

MATH3823 - Solutions to Chapter 4 Exercises

Exercise 4.1 We have that the estimates of the natural parameter is given by, $\hat{\theta} = (b')^{-1}(\bar{y})$, hence

- (a) For the binomial, $b(\theta) = m \log(1 + e^\theta)$ and so $b'(\theta) = m e^\theta / (1 + e^\theta)$. The estimate, $\hat{\theta}$ is given by the solution of

$$b'(\hat{\theta}) = m \frac{e^{\hat{\theta}}}{1 + e^{\hat{\theta}}} = \bar{y}$$

and so θ can be estimated as

$$\hat{\theta} = \log \left\{ \frac{\bar{y}/m}{1 - \bar{y}/m} \right\} = \log \left\{ \frac{\bar{y}}{m - \bar{y}} \right\}.$$

We know that the estimator, $\hat{\theta} = \log \{ \bar{Y} / (m - \bar{Y}) \}$ is asymptotically unbiased, at least. Its variance is given by

$$\text{Var}(\hat{\theta}) = \frac{\phi}{n b''(\theta_0)}$$

where θ_0 is the true value. Here

$$b''(\theta) = m \frac{e^\theta}{(1 + e^\theta)^2}$$

and hence

$$\text{Var}(\hat{\theta}) = \frac{1}{nm} \frac{(1 + e^{\theta_0})^2}{e^{\theta_0}}$$

Notice that this depends on the unknown true value θ_0 .

- (b) For the geometric, $b(\theta) = \theta - \log(1 - e^\theta)$ and so $b'(\theta) = 1 + e^\theta / (1 - e^\theta)$. The estimate, $\hat{\theta}$ is given by the solution of

$$b'(\hat{\theta}) = 1 + e^{\hat{\theta}} / (1 - e^{\hat{\theta}}) = \bar{y}$$

and so

$$\hat{\theta} = \log \{ 1 - 1/\bar{y} \}.$$

We know that the estimator, $\hat{\theta} = \log \{ 1 - 1/\bar{Y} \}$ is asymptotically unbiased, at least. To find the variance, first

$$b''(\theta) = \frac{e^\theta}{(1 - e^\theta)^2}$$

and hence

$$\text{Var}(\hat{\theta}) = \frac{1}{n} \frac{(1 - e^{\theta_0})^2}{e^{\theta_0}}$$

- (c) For the exponential, $\theta = -\lambda$ and $b(\theta) = -\log(-\theta)$ (these are not in our notes and hence need to first write exponential distribution in exponential family form). Hence, $b'(\theta) = -1/\theta$. The estimate, $\hat{\theta}$ is given by the solution of

$$b'(\hat{\theta}) = -\frac{1}{\hat{\theta}} = \bar{y}$$

and so

$$\hat{\theta} = -1/\bar{y}.$$

For the variance,

$$b''(\theta) = \frac{1}{\theta^2}$$

and hence

$$\text{Var}(\hat{\theta}) = \frac{\theta^2}{n}.$$

Exercise 4.2 For the exponential family form of the normal distribution we have $\theta = \mu$, $\phi = \sigma^2$, $b(\theta) = \theta^2/2$, hence $\hat{\theta} = \hat{\mu}$. Then, @eq-ehattheta gives

$$\text{E}[\hat{\mu}] = \text{E}[\hat{\theta}] \approx \theta_0 = \mu.$$

Further, since $b''(\theta) = 1$, then

$$\text{Var}[\hat{\theta}] = \frac{\sigma^2}{n}.$$

Hence, the approach based on the exponential family leads to the familiar results – as we would expect.

Exercise 4.3 Following the principle of maximum likelihood estimation we maximize the log-likelihood function, that is

$$\hat{\beta} = \max_{\beta} l(\beta; \mathbf{y}, \phi).$$

where $l(\beta; \mathbf{y}, \phi)$ is the log-likelihood of the parameters β given data \mathbf{y} and assuming that the nuisance scale parameter is fixed.

For the normal linear regression model the log-likelihood is given by

$$l = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta).$$

Differentiating with respect to vector β gives

$$\frac{dl}{d\beta} = -\frac{1}{\sigma^2} X^T (\mathbf{y} - X\beta)$$

which represents a system of p equations in p unknown. Setting this to zero gives the required result of @eq-linearregressionmle,

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}.$$

Exercise 4.4

- Under the normal model, with $\mu = \theta$, $b(\theta) = \theta^2/2$, $\phi = \sigma^2$, likelihood takes the form

$$l = \sum_{i=1}^n \{[\mu_i y_i - \mu_i^2/2]/\sigma^2 + c(y_i, \phi)\}$$

let \hat{l} denote the maximized likelihood under the generalized linear model of interest, with μ_i replaced by $\hat{\mu}_i$.

Under the saturated model, μ_i is estimated by maximizing the i th term of the likelihood, yielding $\hat{\mu} = y_i$. Substituting these values and noting that $y_i y_i - y_i^2/2 = y_i^2/2$ yields the maximized saturated log-likelihood

$$\tilde{l} = \sum_{i=1}^n \{[y_i^2/2]/\sigma^2 + c(y_i, \phi)\}$$

Hence the deviance reduces to the residual sum of squares

$$D = 2\phi(\tilde{l} - \hat{l}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2.$$

- Binomial data consist of the number of successes y_i out of m_i trials, $i = 1, \dots, n$. The likelihood takes a form that is slightly modified from the general case,

$$\begin{aligned} l &= \sum_{i=1}^n \{\theta_i y_i - b_i(\theta_i) + c(y_i, m_i)\} \\ &= \sum_{i=1}^n \{y_i \log(p_i) + (m_i - y_i) \log(1 - p_i) + c(y_i, m_i)\}. \end{aligned}$$

where $\theta_i = \text{logit}(p_i)$ and

$$p_i = \mu_i/m_i = E(Y_i)/m_i = b'_i(\theta_i)/m_i$$

is the probability of success in the i th set of trials, and $b_i(\theta_i) = m_i \log(1 + e^{\theta_i})$. Also,

$$\exp\{c(y_i, m_i)\} = \binom{m_i}{y_i}.$$

Under the saturated model p_i is estimated by $\tilde{p}_i = y_i/m_i$. Hence the deviance takes the form

$$D = 2(\tilde{l} - \hat{l}) = 2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \log \left(\frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right\}.$$

In general $\hat{p}_i = \text{logit}^{-1}(\hat{\beta}^T x_i)$ will always lie strictly between 0 and 1. However, if $y_i = 0$ or m_i , then the saturated estimate \tilde{p}_i will equal 0 or 1. Suppose that $y_i = 0$ for some i . Then the i th term of the saturated likelihood becomes $m_i + c(0, m_i)$. You should convince yourself that the formula for the deviance remains valid if we interpret $0 \log 0 = 0$. Similarly if $y_i = m_i$.

- In the Poisson case the likelihood becomes

$$\begin{aligned} l &= \sum_{i=1}^n \{\theta_i y_i - b(\theta_i) + c(y_i)\} \\ &= \sum_{i=1}^n \{y_i \log \lambda_i - \lambda_i + c(y_i)\}, \end{aligned}$$

where $\mu_i = \lambda_i = \log \theta_i$ and $b(\theta_i) = \exp(\theta_i)$. Under the saturated model $\tilde{\lambda}_i = y_i$. For most models of interest, it can be shown that $\sum \hat{\lambda}_i = \sum y_i$. Hence the deviance reduces to

$$D = 2(\tilde{l} - \hat{l}) = 2 \sum_{i=1}^n \{y_i \log(y_i/\hat{\mu}_i)\}.$$

Exercise 4.5 In the notes, we analysed the birthweight data using F -tests based on the residual sum of squares R and residual degrees of freedom r' (note the change of notation — r was used for degrees of freedom, but later we have used r for the number of parameters). In other other places, we described tests using the deviance D and number of parameters r .

Since $D = R$ for a normal GLM and $r' = n - r$, where n is the number of observations, the F -statistic comparing Model 0 and Model 1 can be written

$$F_{01} = \frac{(R_0 - R_1)/(r'_0 - r'_1)}{R_1/r'_1} = \frac{(D_0 - D_1)/(r_1 - r_0)}{D_1/(n - r_1)},$$

which we compared to an F -distribution on $r_1 - r_0$ and $n - r_1$ degrees of freedom. This is almost equivalent to the test statistic:

$$\frac{(D_1 - D_2)/(r_2 - r_1)}{D_3/(n - r_3)},$$

which we compared to an $F_{r_2-r_1, n-r_3}$ distribution.

The only differences are (i) the arbitrary indexing of models (0,1) or (1,2) and (ii) that in the second form we use a “large” Model 3 to estimate σ^2 by $\hat{\phi} = D_3/(n - r_3)$. In practice the differences will usually be small and the two approaches to variable selection are equivalent.

Exercise 4.6 Start this question by creating an R-Script file entering appropriately chosen R commands, such as those below.

```
# Define variables, your choice of names, for the data
dose = seq(20,100,by=20)
m     = c(23,18,14,23,22)      # This is the Group size
y     = c(20,14,5,6,1)         # This is the number re-treated

# Now plot the data as a scatter plot
par(mfrow=c(1,2))
plot(dose,y/m, pch=16, ylim=c(0,1), xlim=c(0,120),
     ylab="Proportion re-treated", xlab="Dose, mg/kg")
abline(h=c(0,1), lty=2)

# Create a data frame to hold the data for the lm
data = data.frame(dose=dose, prop=y/m)

my.lm = lm(prop~dose, data=data)
abline(my.lm)

# Now get ready for the glm
plot(dose,y/m, pch=16, ylim=c(0,1), xlim=c(0,120),
     ylab="Proportion re-treated", xlab="Dose, mg/kg")
abline(h=c(0,1), lty=2)

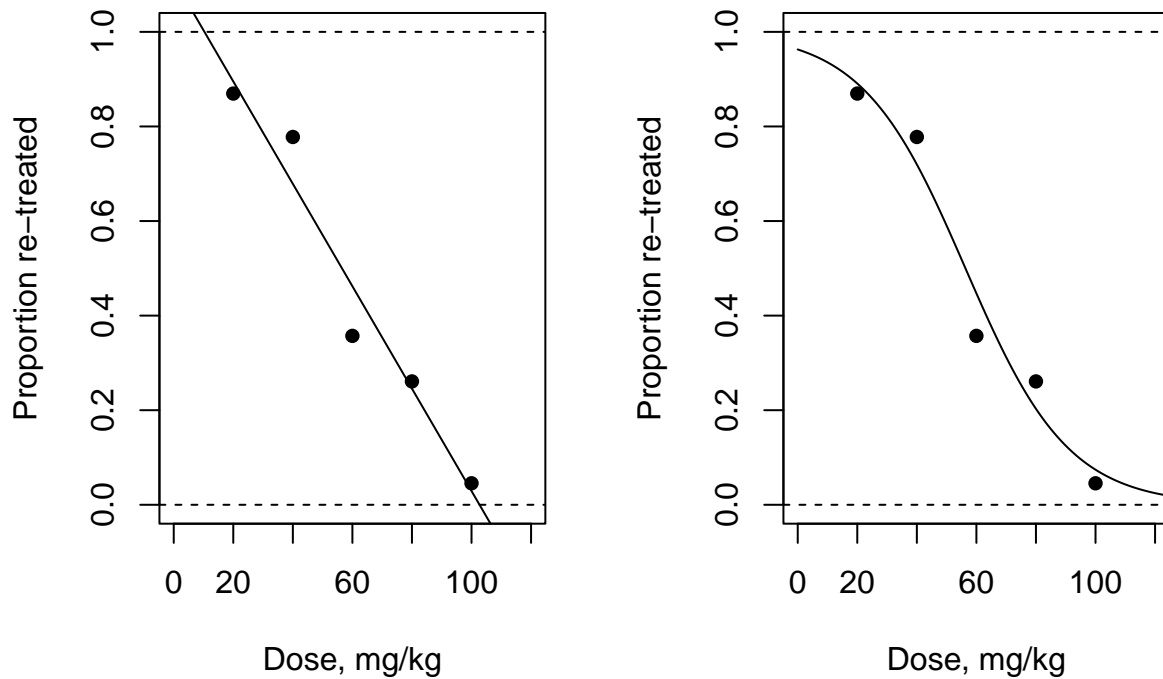
# Now re-define the response as "successes" and "failures"
response = cbind(y,m-y)
my.glm = glm(response~dose, family = "binomial")

# For the fitted curve evaluate the fitted function at many doses
newval = data.frame(dose=seq(0,125,length.out=100))
```

```

predicted = predict(my.glm, newdata = newval, type="response")
lines(newval$dose, predicted)

```



Although both models fit the data well, as the data points are never far from the fitted functions, knowing that proportions can never be outside the interval zero to one, then the binomial logistic model is preferable.

To emphasize the similarity of the fitted model within the range of the data, consider the model residuals as shown in the following graphs.

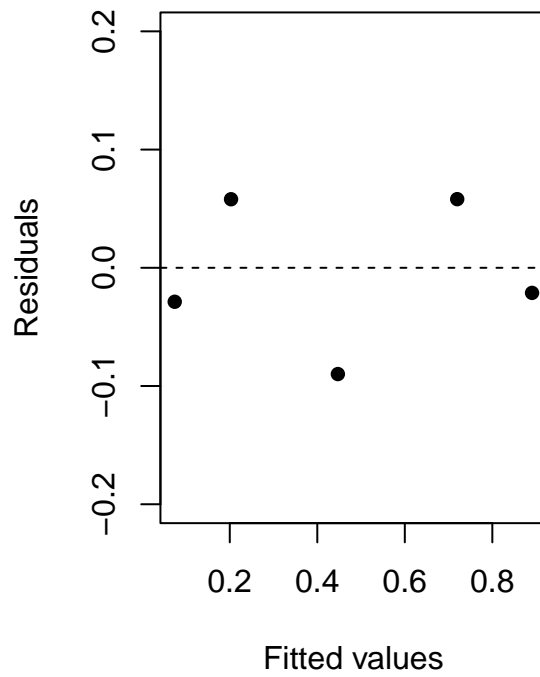
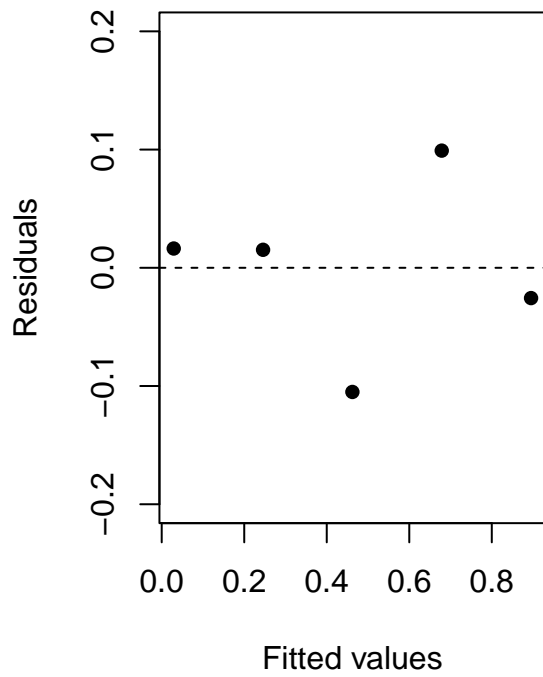
```

par(mfrow=c(1,2))

plot(my.lm$fitted, my.lm$residuals, pch=16, ylim=0.2*c(-1,1),
     xlab="Fitted values", ylab="Residuals")
abline(h=0, lty=2)

glm.resids = residuals(my.glm, type="response")
plot(my.glm$fitted, glm.resids, pch=16, ylim=0.2*c(-1,1),
     xlab="Fitted values", ylab="Residuals")
abline(h=0, lty=2)

```



Although these are not identical, there is no clear *better*. Further, the RSS for the linear model is 0.022 whereas for the logistic model it is 0.0161. From the data alone it is impossible to choose. Knowing the background, however, the better model is clearly the logistic.

End of Solutions to Chapter 4 Exercises