AMS 274 - GLM HW 2

Arthur Lui

27 October, 2016

1a)

For $y_i \sim \text{Poisson}(\mu_i)$, the probability density is $f(y_i|\mu_i) = \dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}$. Also, the Exponential Dispersion Family (EDF) has pdf $p(y_i|\theta_i,\phi) = \exp\left\{\dfrac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i,\phi)\right\}$. And the deviance statistic is defined as

$$D = 2\sum_{i=1}^n w_i \left\{y_i\left(\tilde\theta_i - \hat\theta_i\right) - b(\tilde\theta_i) + b(\hat\theta_i)\right\}.$$

First, note that the Poisson distribution is a member of the EDF with

$$\begin{aligned}
\theta_i &= \log(\mu_i)\\
b(\theta_i) &= \exp(\theta_i) = -\mu_i\\
w_i &= \phi = 1
\end{aligned}$$

So,

$$\begin{aligned}
D &= 2\sum_{i=1}^n \left\{y_i\left(\tilde\theta_i - \hat\theta_i\right) - b(\tilde\theta_i) + b(\hat\theta_i)\right\}\\
&= 2\sum_{i=1}^n \left\{y_i\left\{\log(\tilde\mu_i) - \log(\hat\mu_i)\right\} + \tilde\mu_i - \hat\mu_i\right\}\\
&= 2\sum_{i=1}^n \left\{y_i \log\left(\frac{\tilde\mu_i}{\hat\mu_i}\right) + \tilde\mu_i - \hat\mu_i\right\}\\
&= 2\sum_{i=1}^n \left\{y_i \log\left(\frac{y_i}{\hat\mu_i}\right) + y_i - \hat\mu_i\right\}\\
&= 2\sum_{i=1}^n \left\{y_i \log\left(\frac{y_i}{\hat\mu_i}\right)\right\} + 2\sum_{i=1}^n \left\{y_i - \hat\mu_i\right\}
\end{aligned}$$

where $\hat\mu_i = g^{-1}(x_i^T\hat\beta)$, and $\hat\beta$ is the MLE of β under the reduced model.

1b)

In the special case where $g(\cdot) = \log(\cdot)$, we have $\hat{\mu_i} = g^{-1}(x_i^T \hat{\beta}) = \exp(x_i^T \hat{\beta})$, and the deviance is

$$
\begin{aligned}
D &= 2\sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu_i}}\right)\right\} + 2\sum_{i=1}^{n} \{y_i - \hat{\mu_i}\} \\
&= 2\sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu_i}}\right)\right\} + 2\sum_{i=1}^{n} \left\{y_i - \exp(x_i^T \hat{\beta})\right\} \\
&= 2\sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu_i}}\right)\right\} + 2n\bar{y} - 2\sum_{i=1}^{n} \exp(x_i^T \hat{\beta}) \\
&= 2\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\hat{\mu_i}}\right)
\end{aligned}
$$

2a)

$$
\begin{aligned}
f(y_i|\mu_i, \nu) &= \frac{(\nu/\mu_i)^\nu y_i^{\nu-1}}{\Gamma(\nu)} \exp(-\nu y_i/\mu_i) \\
&= \exp\left\{ -\frac{\nu y_i}{\mu_i} + (\nu - 1)\log y_i + \nu \log\frac{\nu}{\mu_i} - \log\Gamma(\nu)\right\} \\
&= \exp\left\{ \frac{y_i \mu_i^{-1} - \log\dfrac{\mu_i}{\nu}}{-\nu^{-1}} + (\nu-1)\log(y_i) - \log\Gamma(\nu)\right\}
\end{aligned}
$$

which is a member of the EDF with

$$
\begin{aligned}
\theta_i &= \mu_i^{-1} \\
b(\theta_i) &= \log\frac{\mu_i}{\nu} = -\log(\nu\theta_i) \\
w_i &= -1 \\
\phi &= \nu^{-1}
\end{aligned}
$$

2b)

The scaled deviance is

$$
\begin{aligned}
D^* &= \frac{2}{\phi}\sum_{i=1}^{n} w_i \left\{ y_i\left(\tilde{\theta}_i - \hat{\theta}_i\right) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right\} \\
&= -\frac{2}{\phi}\sum_{i=1}^{n} \left\{ y_i\left(\tilde{\mu_i}^{-1} - \hat{\mu_i}^{-1}\right) - \log\frac{\tilde{\mu_i}}{\nu} + \log\frac{\hat{\mu_i}}{\nu} \right\} \\
&= -\frac{2}{\phi}\sum_{i=1}^{n} \left\{ y_i\left(\tilde{\mu_i}^{-1} - \hat{\mu_i}^{-1}\right) + \log\frac{\hat{\mu_i}}{\tilde{\mu_i}} \right\} \\
&= -\frac{2}{\phi}\sum_{i=1}^{n} \left\{ y_i\left(y_i^{-1} - \hat{\mu_i}^{-1}\right) + \log(y_i\hat{\mu_i}) \right\} \\
&= -\frac{2}{\phi}\sum_{i=1}^{n} 1 - \frac{y_i}{\hat{\mu_i}} + \log(y_i\hat{\mu_i}) \\
&= \frac{1}{\phi} D
\end{aligned}
$$

where $D = 2\sum_{i=1}^{n}\left\{\frac{y_i}{\hat{\mu_i}} - \log(y_i\hat{\mu_i}) - 1\right\}$ is the deviance, and $\hat{\mu_i} = g^{-1}(x_i^T\hat{\beta})$.

3a)

```
Call:
glm(formula = faults ~ length, family = "poisson", data = fabric)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.74127  -1.13312  -0.03904   0.66179   3.07446

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.9717506  0.2124693   4.574 4.79e-06 ***
length      0.0019297  0.0003063   6.300 2.97e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  61.758  on 30  degrees of freedom
AIC: 189.06
```

```
Number of Fisher Scoring iterations: 4
```

3b)

```
Call:
glm(formula = faults ~ length, family = "quasipoisson", data = fabric)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.74127  -1.13312  -0.03904   0.66179   3.07446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.9717506  0.3095033   3.140 0.003781 **
length      0.0019297  0.0004462   4.325 0.000155 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 2.121965)

    Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  61.758  on 30  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

3c)

Likelihood:

$$\hat{\eta_0} = x_0^T \hat{\beta}$$

$$\begin{aligned} \mathrm{E}\left[\hat{\eta_0}\right] &= \mathrm{E}\left[x_0^T \hat{\beta}\right] \\ &= x_0^T \mathrm{E}\left[\hat{\beta}\right] \\ &= x_0^T \beta \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(\hat{\eta_0}) &= \mathrm{Var}(x_0^T \hat{\beta}) \\ &= \mathrm{Var}(x_0^T \hat{\beta}) \end{aligned}$$

$$
\begin{aligned}
&= x_0^T \mathrm{Var}(\hat{\beta}) x_0 \\
&= x_0^T J^{-1}(\beta) x_0
\end{aligned}
$$

Therefore, a point estimate for η_0 is $\hat{\eta_0}$, and an interval estimate is $x_0^T\hat{\beta} \pm z_{.025}\sqrt{x_0^T J^{-1}(\hat{\beta})x_0}$.

3b)

Quasi-Likelihood:

$$\tilde{\eta_0} = x_0^T \tilde{\beta}$$

$$\mathrm{E}\left[\tilde{\eta_0}\right] = x_0^T \beta$$

$$\mathrm{Var}(\hat{\eta_0}) = \tilde{\phi} x_0^T J^{-1}(\beta) x_0$$

Therefore, a point estimate for η_0 is $\tilde{\eta_0}$, and an interval estimate is $x_0^T\tilde{\beta} \pm z_{.025}\sqrt{\tilde{\phi} x_0^T J^{-1}(\hat{\beta})x_0}$.