Advanced Methods Homework 2

Bohao Tang

1. Exercise 4.4:

For the model we considered in this chapter, we have log-likelihood

$$l(y_i;\theta_i,\phi_i) = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + C(y_i,\phi)$$

where we only model $E(y_i) = \mu_i = b'(\theta_i)$ to be $g(\mu_i) = \eta_i = X_i^T\cdot\beta$

Therefore we have likelihood equation

$$\sum_i \frac{\partial l}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = 0 \quad \forall j$$

$$\Leftrightarrow \sum_i \frac{y_i - b'(\theta_i)}{a(\phi)}\cdot\left(\frac{\partial \mu_i}{\partial \theta_i}\right)^{-1}\frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \forall j$$

$$\Leftrightarrow \sum_i \frac{y_i - \mu_i}{a(\phi)\, b''(\theta_i)}\cdot\frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \forall j$$

Since $var(y_i) = a(\phi)\, b''(\theta_i)$

$$\Rightarrow \sum_i \frac{y_i - \mu_i}{var(y_i)}\frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \forall j$$

And if we deal $var(y_i)$ as constant then

$$\min\ \Sigma_i\left[(y_i-\mu_i)^2/var(y_i)\right]$$

$$\Leftrightarrow \sum_i \frac{\partial}{\partial \beta_j}\left[(y_i-\mu_i)^2/var(y_i)\right] = 0$$

$$\Leftrightarrow \sum_i \frac{\partial}{\partial \mu_i}(y_i-\mu_i)^2\,\frac{\partial \mu_i}{\partial \beta_j}\cdot\frac{1}{var(y_i)} = 0 \quad \forall j$$

$$\Leftrightarrow \sum_i (y_i-\mu_i)/var(y_i)\cdot\frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \forall j$$

2. Exercise 4.13:

Log likelihood

$$L(y_i;\mu_i) = -\frac{(y_i-\mu_i)^2}{2\sigma^2} + \text{constant}.$$

$$\Rightarrow \text{Deviance} = \sum_i \frac{(y_i-\hat{\mu}_i)^2}{\sigma^2}$$

But since for nomal with given σ, $var(y_i)=\sigma^2$

$$\Rightarrow \sum_i \frac{(y_i-\hat{\mu}_i)^2}{\sigma^2} = \sum_i \frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \text{Pearson chi-square statistics}$$

For model $M_0 \subset M_1$, suppose $\hat{\mu}_{i0}$ is μ_i estimated under M_0
and $\hat{\mu}_{i1}$ for estimated μ_i under M_1

then $Dev(M_0) - Dev(M_1)$

$$= \sum_i \frac{(y_i-\hat{\mu}_{i0})^2}{\sigma^2} - \frac{(y_i-\hat{\mu}_{i1})^2}{\sigma^2}$$

$$= \sum_i \frac{[2y_i-(\hat{\mu}_{i0}+\hat{\mu}_{i1})]\cdot(\hat{\mu}_{i1}-\hat{\mu}_{i0})}{\sigma^2}$$

3. Exercise 4.16:

(a) log likelihood for binomial model is

$$L = \log\left[\binom{n}{y_i} p_i^{y_i}(1-p_i)^{n-y_i}\right] = \log\binom{n}{y_i} + y_i \log\frac{p_i}{1-p_i} + n\log(1-p_i)$$

Since $\mu_i = E(y_i) = np_i$ $\Rightarrow L(y_i;\mu_i) = y_i \log\frac{\mu_i}{n-\mu_i} + n\log(1-\frac{\mu_i}{n}) + \text{constant}$

$$\Rightarrow d_i = 2\left[y_i \log\frac{y_i}{n-y_i} + n\log(n-y_i) - y_i\log\frac{\hat{\mu}_i}{n-\hat{\mu}_i} - n\log(n-\hat{\mu}_i)\right]$$

$$= 2\left[y_i\log\frac{y_i}{\hat{\mu}_i} + (n-y_i)\log\frac{n-y_i}{n-\hat{\mu}_i}\right]$$

$$\Rightarrow \text{Deviance residual} = \sqrt{2\left[y_i\log\frac{y_i}{\hat{\mu}_i} + (n-y_i)\log\frac{n-y_i}{n-\hat{\mu}_i}\right]} \cdot \text{sign}(y_i-\hat{\mu}_i)$$

(b) Log likelihood for Poisson model is

$$L = \log\left[e^{-\lambda_i}\frac{(\lambda_i)^{y_i}}{y_i!}\right] = -\log y_i! - \lambda_i + y_i \log \lambda_i$$

Since $\mu_i = E(y_i) = \lambda_i$

we have $\cancel{L}\cdot L[y_i;\mu_i] = y_i \log \mu_i - \mu_i + \text{constant}$

$$\Rightarrow d_i = 2\left[y_i \log y_i - y_i - y_i \log \hat{\mu}_i + \hat{\mu}_i\right]$$

$\Rightarrow$ Deviance residual $= \sqrt{2\left[y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i)\right]} \cdot \text{sign}(y_i - \hat{\mu}_i)$

4: (a) $\log f(y)$

$$= (\alpha-1)\log y - y/\beta - \alpha\log\beta - \log\Gamma(\alpha)$$

~~$\Rightarrow$ natural parameter $\theta =$~~

$$= (-\alpha)\left[y\cdot\frac{1}{\alpha\beta} + \log\alpha\beta\right] + \alpha\log\alpha + (\alpha-1)\log y - \log\Gamma(\alpha)$$

$\Rightarrow$ natural parameter $\theta = \frac{1}{\alpha\beta}$

Cumulant function $b(\theta) = -\log\frac{1}{\theta} = \log\theta$

dispersion parameter $\phi = \alpha$ (or $-\alpha$ or $\frac{1}{-\alpha}$)

variance function $V(\mu) = \alpha\beta^2 = \alpha^2\beta^2\cdot\frac{1}{\alpha} = \mu^2/\alpha$

(b) Log likelihood function $(\theta = \frac{1}{\alpha\beta}; \phi = \alpha)$

$$L = (y\cdot\theta - \log\theta)/(-\tfrac{1}{\phi}) + C(y,\phi)$$

If we model $g(\mu_i) = \eta_i = X_i^T\cdot\vec{\beta}$

Then

We have score equation for $\vec{\beta}$ to be

$$\sum_i \partial \frac{y_i\theta_i - \log\theta_i}{-\frac{1}{\phi}} \Big/ \partial\theta_i \cdot \frac{\partial\theta_i}{\partial\mu_i} \cdot \frac{\partial\mu_i}{\partial\eta_i} \cdot \frac{\partial\eta_i}{\partial\beta_j} = 0 \quad \forall j$$

$$\Rightarrow \sum_i (-\phi)\cdot\left(y_i - \frac{1}{\theta_i}\right)\cdot\left(-\frac{1}{\mu_i^2}\right)\cdot\frac{\partial\mu_i}{\partial\eta_i}\cdot x_{ij} = 0 \quad \forall j$$

$$\Rightarrow \sum_i \frac{(y_i-\mu_i)\,x_{ij}}{var(y_i)}\frac{\partial\mu_i}{\partial\eta_i} = 0 \quad \forall j$$

$$\Leftrightarrow \sum_i \frac{(y_i - \alpha_i\beta_i)x_{ij}}{\alpha_i\beta_i^2}\frac{\partial\mu_i}{\partial\eta_i} = 0 \quad \forall j$$

And likelihood function for ϕ is

$$\sum_i \frac{\partial}{\partial\phi}\left[(-\phi)\left[y_i\theta_i + \log\frac{1}{\theta_i}\right] + \phi\log\phi + (\phi-1)\log y_i - \log\Gamma(\phi)\right] = 0$$

$$\Leftrightarrow \sum_i \left[\log\theta_i - y_i\theta_i + \log\phi + 1 + \log y_i - \frac{\Gamma'(\phi)}{\Gamma(\phi)}\right] = 0.$$

Also, the information matrix for $\vec{\beta}$ is

$$I_{rs} = -E\left[\frac{\partial}{\partial\beta_s}\sum_i(-\phi)\cdot\left(y_i-\frac{1}{\theta_i}\right)\cdot\left(-\frac{1}{\mu_i^2}\right)\cdot\frac{\partial\mu_i}{\partial\eta_i}\cdot x_{ir}\right]$$

$$= -E\left[\sum_i(-\phi)\left(y_i-\frac{1}{\theta_i}\right)\frac{\partial}{\partial\beta_s}\left(-\frac{1}{\mu_i^2}\cdot\frac{\partial\mu_i}{\partial\eta_i}x_{ir}\right)\right]$$

$$-E\left[\sum_i\frac{\phi}{\mu_i^2}\frac{\partial\mu_i}{\partial\eta_i}x_{ir}\frac{\partial}{\partial\beta_s}\left(y_i-\frac{1}{\theta_i}\right)\right]$$

$$= -E\left[\sum_i\frac{\phi}{\mu_i^2}\frac{\partial\mu_i}{\partial\eta_i}x_{ir}\cdot\frac{\partial\left(y_i-\frac{1}{\theta_i}\right)}{\partial\theta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_s}\right]$$

$$= -E\left[\sum_i\frac{\phi}{\mu_i^2}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ir}x_{is}\cdot\left(-\frac{1}{\mu_i^2}\right)\cdot\frac{1}{\theta_i^2}\right]$$

$$= \sum_i x_{ir}x_{is}\cdot\frac{\phi}{\mu_i^4}\cdot\mu_i^2\cdot\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 = \sum_i x_{ir}x_{is}\cdot\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\Big/var(y_i) = \sum_i\frac{x_{ir}x_{is}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2}{\alpha_i\beta_i^2}.$$

5: Exercise 4.27.

(a) $\log y - \log \mu \approx (\log y)'|_{y=\mu}(y-\mu) + o(|y-\mu|)$

$= \frac{y-\mu}{\mu} + o(|y-\mu|)$

$\Rightarrow var[\log(y)] \approx var \frac{y-\mu}{\mu} = \frac{var\ y}{\mu^2} = \text{constant}$

if $|y-\mu|$ is small overall, which means a small σ for y.

(b) If $\log(y_i) \sim N(\mu_i, \sigma^2)$

Then $E\ y_i = E\ e^{\log y_i} = E\ e^{1 \cdot \log y_i} = M_{\log y_i}(1)$ where M is MGF.

There fore $E\ y_i = e^{[\mu_i + \frac{\sigma^2}{2}]}$

$\Rightarrow \log E[y_i] = E[\log y_i] + \frac{\sigma^2}{2}$

(C) For linear model for $\log(y_i)$

we have $\log y_i = \mu_i + \varepsilon$ where $\varepsilon \sim$ normal distribution.

Therefore $y_i = e^{\mu_i} \cdot e^{\varepsilon}$

Since $\varepsilon \sim$ normal $e^{\mu_i} \cdot e^{\varepsilon}$ has a median of $e^{\mu_i} \cdot e^{0} = e^{\mu_i}$

$\Rightarrow \hat{\mu}_i$ is a fitted value of μ_i then $e^{\hat{\mu}_i}$ is a fitted median of y_i

If ~~a~~ $\log y_i \sim$ Normal, then y_i have a very heavy tail for $x \to +\infty$, therefore the median may be more robust or more representative for the whole data also the log normal model often base on a normal model, if we know that the variance of normal model is constant but unknown then we can't ~~es~~ directly estimate the mean for its exponential but the median is estimatable.

6: Exercise 6.1:

We have the log likelihood be.

$$L = \log\left[\prod_{i=1}^{C-1} \pi_i^{y_i} \cdot \pi_c^{1-\sum_{i=1}^{C-1} y_i}\right]$$

$$= \log \pi_c + \sum_{i=1}^{C-1} y_i \log\frac{\pi_i}{\pi_c}.$$

$$= \vec{y} \cdot \overrightarrow{\text{logits}} + \log \pi_c$$

where $\overrightarrow{\text{logits}} = \left(\log\frac{\pi_1}{\pi_c}, \cdots, \log\frac{\pi_{c-1}}{\pi_c}\right)$ is the baseline-category logits

so the model is a $(c-1)$-parameter exponential dispersion family with baseline-category logits as natural parameter

7: Exercise 6.8:

(a) It's more likely to treat the response as ordinal.

Because now the response is not exchangeable, for example

if $\frac{\pi_{i1}}{\pi_{ic}} > 1$ then $\log\frac{\pi_{ij}}{\pi_{ic}} = x_i \beta_j = j x_i \beta = j \log\frac{\pi_{i1}}{\pi_{ic}} > (j-1)\log\frac{\pi_{i1}}{\pi_{ic}} = \log\frac{\pi_{ij-1}}{\pi_{ic}}.$

So here π_{ij} behave more like a cumulate probability. So the response is more likely to be ordinal.

(b). We have $\log\frac{\pi_{ir}}{\pi_{is}} = (r-s)\vec{x}_i \cdot \vec{\beta}$

When we model that $\log\frac{\pi_{i,j+1}}{\pi_{ij}} = \vec{x}_i \cdot \vec{\beta}$

then $\log\frac{\pi_{ir}}{\pi_{is}} = (r-s)\vec{x}_i \cdot \vec{\beta}$ = that in model of (a)