Exercises 7 solutions

1. (a) For the i^{th} observation, the pmf is $\Pr(Y_i = y_i) = \mu_i^{y_i} \exp(-\mu_i)/y_i!$ where $\mu_i = e^{\beta_0+\beta_1 x_i}$.

   Hence the likelihood function is

$$L = \prod_{i=1}^{N} \mu_i^{y_i} \exp(-\mu_i)/y_i!$$

   and the log-likelihood function:

$$\ell = -\sum_{i=1}^{N} \mu_i + \sum_{i=1}^{N} y_i \log \mu_i + \text{ constant} = -\sum_{i=1}^{N} e^{\beta_0+\beta_1 x_i} + \sum_{i=1}^{N} y_i(\beta_0 + \beta_1 x_i) + \text{ constant}$$

   Note that here, as in the following, $\mu_i = e^{\beta_0+\beta_1 x_i}$, which is important because we can observe x_i and want to estimate β_0 and β_1, but the μ_i are not explicitly needed (only to make formulas look simpler).

   (b)

   (i) For the likelihood equations:

$$\frac{\partial \ell}{\partial \beta_0} = -\sum_{i=1}^{N} e^{\beta_0+\beta_1 x_i} + \sum_{i=1}^{N} y_i, \qquad \frac{\partial \ell}{\partial \beta_1} = -\sum_{i=1}^{N} x_i e^{\beta_0+\beta_1 x_i} + \sum_{i=1}^{N} x_i y_i$$

   Putting these derivatives $= 0$ gives likelihood equations.

   (ii) For the elements of information matrix:

$$\mathrm{E}\left(-\frac{\partial^2 \ell}{\partial \beta_0^2}\right) = \sum_{i=1}^{N} e^{\beta_0+\beta_1 x_i}, \mathrm{E}\left(-\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1}\right) = \sum_{i=1}^{N} x_i e^{\beta_0+\beta_1 x_i}, \mathrm{E}\left(-\frac{\partial^2 \ell}{\partial \beta_1^2}\right) = \sum_{i=1}^{N} x_i^2 e^{\beta_0+\beta_1 x_i}.$$

   (c) Information matrix is $\mathcal{I} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}$ where $\mathbf{W}$ is the $N \times N$ diagonal matrix with diagonal elements

$$w_{ii} = \frac{1}{V_i}\left(\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}\right)^2 = \frac{1}{\mu_i}\mu_i^2 = \mu_i \qquad (i = 1, \ldots, N).$$

   With $$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ . & . \\ . & . \\ . & . \\ 1 & x_N \end{pmatrix}$$ then $$\mathcal{I} = \begin{pmatrix} \sum \mu_i & \sum x_i \mu_i \\ \sum x_i \mu_i & \sum x_i^2 \mu_i \end{pmatrix}$$

   as already obtained in part (b)(ii).

2. (a) The fitted values $\hat{\mu}_1, \ldots, \hat{\mu}_5$ are given at the end of the R output.

$$\hat{\mu}_1 = e^{\hat{\beta}_0+\hat{\beta}_1} = 3.2460.$$

   (b) Simply verify that $\sum y_i = \sum \hat{\mu}_i$ (both are 202) and $\sum x_i y_i = \sum x_i \hat{\mu}_i$ (both are 884).

(c) In the information matrix, $\sum \hat{\mu}_i$ and $\sum x_i \hat{\mu}_i$ have already been obtained in part (b). The other quantity is $\sum x_i^2 \hat{\mu}_i = 4042.89$. Then invert the estimated information matrix $\mathcal{I} = \begin{pmatrix} 202 & 884 \\ 884 & 4042.89 \end{pmatrix}$ to verify the estimated covariance matrix.

The standard errors are the square roots of the two diagonal elements of the estimated covariance matrix.

(d) The deviance is obtained from $D = 2[\hat{\ell}(\text{sat}) - \hat{\ell}(\text{model})]$ where $\hat{\ell}$ denotes the log-likelihood evaluated at the MLE under the model indicated in brackets.

The log-likelihood function ℓ was obtained in question 1(a) above. For the purposes of obtaining the deviance however, it is best left expressed in terms of the expected values, ie

$$\ell = -\sum_{i=1}^{N} \mu_i + \sum_{i=1}^{N} y_i \log \mu_i + c$$

where c is the 'constant' term.

Then for the proposed model, the maximum of ℓ is

$$\hat{\ell}(\text{model}) = -\sum_{i=1}^{N} \hat{\mu}_i + \sum_{i=1}^{N} y_i \log \hat{\mu}_i + c$$

and for the saturated model with $\hat{\mu}_i = y_i$,

$$\hat{\ell}(\text{sat}) = -\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} y_i \log y_i + c.$$

Hence the deviance

$$D = 2[\hat{\ell}(\text{sat}) - \hat{\ell}(\text{model})] = 2\sum_{i=1}^{N} [y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)].$$

When there is a constant term in the predictor (as here), the first likelihood equation gives $\sum y_i = \sum \hat{\mu}_i$ and then D reduces to

$$2\sum_{i=1}^{N} y_i \log(y_i/\hat{\mu}_i).$$

As earlier, use the fitted values given at the end of the R output, to verify that $D = 2.0163$.

(e) Another test of H_0: $\beta_1 = 0$ is given by $D_0 - D$ where D_0 is the deviance under H_0 (this uses the test described in Section 3.2.3 (iii) with $p = 2$ and $q = 1$). In the R output, D is the residual deviance and D_0 is the null deviance. Hence the observed value of $D_0 - D$ is 217.11 with 1df, which gives a very small P-value, just as the 'z-test' does. This implies very strong evidence against H_0 which, as already noted in the discussion of R output 2, is not surprising.

(f) Use the procedure described in Section 3.2.2 (i). The standard error of $\hat{\beta}_1$ is in the R output. Hence an approximate 95% confidence interval for β_1 has limits $0.90598 \pm 1.96 \times 0.07574$, and the interval is (0.758, 1.054).

(g) First note that $\text{var}(\hat{\eta}) = \text{var}(\hat{\beta}_0) + 6\text{cov}(\hat{\beta}_0, \hat{\beta}_1) + 9\text{var}(\hat{\beta}_1)$.

Using the estimates of the variances and covariance given in the R output and verified in part (c) above, gives an estimate of $\mathrm{var}(\hat{\eta})$. Then square root this estimated variance to give $\mathrm{se}(\hat{\eta}) = 0.12575$.

Now $\hat{\eta} = \hat{\beta}_0 + 3\hat{\beta}_1 = 2.98937$, hence an approximate 95% confidence interval for η has limits $2.98937 \pm 1.96 \times 0.12575$, and the interval is (2.74291, 3.23583). (This is an application of Section 3.2.2 (ii).)

An approximate 95% confidence interval for the expected response when $x = 3$ is simply obtained by exponentiation of the limits of the above interval, i.e. (15.53, 25.43).

3. (a) The canonical link has $\eta_i = \theta_i$, so the derivatives of ℓ_i are, for $j = 1, \ldots, p$,

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta j} = \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \eta_i}{\partial \beta j} .$$

But $\eta_i = \sum_{r=1}^{p} \beta_r x_{ir}$, so $\partial \eta_i / \partial \beta_j = x_{ij}$. And $\partial \ell_i / \partial \theta_i = \left(y_i - b'(\theta_i)\right) / a_i(\phi)$ (from the definition of the exponential family) $= \left(y_i - \mu_i\right) / a_i(\phi)$. Thus $\partial \ell_i / \partial \beta_j = \left(y_i - \mu_i\right) x_{ij} / a_i(\phi)$, and the required result follows.

(b) If $a_i(\phi) = \phi$ is constant then the likelihood equations are, for $j = 1, \ldots, p$,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{\phi}\right) x_{ij} = 0 \qquad \Rightarrow \qquad \sum_{i=1}^{N} \left(y_i - \mu_i\right) x_{ij} = 0 .$$

If there is a constant term in the model, there must be some j such that $x_{ij} = 1$ for $i = 1, \ldots, N$. For this j, the likelihood equations now read $\sum_{i=1}^{N} \left(y_i - \mu_i\right) = 0$, and the residuals $\{r_i = y_i - \mu_i\}$sum to zero as required.
For any other j, the sample correlation between the residuals and the covariate is proportional to $\sum_{i=1}^{N} \left(r_i - \bar{r}\right)\left(x_{ij} - \bar{x}_j\right)$ in an obvious notation. But we have just shown that $\bar{r} = 0$, so the correlation is proportional to

$$\sum_{i=1}^{N} r_i \left(x_{ij} - \bar{x}_j\right) = \sum_{i=1}^{N} r_i x_{ij} - \bar{x}_j \sum_{i=1}^{N} r_i .$$

The first term here is zero because it is proportional to $\partial \ell / \partial \beta_j$, and the second is zero because the residuals sum to zero. Hence the sample correlation is zero, as required.

4. (a) Since $\mathbf{M}'$ is $p \times N$, $\mathbf{A}$ is $N \times N$ and $\mathbf{V}$ is $N \times 1$, $\mathbf{M}'\mathbf{A}\mathbf{V}$ must be $p \times 1$. Its jth (think $(j,1)$th) entry is $\sum_{i=1}^{N} \mathbf{M}'_{ji} \left[\mathbf{AV}\right]_{i1} = \sum_{i=1}^{N} m_{ij} \left[\mathbf{AV}\right]_{i1}$. And the $(i,1)$ element of $\mathbf{AV}$ is $\sum_{k=1}^{N} a_{ik} v_k$. But $a_{ik} = 0$ except when $k = i$, so the $(i,1)$ element of $\mathbf{AV}$ is $a_{ii} v_i$. Hence the jth element of $\mathbf{M}'\mathbf{AV}$ is $\sum_{i=1}^{N} m_{ij} a_{ii} v_i$, as required.

(b) From equation (3.8) of the lecture notes, the jth element of $\mathbf{U}(\boldsymbol{\beta})$ is

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta_j} &= \sum_{i=1}^{N} \left(y_i - \mu_i\right) \cdot \frac{1}{V_i}\frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \sum_{i=1}^{N} x_{ij} \left(y_i - \mu_i\right) \cdot \frac{1}{V_i} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{\partial \eta_i}{\partial \mu_i} \\
&= \sum_{i=1}^{N} x_{ij} \left(y_i - \mu_i\right) w_{ii} \frac{\partial \eta_i}{\partial \mu_i} = \sum_{i=1}^{N} x_{ij} w_{ii} \left[\eta_i + \left(y_i - \mu_i\right) \frac{\partial \eta_i}{\partial \mu_i} - \eta_i\right] \\
&= \sum_{i=1}^{N} x_{ij} w_{ii} \left[z_i - \eta_i\right] = \sum_{i=1}^{N} x_{ij} w_{ii} z_i - \sum_{i=1}^{N} x_{ij} w_{ii} \eta_i .
\end{aligned}$$

From the result in part (a) we can recognise the first term here as the jth element of $\mathbf{X}'\mathbf{W}\mathbf{z}$, and the second as the jth element of $\mathbf{X}'\mathbf{W}\boldsymbol{\eta}$. Thus $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{X}'\mathbf{W}\mathbf{z} - \mathbf{X}'\mathbf{W}\boldsymbol{\eta}$, and solving $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ is equivalent to setting $\mathbf{X}'\mathbf{W}\mathbf{z} = \mathbf{X}'\mathbf{W}\boldsymbol{\eta}$. Finally, note that $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ to obtain the given result.

(c) The reason the iterative procedure is necessary is that the elements of $\mathbf{W}$ and $\mathbf{z}$ themselves depend on $\boldsymbol{\beta}$. The iterative procedure can thus be seen as a way of starting from an initial guess to get approximate values for $\mathbf{W}$ and $\mathbf{z}$, and using this to improve the estimate of $\boldsymbol{\beta}$. From this improved estimate of $\boldsymbol{\beta}$, updated values for $\mathbf{W}$ and $\mathbf{z}$ can be obtained, and the procedure is iterated until the equality in part (b) is satisfied (at least, to within some tolerance).