STK3100 Exercises, Week 7

Vinnie Ko

October 11, 2018

Exercise 4.13

i)

From p. 135 of the book, we have that the Pearson chi-squared statistic is defined as $X^2 = \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu}_i)^2}{\text{Var}(Y_i)}$.

Further, it's given that $y_i \sim N(\mu_i, \sigma^2)$. So, $X^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{\mu}_i)^2}{\sigma^2} \sim \chi^2_{n-p}$.

The deviance is

$$
\begin{aligned}
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= -2\left(L\left(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}\right) - L\left(\boldsymbol{y}, \boldsymbol{y}\right)\right) \\
&= -2\left(-\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \widehat{\mu}_i)^2 - \left(-\frac{n}{2}\log(2\pi) - n\log\sigma\right)\right) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \widehat{\mu}_i)^2.
\end{aligned}
$$

So, the Pearson chi-squared statistic is the same as the deviance.

ii)

$$
\begin{aligned}
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}_0) - D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}_1) &= -2\left(L\left(\widehat{\boldsymbol{\mu}}_0, \boldsymbol{y}\right) - L\left(\widehat{\boldsymbol{\mu}}_1, \boldsymbol{y}\right)\right) \\
&= -2\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \widehat{\mu}_{0,i})^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \widehat{\mu}_{1,i})^2\right) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[(y_i - \widehat{\mu}_{0,i})^2 - (y_i - \widehat{\mu}_{1,i})^2\right]
\end{aligned}
$$

Exercise 4.14

The likelihood equation of GLM is $\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{i,j}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0$ (p.124 of the book).

For the intercept. this becomes $\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0$.

Canonical link function $g(\cdot)$ is a link function such that $\theta = g(\mathrm{E}[Y])$. (p.123 of the book).
So, $\dfrac{\partial \mu_i}{\partial \eta_i} = \dfrac{\partial \mu_i}{\partial \theta_i} = \dfrac{\partial b'(\theta_i)}{\partial \theta_i} = b''(\theta_i)$. Further, we know $\mathrm{Var}(Y_i) = b''(\theta_i)a(\phi)$.
Thus,

$$\begin{aligned}\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i=1}^{n} \frac{(y_i-\mu_i)}{\mathrm{Var}(Y_i)}\frac{\partial \mu_i}{\partial \eta_i} \\ &= \sum_{i=1}^{n}(y_i-\mu_i)\frac{1}{a(\phi)} \\ &= 0\end{aligned}$$

In most cases $a(\phi)$ doesn't depend on the data. In that case,

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} = \frac{1}{a(\phi)}\sum_{i=1}^{n}(y_i-\mu_i) = 0.$$

So, $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \widehat{\mu}_i$ and this equality doesn't necessarily hold for a GLM with non-canonical function.
Obviously, when there is no β_0 for a GLM with canonical link function, we can't derive this equality.

Exercise 4.16

Note: I use the binomial distribution as defined in p.122 of the book. This implies that y_i is a proportion of success, instead of a number of success.

a)

The pmf of binomial distribution is

$$\begin{aligned}f(y_i) &= \binom{n_i}{n_i y_i}\pi_i^{n_i y_i}(1-\pi_i)^{n_i-n_i y_i} \\ &= \exp\left[\frac{\theta_i y_i - \log\left(1+e^{\theta_i}\right)}{\frac{1}{n_i}} + \log\binom{n_i}{n_i y_i}\right] \\ &= \exp\left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right]\end{aligned}$$

where $\theta_i = \log\left(\dfrac{\pi_i}{1-\pi_i}\right)$, $\pi_i = \dfrac{e^{\theta_i}}{1+e^{\theta_i}}$, $b(\theta_i) = \log\left(1+e^{\theta_i}\right) = -\log(1-\pi_i)$, $c(y_i) = \log\dbinom{n_i}{n_i y_i}$, $a(\phi) = \dfrac{1}{n_i}$.
Since $a(\phi) = \dfrac{1}{n_i}$, $w_i = n_i$ (p.121 of the book) and we have

$$\begin{aligned}d_i &= 2w_i\left[y_i\left(\widetilde{\theta}_i - \widehat{\theta}_i\right) - b\left(\widetilde{\theta}_i\right) + b\left(\widehat{\theta}_i\right)\right] \\ &= 2n_i\left[y_i\left(\log\left(\frac{y_i}{1-y_i}\right) - \log\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right)\right) + \log(1-y_i) - \log(1-\widehat{\pi}_i)\right] \\ &= 2n_i\left[y_i\log y_i - y_i\log(1-y_i) - y_i\log\widehat{\pi}_i + y_i\log(1-\widehat{\pi}_i) + \log(1-y_i) - \log(1-\widehat{\pi}_i)\right] \\ &= 2n_i\left[y_i\log\frac{y_i}{\widehat{\pi}_i} + (1-y_i)\log\left(\frac{1-y_i}{1-\widehat{\pi}_i}\right)\right] \\ &= 2\left[n_i y_i\log\frac{y_i}{\widehat{\pi}_i} + n_i(1-y_i)\log\left(\frac{1-y_i}{1-\widehat{\pi}_i}\right)\right].\end{aligned}$$

b)

The pmf of Poisson distribution is

$$
\begin{aligned}
f(y_i) &= \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} \\
&= \exp\left[y_i \log \mu_i - \mu_i - \log y_i!\right] \\
&= \exp\left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right]
\end{aligned}
$$

where $\theta_i = \log \mu_i$, $b(\theta_i) = e^{\theta_i} = \mu_i$, $c(y_i) = -\log y_i!$, $a(\phi) = 1$, $w_i = 1$.
Thus, we have

$$
\begin{aligned}
d_i &= 2w_i \left[y_i \left(\widetilde{\theta}_i - \widehat{\theta}_i\right) - b\left(\widetilde{\theta}_i\right) + b\left(\widehat{\theta}_i\right)\right] \\
&= 2\left[y_i \left(\log y_i - \log \widehat{\mu}_i\right) - y_i + \widehat{\mu}_i\right] \\
&= 2\left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - y_i + \widehat{\mu}_i\right].
\end{aligned}
$$

This matches the expression in p.133 of the book.

Exercise 4.19

a)

For $\beta^{(0)}$ that is close to $\widehat{\beta}$, we can use first order Taylor approximation

$$
0 = L'(\widehat{\beta}) = L'(\beta^{(0)}) + \left(\widehat{\beta} - \beta^{(0)}\right) L''(\beta^{(0)}) + o_p(n^{-\frac{1}{2}}).
$$

We can rewrite this as

$$
\widehat{\beta} \approx \beta^{(0)} - \frac{L'(\beta^{(0)})}{L''(\beta^{(0)})}.
$$

Approximating $\widehat{\beta}$ by $\beta^{(1)}$ gives us

$$
\beta^{(1)} = \beta^{(0)} - \frac{L'(\beta^{(0)})}{L''(\beta^{(0)})}.
$$

b)

By generalizing the result from a), we have

$$
\beta^{(t+1)} = \beta^{(t)} - \frac{L'(\beta^{(t)})}{L''(\beta^{(t)})}.
$$

Exercise 5.14

```
set.seed(313)
n = 100
x = sort(runif(n = n, min = 0, max = 100))
logit_pi = -2.0 + 0.04*x
```

```
pi = exp(logit_pi)/(1 + exp(logit_pi))
y = rbinom(n = n, size = 1, prob = pi)
model = glm(y ~ 1 + x, family = binomial(link = "logit"))
plot(pi, residuals(model))
plot(x, residuals(model))
lines(x, pi) abline(h = 0)
abline(h = 1)
```

Exercise 5.14

Note: I use the binomial distribution as defined in p.122 of the book. This implies that y_i is a proportion of success, instead of a number of success.

The binomial pmf with a common π is

$$f(y_i) = \binom{n_i}{n_i y_i} \pi^{n_i y_i} (1-\pi)^{n_i - n_i y_i}$$

Then, the log-likelihood is

$$L(\pi) = \sum_{i=1}^{N} \left[\log \binom{n_i}{n_i y_i} + n_i y_i \log \pi + (n_i - n_i y_i) \log(1-\pi) \right].$$

By solving

$$\frac{\partial L(\pi)}{\partial \pi} = \frac{1}{\pi} \sum_{i=1}^{N} n_i y_i - \frac{1}{1-\pi} \sum_{i=1}^{N} (n_i - n_i y_i) = 0$$

$$(1-\pi) \sum_{i=1}^{N} n_i y_i = \pi \sum_{i=1}^{N} (n_i - n_i y_i)$$

we obtain

$$\widehat{\pi} = \frac{\sum_{i=1}^{N} n_i y_i}{\sum_{i=1}^{N} n_i}.$$

The Pearson chi-squared statistic (p.135 of the book) is

$$\begin{aligned} X^2 &= \sum_{i=1}^{N} \frac{(y_i - \widehat{\pi})^2}{\operatorname{Var}(Y_i)} \\ &= \sum_{i=1}^{N} \frac{(y_i - \widehat{\pi})^2}{\widehat{\pi}(1-\widehat{\pi})/n_i} \end{aligned}$$

When all $n_i = 1$, we obtain $\widehat{\pi} = \dfrac{\sum_{i=1}^N n_i y_i}{\sum_{i=1}^N n_i} = \dfrac{\sum_{i=1}^N y_i}{N}$ and

$$
\begin{aligned}
X^2 &= \sum_{i=1}^N \frac{(y_i - \widehat{\pi})^2}{\mathrm{Var}(Y_i)} \\
&= \frac{1}{\widehat{\pi}(1-\widehat{\pi})} \sum_{i=1}^N (y_i - \widehat{\pi})^2 \\
&= \frac{1}{\widehat{\pi}(1-\widehat{\pi})} \sum_{i=1}^N (y_i^2 - 2\widehat{\pi} y_i + \widehat{\pi}^2) \\
&= \frac{1}{\widehat{\pi}(1-\widehat{\pi})} \left(\sum_{i=1}^N y_i - 2\widehat{\pi} \sum_{i=1}^N y_i + N\widehat{\pi}^2 \right) \\
&= \frac{1}{\widehat{\pi}(1-\widehat{\pi})} \left(N\widehat{\pi} - 2\widehat{\pi} N\widehat{\pi} + N\widehat{\pi}^2 \right) \\
&= \frac{1}{\widehat{\pi}(1-\widehat{\pi})} N\widehat{\pi}\,(1-\widehat{\pi}) \\
&= N.
\end{aligned}
$$

Exercise 5.15

Note: I use the binomial distribution as defined in p.122 of the book. This implies that y_i is a proportion of success, instead of a number of success.

For binomial GLM, the deviance is defined as

$$
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^n n_i y_i \log\left(\frac{n_i y_i}{n_i \widehat{\pi}_i}\right) + 2\sum_{i=1}^n (n_i - n_i y_i) \log\left(\frac{n_i - n_i y_i}{n_i - n_i \widehat{\pi}_i}\right)
$$

We are given that $n_i = 1$. So,

$$
\begin{aligned}
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= 2\sum_{i=1}^n y_i \log\left(\frac{y_i}{\widehat{\pi}_i}\right) + 2\sum_{i=1}^n (1 - y_i) \log\left(\frac{1 - y_i}{1 - \widehat{\pi}_i}\right) \\
&= 2\sum_{i=1}^n \left[y_i \log y_i - y_i \log(\widehat{\pi}_i) + (1-y_i)\log(1-y_i) - (1-y_i)\log(1-\widehat{\pi}_i) \right]
\end{aligned}
$$

By noting that $y_i \log y_i$ and $(1-y_i)\log(1-y_i)$ are always 0.

$$
= 2\sum_{i=1}^n \left[-y_i \log(\widehat{\pi}_i) - (1-y_i)\log(1-\widehat{\pi}_i) \right]
$$

Since $\text{logit}\,(\widehat{\pi}_i) = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, $\widehat{\pi}_i = \dfrac{e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}}$

$$
\begin{aligned}
&= 2\sum_{i=1}^{n}\left[-y_i \log\left(\frac{e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}}\right) - (1-y_i)\log\left(1-\frac{e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}}{1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}}\right)\right] \\
&= 2\sum_{i=1}^{n}\left[-y_i\left(\widehat{\beta}_0+\widehat{\beta}_1 x_i\right) + \log\left(1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}\right)\right] \\
&= 2\left[-\sum_{i=1}^{n} y_i\left(\widehat{\beta}_0+\widehat{\beta}_1 x_i\right) + \sum_{i=1}^{n}\log\left(1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}\right)\right] \\
&= 2\left[-\widehat{\beta}_0\sum_{i=1}^{n} y_i - \widehat{\beta}_1\sum_{i=1}^{n} y_i x_i + \sum_{i=1}^{n}\log\left(1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}\right)\right]
\end{aligned}
$$

According to equation (5.5) from p.173 of the book, we have $\sum_{i=1}^{n} n_i \widehat{\pi}_i x_{i,j} = \sum_{i=1}^{n} n_i y_i x_{i,j}$.

$$
= 2\left[-\widehat{\beta}_0\sum_{i=1}^{n} \widehat{\pi}_i - \widehat{\beta}_1\sum_{i=1}^{n} \widehat{\pi}_i x_i + \sum_{i=1}^{n}\log\left(1+e^{\widehat{\beta}_0+\widehat{\beta}_1 x_i}\right)\right]
$$

Therefore, the deviance depends on $\widehat{\pi}_i$ and not on y_i. This implies that goodness-of-fit statistics are uninformative for ungrounded data.