Supplementary material to Lecture Notes 10

Multinomial Regression

Summary of the last lecture

- Logistic regression models for multinomial response
- Ungrouped binary/Grouped binary

Key terms of this lecture

- Logistic regression (cont'd)
  - Other link function
  - Confounding and interaction
  - Different study designs

Reading

- McCullagh and Nelder (1989) Chapter 5
- Dobson and Bartnett (2008) Chapter 8

- For the nominal logistic regression, we have two R functions to do this: one is to use `multinom()` in the package "nnet". The other is to use `vglm()` in the package "VGAM" with `family=multinomial`. `multinom()` is preferred, since it uses the first category as the reference and its result is consistent with the model form. However, `vglm()` can do both the nominal and ordinal logistic regression by changing the argument in "family=". Note: `multinom()` uses long form data and `vglm()` uses wide form data. [See `Table8_2-VGAM.R`]

```
##  logistic regression for nominal response;
library(VGAM)
car.nl<-vglm(cbind(B_Imp, C_VeImp, A_NoImp) ~ c_age + c_sex,
       family=multinomial, data=car.wide)
```

- In fact, for the proportional odds logistic regression, we have two R functions to do this: one is to use polr() in the package "MASS". The other is to use vglm() in the package "VGAM". vglm() is preferred, since its result is consistent with the model form and the result from SAS, while polr() has a different model form and the signs of the coefficients are opposite from those from vglm(). Note: polr() uses long form data and vglm() uses wide form data. [See Table8_4-VGAM.R]

```
## Cumulative logit for ordinal response;
library(VGAM)
car.cl1 <-vglm(cbind(A_NoImp, B_Imp, C_VeImp) ~ c_age + c_sex,
          family=cumulative(parallel=TRUE), data=car.wide)
```


More Examples for Multi-category Logit Models: Political Ideology and Party Affiliation

In Agresti's 2007 book, Table 6.7, from a General Social Survey, relates political ideology to political party affiliation. Political ideology has a five-point ordinal scale, ranging from very liberal to very conservative.

		Ideology				
Gender	Party	VLib	SLib	Mod	SCon	VCon
Female	Dem	44	47	118	23	32
	Rep	18	28	86	39	48
Male	Dem	36	34	53	18	23
	Rep	12	18	62	45	51

Y = political ideology (very liberal, slightly liberal, moderate, slightly conservative, very conservative)

x_1 = gender(1=M,0=F); x_2 = political party (1 = Rep, 0 = Dem)

Cumulative Logit Model:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \quad j = 1, 2, 3, 4.$$

In R, we will fit these models using the VGAM package. Another R package `ordinal` also does the work. Install R package `icda` to get the data. See R code `ideology.R`. Note: VGAM treats the last category as the reference.

```
install.packages("icda",repos="http://www.stat.ufl.edu/~presnell/R",type="source")
> library(VGAM)
> library(icda)
> data(ideology)
> head(ideology)
  Party Gender Ideology Freq
1   Dem Female     VLib   44
2   Rep Female     VLib   18
3   Dem   Male     VLib   36
4   Rep   Male     VLib   12
5   Dem Female     SLib   47
6   Rep Female     SLib   28

> library(reshape2)
> ideow <- dcast(ideology, Gender + Party ~ Ideology,
```



```
                value.var="Freq")
> ideow
  Gender Party VLib SLib Mod SCon VCon
1 Female   Dem   44   47 118   23   32
2 Female   Rep   18   28  86   39   48
3   Male   Dem   36   34  53   18   23
4   Male   Rep   12   18  62   45   51
> ideo.cl1 <-
    vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender + Party,
        family=cumulative(parallel=TRUE), data=ideow)
> summary(ideo.cl1)

Call:
vglm(formula = cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender +
    Party, family = cumulative(parallel = TRUE), data = ideow)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  -1.4518     0.1228 -11.818  < 2e-16 ***
(Intercept):2  -0.4583     0.1058  -4.333 1.47e-05 ***
(Intercept):3   1.2550     0.1145  10.956  < 2e-16 ***
(Intercept):4   2.0890     0.1292  15.174  < 2e-16 ***
GenderMale     -0.1169     0.1268  -0.921    0.357
PartyRep       -0.9636     0.1294  -7.449 9.39e-14 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

```
Number of linear predictors:  4

Names of linear predictors:
logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3]), logit(P[Y<=4])

Residual deviance: 15.0556 on 10 degrees of freedom

Log-likelihood: -47.415 on 10 degrees of freedom

> deviance(ideo.cl1)
[1] 15.056
> df.residual(ideo.cl1)
[1] 10
> pchisq(deviance(ideo.cl1), df.residual(ideo.cl1),
        lower.tail = FALSE)
[1] 0.13005
```


Political Ideology and Party Affiliation (cont'd)

- Cumulative logit model fit: for $j = 1, 2, 3, 4$.

$$\text{logit}[\hat{P}(Y \le j)] = \boxed{\hat{\alpha}_j + \underbrace{(-0.117)x_1}_{\text{gender}} + (-0.964)}\underbrace{x_2}_{\text{party}}$$

$\hat{\alpha}_1 = -1.452$, $\hat{\alpha}_2 = -0.458$, $\hat{\alpha}_3 = 1.255$, $\hat{\alpha}_4 = 2.089$

- Controlling for gender, estimated odds that a Rep's response is in liberal direction ($Y \le j$) rather than conservative ($Y > j$) are $\boxed{\exp(-0.964)} = 0.38$ times estimated odds for a Dem. (0.38136)
- Equivalently: controlling for gender, estimated odds that a Dem's response is in liberal direction ($Y \le j$) rather than conservative ($Y > j$) are

$\exp\boxed{(0.964)} = 2.62$ times estimated odds for a Rep.

- 95% CI for true odds ratio is $\boxed{\exp(\hat{\beta} \pm SE(\hat{\beta}) \times 1.96}$ = (0.30, 0.49).
$= (\exp(-1.21684), \exp(-0.7116)) = (0.296, 0.491) \approx (0.30, 0.49)$
- Contingency table data. No evidence of lack of fit: deviance = 15.1, df = 10, p-value = 0.13

- Test for party effect (controlling for gender), i.e., H_0 : $\boxed{\beta_2 = 0}$

  Wald: $z = \boxed{\frac{\hat\beta_2 - 0}{SE(\hat\beta_2)}}$ $(z^2 = 55.49)$. $= \frac{-0.964}{0.129} = -7.4729$

  LR: $D_0 - D_1 = \boxed{71.902 - 15.056}$ $= 56.846$, $df = \boxed{11 - 10} = 1$

  p-value < 0.0001, (either test)

  Strong evidence that Republicans tend to be less liberal (more conservative than Democrats (for each gender).

- No evidence of gender effect (controlling for party). (p-value ≈ 0.36 using either Wald or LR test).

```
> ideo.cl2 <-
    vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender,
         family=cumulative(parallel=TRUE), data=ideow)
> deviance(ideo.cl2)
[1] 71.902
> df.residual(ideo.cl2)
[1] 11
```



```
> deviance(ideo.cl2) - deviance(ideo.cl1)
[1] 56.847
> pchisq(deviance(ideo.cl2) - deviance(ideo.cl1),
         df.residual(ideo.cl2) - df.residual(ideo.cl1),
         lower.tail=FALSE)
[1] 4.711e-14
```

- The ordinal logistic regression model can also be fitted by `polr()` in the R package "MASS". Note the opposite signs in coefficients, but the intercept are the same.

```
##Mehtod II: Using polr()
library(MASS)
##Use the long form data
## Variables:  Party Gender Ideology Freq
ideo.polr<-polr(factor(Ideology)~factor(Gender)
   +factor(Party), weights=Freq, data=ideology)
```

```
print(summary(ideo.polr))

Call:
polr(formula = factor(Ideology) ~ factor(Gender)
      + factor(Party), data = ideology, weights = Freq)

Coefficients:
                          Value Std. Error t value
factor(Gender)Male 0.1169      0.1273  0.9177
factor(Party)Rep   0.9636      0.1297  7.4311
```

Compare this to that from vglm(), the signs are opposite.

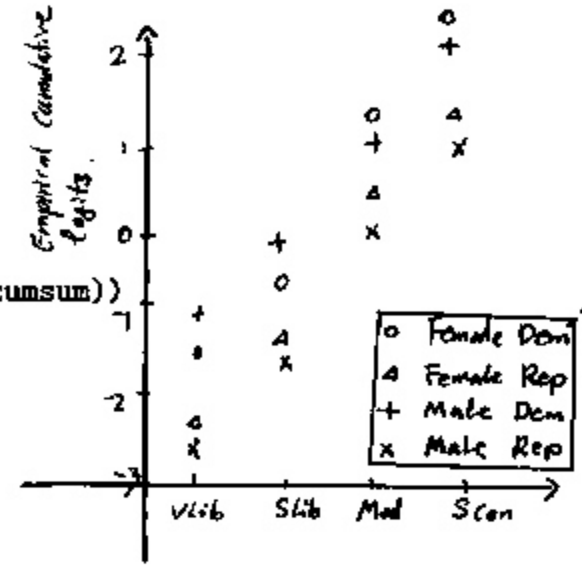
```
Intercepts:
           Value    Std. Error t value
VLib|SLib  -1.4518   0.1226    -11.8373
SLib|Mod   -0.4583   0.1048     -4.3746
Mod|SCon    1.2550   0.1142     10.9874
SCon|VCon   2.0890   0.1293     16.1588
```

Compare this to that from vglm(), they are the same!


```
Residual Deviance: 2474.142
AIC: 2486.142
```

Political Ideology and Party Affiliation (cont'd): Party-Gender interaction?

```
> ideow
  Gender Party VLib SLib Mod SCon VCon
1 Female   Dem   44   47 118   23   32
2 Female   Rep   18   28  86   39   48
3   Male   Dem   36   34  53   18   23
4   Male   Rep   12   18  62   45   51
> ideo.csum <- t(apply(ideow[,-(1:2)], 1, cumsum))
>  ideo.csum
     VLib SLib Mod SCon VCon
[1,]   44   91 209  232  264
[2,]   18   46 132  171  219
[3,]   36   70 123  141  164
[4,]   12   30  92  137  188
```




```
> ideo.cprop <- ideo.csum[,1:4]/ideo.csum[,5]
> ideo.ecl <- qlogis(ideo.cprop) # empirical cumul. logits
> ideo.cl3 <-
+     vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender*Party,
+          family=cumulative(parallel=TRUE), data=ideow)
> coef(summary(ideo.cl3))
                     Estimate Std. Error     z value     Pr(>|z|)
(Intercept):1      -1.5520853  0.1335312 -11.6233940 3.134323e-31
(Intercept):2      -0.5549908  0.1170310  -4.7422535 2.113539e-06
(Intercept):3       1.1646526  0.1233734   9.4400633 3.725543e-21
(Intercept):4       2.0012144  0.1368228  14.6263266 1.908149e-48
GenderMale          0.1430828  0.1793647   0.7977197 4.250331e-01
PartyRep           -0.7562072  0.1669103  -4.5306197 5.881093e-06
GenderMale:PartyRep -0.5091332 0.2540825  -2.0038111 4.509030e-02
> deviance(ideo.cl3)
[1] 11.06338
>  df.residual(ideo.cl3)
```

```
[1] 9
> deviance(ideo.cl1) - deviance(ideo.cl3)
[1] 3.992192
> pchisq(deviance(ideo.cl1) - deviance(ideo.cl3),
+         df.residual(ideo.cl1) - df.residual(ideo.cl3),
+         lower.tail=FALSE)
[1] 0.04571157
```


Political Ideology and Party Affiliation (with interaction)

Plot of empirical logits suggests interaction between party and gender. The model is

$$\text{logit}[\hat{P}(Y \le j)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \quad j = 1, 2, 3, 4.$$

- ML fit:

Gender Party Gender × Party

$$\text{logit}[\hat{P}(Y \le j)] = \boxed{\hat{\alpha}_j + 0.143\, x_1 + (-0.756)\, X_2 +} (-0.509)\, x_1 * x_2$$

$\hat{\alpha}_1 = -1.552$, $\hat{\alpha}_2 = -0.555$, $\hat{\alpha}_3 = 1.165$, $\hat{\alpha}_4 = 2.006$.

- Test for party × gender interaction ($H_0: \beta_3 = 0$):

  LR: $D_0 - D_1 = \boxed{15.056 - 11.063} = \boxed{3.993}$

  $df = \boxed{1\ (10-9)}$, p − value = $\boxed{0.0457}$.

- Some evidence (significant at 0.05 level) that effect of Party varies with Gender (and vice versa).

Political Ideology and Party Affiliation (with interaction)

- Estimated odds ratio for party effect (x_2) is

Rep = 1

$$e^{-0.756} = \quad 0.47 \quad when\ x_1 = 0\ (F)$$
$$e^{-0.756-0.509} = e^{-1.265} = \quad 0.28 \quad when\ x_1 = 1\ (M)$$

Implying: among males and females, Reps are more conservative than Dems.

- Estimated odds ratio for gender effect (x_1) is

M = 1

$$e^{0.143} = \quad 1.15 \quad when\ x_2 = 0\ (Dem)$$
$$e^{0.143-0.509} = e^{-0.336} = \quad 0.69 \quad when\ x_2 = 1\ (Rep)$$

OR > 1

- Among Dems, males tend to be more liberal than females. Among Reps, males tend to be more conservative than females.

OR < 1

Political Ideology and Party Affiliation (with interaction) (cont'd

- Probability of very liberal for male and female Republicans, that is $\hat{P}(Y = 1)$ when $x_1 = 1$ or 0 and $x_2 = 1$.
- Notice $\hat{P}(Y \leq j) = \frac{\exp(\hat{\alpha}_j + 0.143x_1 - 0.756x_2 - 0.509x_1x_2)}{1+\exp(\hat{\alpha}_j + 0.143x_1 - 0.756x_2 - 0.509x_1x_2)}$. For $j = 1$, $\hat{\alpha}_1 = -1.55$.
- Male Republicans ($x_1 = 1, x_2 = 1$):

$$\hat{P}(Y = 1) = \frac{\exp(-1.55 + 0.143 - 0.756 - 0.509)}{1 + \exp(-1.55 + 0.143 - 0.756 - 0.509)} = \frac{\exp(-2.67)}{1 + \exp(-2.67)} = 0.065.$$

- Female Republicans ($x_1 = 0, x_2 = 1$):

$$\hat{P}(Y = 1) = \frac{\exp(-1.55 - 0.756)}{1 + \exp(-1.55 - 0.756)} = \frac{\exp(-2.31)}{1 + \exp(-2.31)} = 0.090.$$

- Similarly, $\hat{P}(Y = 2) = \hat{P}(Y \leq 2) - \hat{P}(Y \leq 1)$, etc. Note: $\hat{P}(Y = 5) = \hat{P}(Y \leq 5) - \hat{P}(Y \leq 4) = 1 - \hat{P}(Y \leq 4)$.

Remarks

- Reversing order of response categories changes signs of "slope" estimates (cumulative odds ratio $\rightarrow$ 1/cumulative odds ratio). To show this, let $Z = J - Y + 1$, $k = J - j$, when $j = 1, \ldots, J-1$, $k = J-1, \ldots, 1$, then, for example,

$$\text{logit}[P(Z \leq k)] = -\text{logit}[P(Y \leq j)] = -\alpha_{J-k} - \beta_1 x_1 - \beta_2 x_2.$$

- For ordinal response, only two sensible orderings, either $1, 2, \ldots, J$ or $J, J-1, \ldots, 1$