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

$$\sum_i \frac{\partial b}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = 0 \qquad \forall j$$

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

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

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

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

And if we deal $\mathrm{var}(y_i)$ as constant then

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

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

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

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

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 normal with given σ, $\text{var}(y_i)=\sigma^2$

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

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 $\text{Dev}(M_0) - \text{Dev}(M_1)$

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

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

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

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

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

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

(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 $L[y_i; \mu_i] = y_i \log \mu_i - \mu_i + \text{constant}$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5: Exercise 4.27.

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

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

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

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 $y_i \approx e^{\mu_i} \cdot e^{\varepsilon}$

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 $\log y_i \sim$ 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 $\overrightarrow{\text{logits}} = (\log \frac{\pi_1}{\pi_c}, \cdots, \log \frac{\pi_{c-1}}{\pi_c})$ 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 $\log \frac{\pi_{ij}}{\pi_{ic}} = x_i \beta_j = j x_i \beta = j \log \frac{\pi_{i1}}{\pi_{ic}} \Rightarrow \log \frac{\pi_{ij}}{\pi_{ic}} = \log \frac{\pi_{ij-1}}{\pi_{ic}}$.

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 $\log \frac{\pi_{ij+1}}{\pi_{ij}} = \vec{x}_i \cdot \vec{\beta}$

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

Coding

Bohao Tang

February 25, 2018

6.18

Here we do a similiar thing as in Example 6.3.2. The basic main effect model is:

$$\log \frac{\pi_{ij}}{\pi_{i1}} = \beta_{j0} + \beta_{j1} s_i + \beta_{j2} g_i + \beta_{j3} z_i^O + \beta_{j4} z_i^T + \beta_{j5} z_i^G$$

For $j = 2, 3, 4, 5$. Here the notation is the same as in Eample 6.3.2. And g_i is the gender for alligator i. Then we fit the model:

```
library("VGAM")

Alligators = read.table("Alligators2.txt", header = TRUE)

fit <- vglm(formula = cbind(y2,y3,y4,y5,y1) ~ size + gender + factor(lake),
            family = multinomial, data = Alligators)

summary(fit)
```

```
##
## Call:
## vglm(formula = cbind(y2, y3, y4, y5, y1) ~ size + gender + factor(lake),
##     family = multinomial, data = Alligators)
##
##
## Pearson residuals:
##                          Min      1Q   Median     3Q   Max
## log(mu[,1]/mu[,5]) -1.3218 -0.4611  0.01054 0.3810 1.866
## log(mu[,2]/mu[,5]) -0.7033 -0.5751 -0.35511 0.2610 2.064
## log(mu[,3]/mu[,5]) -1.1985 -0.5478 -0.22421 0.3678 3.478
## log(mu[,4]/mu[,5]) -1.6945 -0.2893 -0.10807 1.1236 1.367
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept):1    -2.9477     0.6741  -4.373 1.23e-05 ***
## (Intercept):2    -1.7295     0.7928  -2.182  0.02914 *
## (Intercept):3    -1.1266     0.6696  -1.682  0.09248 .
## (Intercept):4    -0.9547     0.5295  -1.803  0.07138 .
## size:1            1.3363     0.4112   3.250  0.00116 **
## size:2           -0.5570     0.6466  -0.861  0.38898
## size:3           -0.7302     0.6523  -1.120  0.26292
## size:4            0.2906     0.4599   0.632  0.52751
## gender:1         -0.4630     0.3955  -1.171  0.24180
## gender:2         -0.6276     0.6853  -0.916  0.35978
## gender:3         -0.6064     0.6888  -0.880  0.37867
## gender:4         -0.2526     0.4663  -0.542  0.58810
## factor(lake)2:1   2.6937     0.6693   4.025 5.70e-05 ***
## factor(lake)2:2   1.4008     0.8105   1.728  0.08393 .
## factor(lake)2:3  -1.1256     1.1923  -0.944  0.34513
```

```
## factor(lake)2:4  -0.7405     0.7421  -0.998  0.31837
## factor(lake)3:1   2.9363     0.6874   4.272 1.94e-05 ***
## factor(lake)3:2   1.9316     0.8253   2.340  0.01926 *
## factor(lake)3:3   0.6617     0.8461   0.782  0.43415
## factor(lake)3:4   0.7912     0.5879   1.346  0.17840
## factor(lake)4:1   1.7805     0.6232   2.857  0.00428 **
## factor(lake)4:2  -1.1295     1.1928  -0.947  0.34369
## factor(lake)4:3  -0.5753     0.7952  -0.723  0.46943
## factor(lake)4:4  -0.7666     0.5686  -1.348  0.17756
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  4
##
## Names of linear predictors:
## log(mu[,1]/mu[,5]), log(mu[,2]/mu[,5]), log(mu[,3]/mu[,5]), log(mu[,4]/mu[,5])
##
## Residual deviance: 50.2637 on 40 degrees of freedom
##
## Log-likelihood: -73.3221 on 40 degrees of freedom
##
## Number of iterations: 5
##
## No Hauck-Donner effect found in any of the estimates
##
## Reference group is level  5  of the response
```

The residual deviance 50.2637 with $df = 40$ doesn't give much evidence against our model, because the p value is not small:

```
1 - pchisq(50.2637,40)
```

```
## [1] 0.1281848
```

Drop `size` or `gender` will cause a significant loss while no significant loss to drop `gender`. But dropping `gender` we will get the same result as in Example 6.3.2. Therefore here we accept our original model.

We have:

$$\log \frac{\hat{\pi}_{i2}}{\hat{\pi}_{i1}} = -2.9477 + 1.3363 s_2 - 0.463 g_2 + 2.6937 z_2^O + 2.9363 z_2^T + 1.7805 z_2^G$$

For a given lake and gender, for small alligators the estimated odds that primary food choice was invertebrates instead of fish are $\exp(1.3363) = 3.8$ times the estimated odds for large alligators. The estimated effect is imprecise, as the Wald 95% con dence interval is $\exp(1.3363 \pm 1.96(0.4412)) = (1.602, 9.034)$. For a given lake and size, for gender = 0 alligators the estimated odds that primary food choice was invertebrates instead of fish are $\exp(0.463) = 1.5888$ times the estimated odds for gender=1 alligators. The estimated effect is also imprecise, as the Wald 95% con dence interval is $\exp(0.463 \pm 1.96(0.3955)) = (0.7318, 3.44933)$. The lake effects indicate that the estimated odds that the primary food choice was invertebrates instead of fish are relatively higher at lakes Ocklawaha, Trafford and George than they are at Lake Hancock.

7.31

a.

We fit the model like below:

$$\log \mu_i = \beta_0 + \beta_1 x_i$$

```r
Homicide = read.table("Homicide.txt", header = TRUE)

fit.poi <- glm(count ~ race, family=poisson, data=Homicide)

summary(fit.poi)
```

```
##
## Call:
## glm(formula = count ~ race, family = poisson, data = Homicide)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0218  -0.4295  -0.4295  -0.4295   6.1874
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.38321    0.09713  -24.54   <2e-16 ***
## race         1.73314    0.14657   11.82   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 962.80  on 1307  degrees of freedom
## Residual deviance: 844.71  on 1306  degrees of freedom
## AIC: 1122
##
## Number of Fisher Scoring iterations: 6
```

Then the $\hat{\beta} = 1.73314$ means the estimated log ratio of mean for black and white people is 1.73314.

b.

Maybe the time interval to collect the response for each people is not fixed but rather a random variable. Then the Poisson assumption is inadequate.

Here we fit the corresponding negative binomial model:

```r
library("MASS")

fit.nb <- glm.nb(count ~ race, data=Homicide)

summary(fit.nb)
```

```
##
## Call:
## glm.nb(formula = count ~ race, data = Homicide, init.theta = 0.2023119205,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7184  -0.3899  -0.3899  -0.3899   3.5072
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  -2.3832     0.1172 -20.335  < 2e-16 ***
## race          1.7331     0.2385   7.268 3.66e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.2023) family taken to be 1)
##
##     Null deviance: 471.57  on 1307  degrees of freedom
## Residual deviance: 412.60  on 1306  degrees of freedom
## AIC: 1001.8
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  0.2023
##           Std. Err.:  0.0409
##
##  2 x log-likelihood:  -995.7980
```

Here, the AIC for NB2 is much smaller than that for Poisson (1001.8 vs 1122). The much larger SE(0.2385 vs 0.14657) also the estimated NB2 dispersion parameter is $1/0.2023 = 4.943$ is much larger than 1. We can notice that Poisson model had overdispersion.

c.

Here we are calculating the Wald 95% CI for β. We have:

For Poisson model:

```
print(exp(1.7331+0.14657*c(qnorm(0.025),qnorm(0.975))))
```

```
## [1] 4.245366 7.541129
```

For Negative Binomial model:

```
print(exp(1.7331+0.2385*c(qnorm(0.025),qnorm(0.975))))
```

```
## [1] 3.545391 9.029991
```

Since we have enough evidence that Poisson GLM have an overdispersion, we rather believe the larger interval for $\hat{\beta}$. Also the AIC shows that NB GLM is a better model, so we choose (3.5,9.0) to be our CI.

overdispersion

The only thing we need to do is to estimate the overdispersion term σ^2 for $var(Y_i) = \sigma^2\mu_i$.

We have:

```
n = length(Homicide$Obs)
p = 1
sigma_2 = 1 / (n-p) * sum((Homicide$count - fitted(fit.poi))^2 / fitted(fit.poi))
sigma_2
```

```
## [1] 1.744356
```

Therefore the ture estimate variance for $\hat{\beta}$ is 0.14657 * 1.744356 = 0.2556703. So the CI is

```
print(exp(1.7331+0.2556703*c(qnorm(0.025),qnorm(0.975))))
```

```
## [1] 3.428063 9.339050
```

10.

(i)

For the question one, if the books are consistent, than they follow the main effect model as:

$$\log \mu_i = \beta_0 + \beta_1 b_i^{SI} + \beta_2 b_i^{SS} + \eta_1 w_i^{an} + \eta_2 w_i^{that} + \eta_3 w_i^{this} + \eta_4 w_i^{with} + \eta_5 w_i^{without}$$

Here μ_i is the mean count for i^{th} data. $b_i^{()}, w_i^{()}$ are the flag for book and word. If this model holds, then the relative appearence (ratio of mean) for each word is $\eta_i - \eta_j$ ($\eta_0 = 0$ for word `a`) which does not influenced by the book flag β_i. So the books are consistent. Therefore we compare this model with the full model to make the decision.

```
library(tidyverse)
Jane = read_csv("Ex0205.csv")

Jane_herself = Jane %>% filter(Book != "SanditonII")

fit.jane11.full = glm(Count ~ Book:Word, family=poisson, data=Jane_herself)
fit.jane11.homo = glm(Count ~ factor(Book) + factor(Word),
                      family=poisson, data=Jane_herself)
summary(fit.jane11.homo)
```

```
##
## Call:
## glm(formula = Count ~ factor(Book) + factor(Word), family = poisson,
##     data = Jane_herself)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.71399  -0.48124   0.00187   0.49657   1.55116
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     5.23521    0.05995  87.331  < 2e-16 ***
## factor(Book)SanditonI          -0.77851    0.08499  -9.160  < 2e-16 ***
## factor(Book)Sense&Sensibility -0.15985    0.07028  -2.274   0.0229 *
## factor(Word)an                 -1.94591    0.13577 -14.333  < 2e-16 ***
## factor(Word)that               -0.60921    0.08088  -7.532 4.98e-14 ***
## factor(Word)this               -1.61870    0.11803 -13.714  < 2e-16 ***
## factor(Word)with               -0.99164    0.09228 -10.746  < 2e-16 ***
## factor(Word)without            -2.43546    0.16917 -14.396  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 709.289  on 17  degrees of freedom
## Residual deviance:  12.587  on 10  degrees of freedom
## AIC: 127.34
##
## Number of Fisher Scoring iterations: 4
```

Then we find that the AIC for main effect model is less than that of full model (127.34 vs 134.76). Also the residul devience for main effect model is 12.587 with df=10, which has a p value of 0.2476881, no significant loss. So we may choose the main effect model, which means that the books are consistent.

For the question two, we also compare the main effect model:

$$\log \mu_i = \beta_0 + \beta_1 o_i + \eta_1 w_i^{an} + \eta_2 w_i^{that} + \eta_3 w_i^{this} + \eta_4 w_i^{with} + \eta_5 w_i^{without}$$

with the full model, where here the o_i is the flag for whether the book is written by Jane. Then we have:

```
Jane_compare = Jane %>%
               mutate(Original = (Book != "SanditonII"))
fit.jane12.full <- glm(Count ~ Original:Word,
                       family = poisson, data = Jane_compare)
fit.jane12.homo <- glm(Count ~ factor(Original) + factor(Word),
                       family = poisson, data = Jane_compare)
summary(fit.jane12.full)
```

```
##
## Call:
## glm(formula = Count ~ Original:Word, family = poisson, data = Jane_compare)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.2455  -0.7782   0.0000   1.1977   3.2897
##
## Coefficients: (1 not defined because of singularities)
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  2.5390     0.1622  15.651  < 2e-16 ***
## OriginalFALSE:Worda          1.8799     0.1959   9.598  < 2e-16 ***
## OriginalTRUE:Worda           2.4355     0.1692  14.396  < 2e-16 ***
## OriginalFALSE:Wordan         0.8283     0.2466   3.359 0.000781 ***
## OriginalTRUE:Wordan          0.4895     0.2060   2.376 0.017492 *
## OriginalFALSE:Wordthat       0.5521     0.2679   2.061 0.039329 *
## OriginalTRUE:Wordthat        1.8262     0.1748  10.448  < 2e-16 ***
## OriginalFALSE:Wordthis       0.1691     0.3049   0.554 0.579253
## OriginalTRUE:Wordthis        0.8168     0.1948   4.193 2.75e-05 ***
## OriginalFALSE:Wordwith       1.2222     0.2226   5.490 4.03e-08 ***
## OriginalTRUE:Wordwith        1.4438     0.1804   8.006 1.19e-15 ***
## OriginalFALSE:Wordwithout   -1.1527     0.5257  -2.193 0.028319 *
## OriginalTRUE:Wordwithout        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 878.86  on 23  degrees of freedom
## Residual deviance: 108.60  on 12  degrees of freedom
## AIC: 261.19
##
## Number of Fisher Scoring iterations: 4
```

```
summary(fit.jane12.homo)
```

```
##
## Call:
## glm(formula = Count ~ factor(Original) + factor(Word), family = poisson,
##     data = Jane_compare)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -4.5646  -1.5535   0.0266   1.7490   3.6245
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            4.42531    0.07882  56.148  < 2e-16 ***
## factor(Original)TRUE   0.54789    0.07801   7.023 2.16e-12 ***
## factor(Word)an        -1.73718    0.11368 -15.281  < 2e-16 ***
## factor(Word)that      -0.69508    0.07622  -9.119  < 2e-16 ***
## factor(Word)this      -1.63292    0.10879 -15.010  < 2e-16 ***
## factor(Word)with      -0.92992    0.08268 -11.247  < 2e-16 ***
## factor(Word)without   -2.51037    0.16045 -15.646  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 878.86  on 23  degrees of freedom
## Residual deviance: 140.34  on 17  degrees of freedom
## AIC: 282.92
##
## Number of Fisher Scoring iterations: 4
```

Now the residual deviance is 140.34-108.6=31.74 with df=17-12=5. So the p value is 6.688744e-06, there is a significant difference between the style of Jane and the ghost writer.

For question three, the model is the same as in question two. But the data for Jane here only contains SantitonI:

```
Jane_s12 = Jane %>%
            filter(Book=="SanditonI" | Book=="SanditonII")
fit.jane13.full <- glm(Count ~ factor(Book):factor(Word),
                       family = poisson, data = Jane_s12)
fit.jane13.homo <- glm(Count ~ factor(Book) + factor(Word),
                       family = poisson, data = Jane_s12)
summary(fit.jane13.homo)
```

```
##
## Call:
## glm(formula = Count ~ factor(Book) + factor(Word), family = poisson,
##     data = Jane_s12)
##
## Deviance Residuals:
##        1         2         3         4         5         6         7
##  0.77745  -2.26305  -0.05810   1.24308  -1.39358   1.02240  -0.81138
##        8         9        10        11        12
##  1.95671   0.05868  -1.36807   1.31126  -1.19731
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)              4.53675    0.08873  51.131  < 2e-16 ***
## factor(Book)SanditonII -0.03015    0.10026  -0.301    0.764
## factor(Word)an         -1.52606    0.17445  -8.748  < 2e-16 ***
## factor(Word)that       -1.13740    0.14961  -7.602 2.91e-14 ***
## factor(Word)this       -1.81374    0.19690  -9.212  < 2e-16 ***
## factor(Word)with       -0.95226    0.13971  -6.816 9.37e-12 ***
```

```
## factor(Word)without    -2.57588    0.27724  -9.291  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 259.514  on 11  degrees of freedom
## Residual deviance:  19.777  on  5  degrees of freedom
## AIC: 93.658
##
## Number of Fisher Scoring iterations: 4
```

We get a residual deviance of 19.777 with df=5, therefore a p value of 0.001376036, still a significant loss. So the ghost writer didn't follow the style in SanditonI.

(ii)

For the first question, we build the Multinormal model like this:

$$\log \frac{\pi_{ij}}{\pi_{i1}} = \beta_{0j} + \beta_{1j} b_i^{SI} + \beta_{2j} b_i^{SS}$$

Where π_{ij} is the probability for j^{th} word in i^{th} data, $b_i^{()}$ are the flags for book name

Then we fit the model:

```
Jane_herself_2 = Jane %>%
                    filter(Book != "SanditonII") %>%
                    spread(Word,Count)

fit.jane21 <- vglm(formula = cbind(an,that,this,with,without,a) ~ factor(Book),
            family = multinomial, data = Jane_herself_2)

summary(fit.jane21)
```

```
##
## Call:
## vglm(formula = cbind(an, that, this, with, without, a) ~ factor(Book),
##     family = multinomial, data = Jane_herself_2)
##
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept):1                   -1.967650   0.209375  -9.398  < 2e-16 ***
## (Intercept):2                   -0.571786   0.122066  -4.684 2.81e-06 ***
## (Intercept):3                   -1.562185   0.176117  -8.870  < 2e-16 ***
## (Intercept):4                   -0.921682   0.137440  -6.706 2.00e-11 ***
## (Intercept):5                   -2.923162   0.324617  -9.005  < 2e-16 ***
## factor(Book)SanditonI:1         -0.249575   0.380326  -0.656   0.5117
## factor(Book)SanditonI:2         -0.432416   0.227658  -1.899   0.0575 .
## factor(Book)SanditonI:3         -0.344885   0.328002  -1.051   0.2930
## factor(Book)SanditonI:4         -0.361234   0.253979  -1.422   0.1549
## factor(Book)SanditonI:5          0.610626   0.463980   1.316   0.1882
## factor(Book)Sense&Sensibility:1  0.196093   0.301066   0.651   0.5148
## factor(Book)Sense&Sensibility:2  0.124649   0.179837   0.693   0.4882
## factor(Book)Sense&Sensibility:3  0.037488   0.262812   0.143   0.8866
```

```
## factor(Book)Sense&Sensibility:4  0.008786   0.206499   0.043   0.9661
## factor(Book)Sense&Sensibility:5  0.823101   0.409554   2.010   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  5
##
## Names of linear predictors:
## log(mu[,1]/mu[,6]), log(mu[,2]/mu[,6]), log(mu[,3]/mu[,6]), log(mu[,4]/mu[,6]), log(mu[,5]/mu[,6])
##
## Residual deviance: -6.628e-14 on 0 degrees of freedom
##
## Log-likelihood: -37.9598 on 0 degrees of freedom
##
## Number of iterations: 4
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):5'
##
## Reference group is level  6  of the response
```

We notice that the estimate $\hat{\beta_{1j}}$ and $\hat{\beta_{2j}}$ are mostly not significant away from zero(see the p value), which means for many words the ratio of odds between different book is not significant away from one. Therefore we can regard a consistence between Jane's writing.

For the second question, we build model like this:

$$\log \frac{\pi_{ij}}{\pi_{i1}} = \beta_{0j} + \beta_{1j} o_i$$

Where o_i is the flag for i^{th} whether or not belong to Jane.

Then we fit the model:

```
Jane_compare_2 = Jane %>%
                mutate(Original = (Book != "SanditonII")) %>%
                spread(Word,Count)
fit.jane22 <- vglm(formula = cbind(an,that,this,with,without,a) ~ Original,
            family = multinomial, data = Jane_compare_2)
summary(fit.jane22)
```

```
##
## Call:
## vglm(formula = cbind(an, that, this, with, without, a) ~ Original,
##     family = multinomial, data = Jane_compare_2)
##
##
## Pearson residuals:
##   log(mu[,1]/mu[,6]) log(mu[,2]/mu[,6]) log(mu[,3]/mu[,6])
## 1         -1.113e-01          3.593e-01          3.476e-01
## 2         -6.924e-01         -1.873e+00         -8.495e-01
## 3         -1.542e-16          9.135e-16          7.619e-16
## 4          6.287e-01          9.858e-01          2.470e-01
##   log(mu[,4]/mu[,6]) log(mu[,5]/mu[,6])
## 1          5.860e-01         -1.549e+00
## 2         -1.103e+00          6.159e-01
## 3         -1.995e-16         -3.643e-17
```

```
## 4          1.747e-01          1.226e+00
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -1.05154    0.21571  -4.875 1.09e-06 ***
## (Intercept):2  -1.32780    0.23980  -5.537 3.07e-08 ***
## (Intercept):3  -1.71079    0.28056  -6.098 1.08e-09 ***
## (Intercept):4  -0.65764    0.18789  -3.500 0.000465 ***
## (Intercept):5  -3.03255    0.51191  -5.924 3.14e-09 ***
## OriginalTRUE:1 -0.89437    0.25488  -3.509 0.000450 ***
## OriginalTRUE:2  0.71859    0.25307   2.839 0.004519 **
## OriginalTRUE:3  0.09209    0.30438   0.303 0.762225
## OriginalTRUE:4 -0.33400    0.20933  -1.596 0.110587
## OriginalTRUE:5  0.59709    0.53914   1.107 0.268082
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  5
##
## Names of linear predictors:
## log(mu[,1]/mu[,6]), log(mu[,2]/mu[,6]), log(mu[,3]/mu[,6]), log(mu[,4]/mu[,6]), log(mu[,5]/mu[,6])
##
## Residual deviance: 12.5873 on 10 degrees of freedom
##
## Log-likelihood: -55.6109 on 10 degrees of freedom
##
## Number of iterations: 4
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):5'
##
## Reference group is level  6  of the response
```

Here we can see some significant value for β_{1j}(see the p value). For this result we can say that compared to the ghost writer Jane used less `an` and more `that` (the estimated odds ratio for `an` and `that` between two writer based on `a` is exp(-0.89437)=0.4089 and exp(0.71859)=2.05154), and no significant difference for other words (based on the odds). Therefore we can say that the ghost writer didn't strictly follow Jane's pattern.

For the third question, the model is same as in second question, but now Jane's data only contains the book SanditonI We fit the model.

```
Jane_s12_2 = Jane %>%
               filter(Book=="SanditonI" | Book=="SanditonII") %>%
               spread(Word,Count)
fit.jane23 <- vglm(formula = cbind(an,that,this,with,without,a) ~ Book,
            family = multinomial, data = Jane_s12_2)
summary(fit.jane23)
```

```
##
## Call:
## vglm(formula = cbind(an, that, this, with, without, a) ~ Book,
##     family = multinomial, data = Jane_s12_2)
##
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept):1     -2.2172     0.3175  -6.983 2.88e-12 ***
## (Intercept):2     -1.0042     0.1922  -5.226 1.74e-07 ***
## (Intercept):3     -1.9071     0.2767  -6.892 5.50e-12 ***
## (Intercept):4     -1.2829     0.2136  -6.007 1.89e-09 ***
## (Intercept):5     -2.3125     0.3315  -6.976 3.04e-12 ***
## BookSanditonII:1   1.1657     0.3839   3.037  0.00239 **
## BookSanditonII:2  -0.3236     0.3073  -1.053  0.29232
## BookSanditonII:3   0.1963     0.3941   0.498  0.61842
## BookSanditonII:4   0.6253     0.2845   2.198  0.02794 *
## BookSanditonII:5  -0.7200     0.6099  -1.181  0.23777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  5
##
## Names of linear predictors:
## log(mu[,1]/mu[,6]), log(mu[,2]/mu[,6]), log(mu[,3]/mu[,6]), log(mu[,4]/mu[,6]), log(mu[,5]/mu[,6])
##
## Residual deviance: -3.297e-14 on 0 degrees of freedom
##
## Log-likelihood: -22.8086 on 0 degrees of freedom
##
## Number of iterations: 7
##
## No Hauck-Donner effect found in any of the estimates
##
## Reference group is level  6  of the response
```

The result is similiar to question 2. We can see some significant value for β_{1j}(see the p value). We can say that compared to the SanditonI, the ghost writer used more `an` and more `with` (the estimated odds ratio for `an` and `with` between two writer based on `a` is exp(1.1657)=3.208 and exp(0.6253)=1.8688), and no significant difference for other words (based on the odds). Therefore we can say this ghost writer's work is not good.