, for an arbitrary i, h and j,

$$\begin{aligned}
0 &= \frac{\partial \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta})\right]}{\partial \beta_j} \\
&= \frac{\partial}{\partial \beta_j} \int \frac{\partial \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h} f_{Y_i}\, dy \\
&= \int \frac{\partial^2 \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h \partial \beta_j} f_{Y_i} + \frac{\partial \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h} \frac{\partial f_{Y_i}}{\partial \beta_j}\, dy \\
&= \int \frac{\partial^2 \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h \partial \beta_j} f_{Y_i} + \frac{\partial \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_h} \frac{\partial \log f_{Y_i}(y_i, \boldsymbol{\beta})}{\partial \beta_j} f_{Y_i}\, dy \\
&= -\mathcal{J}_{i,h,j}(\boldsymbol{\beta}) + \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta}) S_{i,j}(\boldsymbol{\beta})\right].
\end{aligned}$$

So, $\mathcal{J}_{i,h,j}(\boldsymbol{\beta}) = \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta}) S_{i,j}(\boldsymbol{\beta})\right]$. $\square$

Finally, we have

$$\begin{aligned}
\mathrm{Cov}\left(S_h(\boldsymbol{\beta}), S_j(\boldsymbol{\beta})\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} S_{i,h}(\boldsymbol{\beta}),\ \ \sum_{i=1}^{n} S_{i,j}(\boldsymbol{\beta})\right) \\
&= \sum_{i=1}^{n} \mathrm{Cov}\left(S_{i,h}(\boldsymbol{\beta}) \cdot S_{i,j}(\boldsymbol{\beta})\right) && (\because \text{ independence}) \\
&= \sum_{i=1}^{n} \left[\mathrm{E}\left[S_{i,h}(\boldsymbol{\beta}) \cdot S_{i,j}(\boldsymbol{\beta})\right] - \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta})\right] \mathrm{E}\left[S_{i,j}(\boldsymbol{\beta})\right]\right] \\
&= \sum_{i=1}^{n} \mathrm{E}\left[S_{i,h}(\boldsymbol{\beta}) \cdot S_{i,j}(\boldsymbol{\beta})\right] && (\text{by lemma 1}) \\
&= \sum_{i=1}^{n} \mathcal{J}_{i,h,j}(\boldsymbol{\beta}) && (\text{by lemma 2}) \\
&= \mathcal{J}_{h,j}(\boldsymbol{\beta})
\end{aligned}$$

where $\mathcal{J}_{h,j}(\boldsymbol{\beta}) = \mathrm{E}\left[-\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_h \partial \beta_j}\right]$.

ii)

$$\begin{aligned}
\boldsymbol{\mathcal{J}}(\boldsymbol{\beta}) &= \mathrm{Cov}\left(\boldsymbol{S}(\boldsymbol{\beta}), \boldsymbol{S}(\boldsymbol{\beta})\right) \\
&= \mathrm{E}\left[\boldsymbol{S}(\boldsymbol{\beta})\boldsymbol{S}(\boldsymbol{\beta})^{\mathrm{T}}\right]
\end{aligned}$$

Let $\boldsymbol{a}$ be a arbitrary non-zero vector, then

$$\begin{aligned}
\boldsymbol{a}^{\mathrm{T}} \boldsymbol{\mathcal{J}}(\boldsymbol{\beta}) \boldsymbol{a} &= \boldsymbol{a}^{\mathrm{T}} \mathrm{Cov}\left(\boldsymbol{S}(\boldsymbol{\beta}), \boldsymbol{S}(\boldsymbol{\beta})\right) \boldsymbol{a} \\
&= \mathrm{Cov}\left(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{S}(\boldsymbol{\beta}), \boldsymbol{a}^{\mathrm{T}} \boldsymbol{S}(\boldsymbol{\beta})\right) \\
&= \mathrm{Var}\left(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{S}(\boldsymbol{\beta})\right) \\
&\geq 0.
\end{aligned}$$

So, $\boldsymbol{\mathcal{J}}(\boldsymbol{\beta})$ is positive semi-definite.