Problem Set 5: Solution

Part One: Hand-Written Exercise

1. Since $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, we have:

$$\ell(y_i, x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2}.$$

Therefore, the log likelihood function is given by:

$$L_N(\beta, \sigma^2) = \ln \Pi_{i=1}^N \ell(y_i, x_i, \beta, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2.$$

(a) The $\hat{\beta}_{ML}$ that maximizes $L_N(\beta, \sigma^2)$ would be the one that minimizes $\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$, which is the OLS criteria! Therefore, $\hat{\beta}_{ML} = \hat{\beta}_{OLS}$.

(b) To obtain $\hat{\sigma}^2_{ML}$, we fist derive the F.O.C of $L_N(\beta, \sigma^2)$ with respect to σ^2.

$$\frac{\partial}{\partial \sigma^2} L_N(\beta, \sigma^2) = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2 = 0$$
$$\Rightarrow \hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^N (y_i - \hat{\beta}_{0,ML} - \hat{\beta}_{1,ML} x_i)^2}{N},$$

which is different from the OLS estimator $\hat{\sigma}^2_{OLS} = \frac{\sum_{i=1}^N (y_i - \hat{\beta}_{0,OLS} - \hat{\beta}_{1,OLS} x_i)^2}{N-2}$.

2. First, we have

$$\begin{aligned}
\mathbf{H}(\boldsymbol{\theta}_0) &= \mathbb{E}[\nabla^2 \ln \ell(\boldsymbol{\theta}_0)] \\
&= \mathbb{E}\left(\nabla\left[\frac{\left[y_i - \Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]\phi(\mathbf{x}_i'\boldsymbol{\theta}_0)}{\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]}\mathbf{x}_i\right]\right) \\
&= \mathbb{E}\left(\frac{\phi'(\mathbf{x}_i'\boldsymbol{\theta}_0)\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)y_i - \phi'(\mathbf{x}_i'\boldsymbol{\theta}_0)\Phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0) - \phi'(\mathbf{x}_i'\boldsymbol{\theta}_0)\Phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0)y_i + \phi'(\mathbf{x}_i'\boldsymbol{\theta}_0)\Phi^3(\mathbf{x}_i'\boldsymbol{\theta}_0)}{\left(\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]\right)^2}\mathbf{x}_i\mathbf{x}_i'\right) \\
&\quad - \mathbb{E}\left(\frac{\left[y_i - 2y_i\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0) + \Phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]\phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0)}{\left(\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]\right)^2}\mathbf{x}_i\mathbf{x}_i'\right) \\
&= 0 - \mathbb{E}\left(\frac{\left[y_i - 2y_i\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0) + \Phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]\phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0)}{\left(\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]\right)^2}\mathbf{x}_i\mathbf{x}_i'\right) \\
&= -\mathbb{E}\left(\frac{\phi^2(\mathbf{x}_i'\boldsymbol{\theta}_0)}{\Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\theta}_0)\right]}\mathbf{x}_i\mathbf{x}_i'\right),
\end{aligned}$$

where that last two equations are by the law of iterated expectation and the fact that $\mathbb{E}\left(y_i|\mathbf{x}_i\right) = \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)$.

Now, for the information matrix $\mathbf{B}(\boldsymbol{\theta}_0)$, we have:

$$\begin{aligned}
\mathbf{B}(\boldsymbol{\theta}_0) &= \mathbb{E}\left(\frac{1}{N}(\nabla L_n(\boldsymbol{\theta}_0))(\nabla L_n(\boldsymbol{\theta}_0))^{'}\right) \\
&= \mathbb{E}\left(\frac{\left[y_i - \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\right]^2 \phi^2(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)}{\left(\Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\right]\right)^2}\mathbf{x}_i\mathbf{x}_i^{'}\right) \quad \text{when } (y_i, \mathbf{x}_i^{'})' \text{ are iid data.} \\
&= \mathbb{E}\left(\frac{\left[y_i^2 - 2y_i\Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0) + \Phi^2(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\right] \phi^2(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)}{\left(\Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\right]\right)^2}\mathbf{x}_i\mathbf{x}_i^{'}\right) \\
&= \mathbb{E}\left(\frac{\phi^2(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)}{\Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\left[1 - \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\right]}\mathbf{x}_i\mathbf{x}_i^{'}\right),
\end{aligned}$$

where the last equation is due to the fact that:

$$\text{Var}\left(y_i|\mathbf{x}_i\right) = \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\left(1 - \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0)\right),$$

and

$$\begin{aligned}
\text{Var}\left(y_i|\mathbf{x}_i\right) &= \mathbb{E}\left(y_i^2|\mathbf{x}_i\right) - \mathbb{E}^2\left(y_i|\mathbf{x}_i\right) \\
&= \mathbb{E}\left(y_i^2|\mathbf{x}_i\right) - \Phi^2(\mathbf{x}_i^{'}\boldsymbol{\theta}_0),
\end{aligned}$$

which leads to the fact

$$\mathbb{E}\left(y_i^2|\mathbf{x}_i\right) = \Phi(\mathbf{x}_i^{'}\boldsymbol{\theta}_0) = \mathbb{E}\left(y_i|\mathbf{x}_i\right).$$

It follows that the information equality holds: $\mathbf{H}(\boldsymbol{\theta}_0) + \mathbf{B}(\boldsymbol{\theta}_0) = 0$

3. (a) (Method 1)

   The pdf of Bernoulli distribution is

$$f(y_i) = G(\boldsymbol{x_i'\beta})^{y_i}(1 - G(\boldsymbol{x_i'\beta}))^{1-y_i}, \text{ where } y_i = 0, 1.$$

   So the likelihood function is

$$\begin{aligned}
L &:= \prod_{i=1}^{12} f(y_i) \\
&= (\frac{1}{1 + exp(-\beta_0 - \beta_1)})^3(\frac{exp(-\beta_0 - \beta_1)}{1 + exp(-\beta_0 - \beta_1)})^3(\frac{1}{1 + exp(-\beta_0)})^4(\frac{exp(-\beta_0)}{1 + exp(-\beta_0)})^2 \\
&= (1 + exp(-\beta_0 - \beta_1))^{-6}(1 + exp(-\beta_0))^{-6}(exp(-\beta_0 - \beta_1))^3(exp(-\beta_0))^2
\end{aligned}$$

And the log-likelihood function is

$$
\begin{aligned}
l &:= \log L \\
&= -6\log(1+e^{-\beta_0-\beta_1}) - 6\log(1+e^{-\beta_0}) - 5\beta_0 - 3\beta_1
\end{aligned}
$$

Then we can get the f.o.c.

$$
\frac{\partial l}{\partial \beta_0} = 6\frac{exp(-\beta_0-\beta_1)}{1+exp(-\beta_0-\beta_1)} + 6\frac{exp(-\beta_0)}{1+exp(-\beta_0)} - 5 = 0 \tag{1}
$$

$$
\frac{\partial l}{\partial \beta_1} = 6\frac{exp(-\beta_0-\beta_1)}{1+exp(-\beta_0-\beta_1)} - 3 = 0 \tag{2}
$$

By (1) - (2), we get

$$
\begin{aligned}
6\frac{exp(-\beta_0)}{1+exp(-\beta_0)} &= 2 \\
\Rightarrow \hat{\beta}_0 = -\log\frac{1}{2} &= 0.6931
\end{aligned}
$$

Replace $\hat{\beta}_0$ in (2), we get $\hat{\beta}_1 = -\log 2 = -0.6931$.

(b) (Method 2)

By the result on slide 17, Lecture 5, the f.o.c. is

$$
\begin{aligned}
\nabla L_n(\boldsymbol{\beta}) &= \begin{bmatrix} \frac{\partial L_n(\beta_0)}{\partial \beta_0} \\ \frac{\partial L_n(\beta_1)}{\partial \beta_1} \end{bmatrix} \\
&= \sum_{i=1}^{n}[y_i - G(\boldsymbol{x_i'\beta})]\boldsymbol{x_i} \\
&= \sum_{i=1}^{n}[y_i - G(\boldsymbol{x_i'\beta})]\begin{bmatrix} 1 \\ x_i \end{bmatrix} \\
&= \sum_{i=1}^{n}\begin{bmatrix} y_i - G(\boldsymbol{x_i'\beta}) \\ (y_i - G(\boldsymbol{x_i'\beta}))x_i \end{bmatrix} = 0
\end{aligned}
$$

$$
\Rightarrow \begin{cases} \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} G(\boldsymbol{x_i'\beta}) = \sum_{i=1}^{n} \frac{1}{1+exp(-\beta_0-\beta_1 x_i)} \\ \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} G(\boldsymbol{x_i'\beta})x_i = \sum_{i=1}^{n} \frac{1}{1+exp(-\beta_0-\beta_1 x_i)}x_i \end{cases}
$$

$$
\Rightarrow \begin{cases} 7 = \frac{6}{1+exp(-\beta_0)} + \frac{6}{1+exp(-\beta_0-\beta_1)} \\ 3 = \frac{6}{1+exp(-\beta_0-\beta_1)} \end{cases}
$$

$$
\Rightarrow \begin{cases} \hat{\beta}_0 = -\log(\frac{1}{2}) = 0.6931 \\ \hat{\beta}_1 = -\log(2) = -0.6931 \end{cases}
$$