Handout 11

STA 138

Multiple Logistic Regression

We now discuss multiple logistic regression. Consider the "Arthritis" data set. This has three predictors: treatment, gender, age. The response is Y: 0 (no improvement), 1 (improvement). Note that we have merged "some improvement" and "marked improvement" in one category. We have also decided to include the interactions: treatment*gender, treatment*age and gender*age. Thus the logistic regression has 6 predictor variables: call them $X_1, \ldots, X_6$. Denoting $\pi_i = P(Y_i = 1)$ and π'_i as its logit transform, the model is

$$\pi'_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}, i = 1, \ldots, n = 84, \quad (1)$$

where X_{i1} is 1 for treatment and 0 for placebo,

X_{i2} is 1 for male and 0 for female,

X_{i3} is age, $X_{i4} = X_{i1}X_{i2}, X_{i5} = X_{i1}X_{i3}, X_{i6} = X_{i2}X_{i3}$.

The likelihood function of for the multiple logistic regression is exactly the same as in the simple linear logistic regression, i.e.,

$$\mathbf{X}^T \boldsymbol{\pi} = \mathbf{X}^T \mathbf{Y},$$

where logit of π_i is $\pi'_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}$ (see the Appendix). There is no explicit solution and computer packages solve this vector equation via iterative methods.

In Handout 10, we had introduced the concepts of Model 1, Model 0 in order to understand the computer outputs. Here Model 1 is the same as given above, and Model 0 is $\pi'_i = \beta_0$ for all i. The saturated model is the one whereby all the π_i's are allowed to be arbitrary, ie, there are n parameters to be estimated under the saturated model. From now onward, we will not discuss Model 1 and Model 0 anymore, and $\hat{\pi}_i$ will denote the estimated π_i under the model given in (1).

The output from the R function glm is given below.

Call:

glm(formula = y ~trtmnt + gender + age + trtmnt * gender + trtmnt * age + gender * age, family = 'binomial')

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2672	-0.9176	0.0048	0.9008	2.2100

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.334339	1.969125	-2.201	0.0277
trtmnt	1.411300	2.781719	0.507	0.6119
gender	2.310981	2.916685	0.792	0.4282
age	0.074478	0.035318	2.109	0.0350
trtmnt:gender	0.676070	1.344351	0.503	0.6150
trtmnt:age	0.003978	0.051174	0.078	0.9380
gender:age	-0.079679	0.050214	-1.587	0.1126

—

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 116.449 on 83 degrees of freedom

Residual deviance: 88.363 on 77 degrees of freedom

AIC: 102.36

Number of Fisher Scoring iterations: 4

xxxxxxxxxxxxxxxxxxxxxxx

Clearly the results show that there are a number of terms that are not significant.

There are two important testing problems:

(a) Can a particular variable, say treatment*age, be dropped?

(b) Can two variables, say treatment*gender and treatment*age, be dropped?

Can a particular variable be dropped from the model?

We want to test if treatment*age can be dropped, i.e., $H_0 : \beta_5 = 0$ vs $H_1 : \beta_5 \neq 0$.

We may carry out the test in two different (and equivalent) ways: z-test or χ^2-test.

(a) z-test; The R output gives us the z-statistic $z^* = b_5/s(b_5) = 0.078$ and the p-value is 0.9380. Clearly, the variable X_5 can be dropped.

(b) χ^2-test (likelihood ratio test): The full model is $\pi_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6}$.

The reduced models is $\pi_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_6 X_{i6}$.

Let L_F and L_R be the likelihoods for the full and the reduced models respectively. The degrees of freedom associated with the likelihood for any model is $n-$(#of beta parameters estimated). So the df's for the full and the reduced models are $df(F) = n - 7 = 77$ and $df(R) = n - 6 = 78$. The test criterion is

$$\begin{aligned} G^2 &= -2[\log(L_R) - \log(L_F)] \text{ with} \\ df &= df(R) - df(F) = 1. \end{aligned}$$

This tell us that we need to run "glm" twice: once for the full model and once for the reduced model.

R does not give a value for $-2log(L_R)$ or $-2log(L_F)$. However, R gives "residual deviance" for the reduced as well as the full models, and they are $-2[\log(L_R) - \log(L_S)]$ and $-2[\log(L_F) - \log(L_S)]$, where L_S is the likelihood for the saturated model. Thus,

$$\begin{aligned} G^2 &= -2[\log(L_R) - \log(L_F)] \\ &= \text{(residual deviance for the reduced model)} \\ &\quad - \text{(residual deviance for the full model)} \\ \text{with } df &= df(R) - df(F) = 78 - 77 = 1. \end{aligned}$$

From the output above we have: residual deviance for the full model=88.363.

We ran the reduced model and a residual deviance for the reduced model to be equal to 88.369. Thus we have

$$G^2 = 88.369 - 88.363 = 0.006.$$

Clearly, $G^2 = 0.006$ with $df = 1$ and the p-value is larger than 0.5. Clearly we cannot reject H_0. Conclusion: we may drop variable X_5.

Why are the chi-square test and the z-test are similar? The answer is simple: check that $G^2 = z^{*2}$.

Remark: The G^2 value can be also be obtained from R as follows. Run the full model using glm and store it as object "full". Run the reduced model using glm and store as object "red". Then give the R command "anova(red,full, test="Chisq")". This will result in the following output:

Analysis of Deviance Table

Model 1: y ~trtmnt + gender + age + trtmnt * gender + gender * age

Model 2: y ~trtmnt + gender + age + trtmnt * gender + trtmnt * age + gender * age

Resid	Resid.Df	Resid.Dev	Df	Deviance	Pr(>Chi)
1	78	88.369			
2	77	88.363	1	0.00604	0.938

Can we drop more than a few variables at a time?

We want to test if we can drop variables X_4 (treatment*gender) and X_5 (treatment*age) from the model, ie, test $H_0 : \beta_4 = \beta_5 = 0$ against H_1 :at least one of β_4 and β_5 is not zero.

Here the only option is the chi-square test (likelihood ratio test). The method is the same as described before. The full and the reduced models are

full: $\pi_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6}$.

reduced $\pi_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_6 X_{i6}$.

Let L_F and L_R be the likelihoods for the full and the reduced models respectively. Their df's are $df(F) = n - 7 = 77$, $df(R) = n - 5 = 79$. The test statistic is

$$
\begin{aligned}
G^2 &= -2[\log(L_R) - \log(L_F)] \\
&= (\text{residual deviance for the reduced model}) \\
&\quad - (\text{residual deviance for the full model}) \\
\text{with } df &= df(R) - df(F) = 77 - 75 = 2.
\end{aligned}
$$

We ran the glm for the reduced model and the residual deviance is 88.640. Thus the value of the test statistic is

$$G^2 = 88.640 - 88.363 = 0.27686.$$

The p-value using the χ^2_2 distribution is 0.871. This tell us we cannot reject H_0. Thus we may drop variables X_4 and X_5.

Remark. If we run the full and the reduced models, store them as objects "full" and "red", the R command yields the output

Analysis of Deviance Table

Model 1: y ~trtmnt + gender + age + gender * age

Model 2: y ~trtmnt + gender + age + trtmnt * gender + trtmnt * age +
gender * age

Resid	Resid.Df	Resid.Dev	Df	Deviance	Pr(>Chi)
1	79	88.640			
2	77	88.363	2	0.27686	0.8707

Stepwise logistic regression:

As in the linear regression case, we can run a stepwise procedure. We will describe the backward stepwise regression procedure using AIC. However, one can change the model selection criterion to BIC and also change the method from the backward stepwise to forward stepwise.[R command: "step(full)", where "full" is the glm object when all the 6 predictor variables are in the model].

Start: AIC=102.36

y ~trtmnt + gender + age + trtmnt * gender + trtmnt * age +
gender * age

	Df	Deviance	AIC
trtmnt:age	1	88.369	100.37
trtmnt:gender	1	88.631	100.63
<none>		88.363	102.36
gender:age	1	90.963	102.96

Step: AIC=100.37

y ~trtmnt + gender + age + trtmnt:gender + gender:age

	Df	Deviance	AIC
trtmnt:gender	1	88.640	98.64
<none>	1	88.369	100.37
gender:age	1	91.363	101.64

Step: AIC=98.64

y ~trtmnt + gender + age + gender:age

	Df	Deviance	AIC
<none>	1	88.640	98.64
gender:age	1	92.063	100.06
trtmnt	1	100.540	108.54

Call: glm(formula = y ~trtmnt + gender + age + gender:age, family = binomial)

Coefficients:

(Intercept)	trtmnt	gender	age	gender:age
-4.55477	1.79705	2.75066	0.07734	-0.07945

xxxxxxxxxxxxxxxxxxxxxxxx

The output clearly points out that the backward stepwise regression with the AIC criterion has chosen a model with the variables: X_1 (treatment), X_2(gender), X_3 (age) and $X_4 = X_1X_2$.

Confidence interval for the probability.

We have now a model which keeps variables treatment, gender, age and the interaction gender and age. For this model, we wish to estimate the probability π of improvement for a female of age 55 who has received the treatment, and also obtain a 95% confidence interval for this probability. In order to obtain a confidence interval we go through the following steps:

(i) Obtain an estimate $\hat{\pi}'$ of π'.

(ii) Obtain the standard error $s(\hat{\pi}')$ of $\hat{\pi}'$.

(iii) Construct a 95% confidence interval for π': suppose this interval is $\{L, U)$.

(iv) 95% confidence interval for π then is $(\exp(L)/(1+\exp(L)),\ \exp(U)/(1+\exp(U)))$.

From R we have

$$\hat{\pi}' = 1.4961,\ s(\hat{\pi}') = 0.4801.$$

So an approximate 95% confidence interval for π' is

$$\begin{aligned}&\hat{\pi}' \pm 1.96 s(\hat{\pi}'),\ \text{ie, } 1.4961 \pm (1.96)(0.4801), \text{ie,}\\ &1.4961 \pm 0.9410, \text{ ie, } (0.5551, 2.4371).\end{aligned}$$

So an approximate 95% confidence interval for π is

$$\begin{aligned}&(\exp(0.5551)/(1+\exp(0.5551)),\ \exp(2.4371)/(1+\exp(2.4371)))\\ =\ &(0.6353, 0.9196).\end{aligned}$$

Here are the R commands:

```
> arthritis=glm(improve~trtmnt+gender+age+gender*age,family='binomial')
> new=data.frame(trtmnt="Treated",Gender="Female",age=55)
> predict(arthritis,newdata=new,se.fit=TRUE,type='link')
$fit
1
1.496094
$se.fit
[1] 0.4801197
$residual.scale
[1] 1.
```

Multicategory (or multinomial) logistic regression.

Recall that in the "Arthritis" data, the response Y has three categories: no improvement, some improvement and marked improvement. Multicategory (or polytomous) logistic regression seeks to deal with this kind of data. Note that the logistic models we have dealt with so far have the response Y 0-1 valued, ie, $Y_i \sim$Bernoulli(π_i), i.e., $Y_i \sim$Binomial(1,π_i). For the Arthritis data, Y has $J = 3$ categories which we will write as none (some), 2 (some) and 3 (marked). Let $\pi_{ij} = P(Y_i = j), j = 1, \ldots, J = 3$, where $\pi_{i1} + \pi_{i2} + \pi_{i3} = 1$. Then $Y_i \sim$Multinomial$(1; \pi_{i1}, \pi_{i2}, \pi_{i3})$. Thus the modeling involves

$$\begin{aligned}\log(\pi_{i1}/\pi_{i3}) &= \beta_{01} + \beta_{11} X_{i1} + \beta_{21} X_{i2} + \beta_{31} X_{i3} = \boldsymbol{\beta}_1^T \mathbf{X}_i,\\ \log(\pi_{i2}/\pi_{i3}) &= \beta_{02} + \beta_{12} X_{i1} + \beta_{22} X_{i2} + \beta_{32} X_{i3} = \boldsymbol{\beta}_2^T \mathbf{X}_i,\end{aligned}$$

where $\mathbf{X}_i^T = (1, X_{i1}, X_{i2}, X_{i3})$, $\boldsymbol{\beta}_1^T = (\beta_{01}, \beta_{11}, \beta_{21}, \beta_{31})$ and $\boldsymbol{\beta}_2^T = (\beta_{02}, \beta_{12}, \beta_{22}, \beta_{32})$. Thus there are two sets of beta parameters, i.e., a total of 8 beta parameters. In general, when Y has J categories, we may write

$$\begin{aligned}
\log(\pi_{i1}/\pi_{iJ}) &= \boldsymbol{\beta}_1^T \boldsymbol{X}_i, \\
\log(\pi_{i2}/\pi_{iJ}) &= \boldsymbol{\beta}_2^T \boldsymbol{X}_i, \\
&\vdots \\
\log(\pi_{i,J-1}/\pi_{iJ}) &= \boldsymbol{\beta}_{J-1}^T \boldsymbol{X}_i, \text{ or} \\
\log(\pi_{ij}/\pi_{iJ}) &= \boldsymbol{\beta}_j^T \boldsymbol{X}_i, j = 1, \ldots, J-1.
\end{aligned}$$

Note that there are $J-1$ sets of betas leading to a total of $4(J-1)$ beta parameters.

Note that we have taken the ratios of π_{i1} and π_{i2} with respect π_{i3}. Thus we have made the third category "marked" as the base category. It is clearly possible to make the first response (or the second response) category as the base category.

Also note that

$$\begin{aligned}
\pi_{i1}/\pi_{i3} = \exp(\boldsymbol{\beta}_1 \boldsymbol{X}_i),\ \pi_{i2}/\pi_{i3} &= \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_i), \\
\exp(\boldsymbol{\beta}_1 \boldsymbol{X}_i) + \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_i) &= \pi_{i1}/\pi_{i3} + \pi_{i2}/\pi_{i3} \\
&= \frac{\pi_{i1} + \pi_{i2}}{\pi_{i3}} = \frac{1 - \pi_{i3}}{\pi_{i3}} = 1/\pi_{i3} - 1.
\end{aligned}$$

Solving for π_{i3} we thus have

$$\pi_{i3} = \frac{1}{1 + \exp(\boldsymbol{\beta}_1 \boldsymbol{X}_i) + \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_i)}.$$

Thus we may write

$$\begin{aligned}
\pi_{i1}/\pi_{i3} = \exp(\beta_1^T X_i)\,, \text{ or } \pi_{i1} &= \frac{\exp(\beta_1^T X_i)}{1 + \exp(\beta_1 X_i) + \exp(\beta_2^T X_i)}, \text{ and similarly} \\
\pi_{i2} &= \frac{\exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_i)}{1 + \exp(\boldsymbol{\beta}_1 \boldsymbol{X}_i) + \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_i)}, \text{ or} \\
\pi_{ij} &= \frac{\exp(\boldsymbol{\beta}_j^T \boldsymbol{X}_i)}{1 + \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_i) + \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_i)}, j = 1, 2.
\end{aligned}$$

In the case there are J response categories and if we make the last (i.e., the J^{th}) response category as the base category, then we may write the model as

$$\pi_{ij} = \frac{\exp(\boldsymbol{\beta}_j^T \boldsymbol{X}_i)}{1 + \exp(\boldsymbol{\beta}_1 \boldsymbol{X}_i) + \cdots + \exp(\boldsymbol{\beta}_{J-1}^T \boldsymbol{X}_i)}, j = 1, \ldots, J-1.$$

Analysis of Arthritis data.

R package "nnet" is required for the analysis.

The first is to create a base category for Y. We created the base as "Marked" and the R command is "relevel(y,ref = "Marked")".

Then the following R command runs a multinomial logistic regression, also given is the output from the (backward with AIC) stepwise procedure. They are

(a) ff=multinom(y~trtmnt+gender+age+trtmnt*gender+trtmnt*age+gender*age)

(b) step(ff)

xxxxxxxxxxxxxxxxxxxxxx

Call:

multinom(formula = y ~trtmnt + gender + age + trtmnt * gender +
trtmnt * age + gender * age)

Coefficients:

	(Intercept)	trtmnt	gender	age	trtmnt:gender	trtmnt:age	gender:age
None	5.2576	-2.4135	-2.2523	-0.0769877	0.211871	0.0042747	0.0639831
Some	0.54730	-4.5142	-4.3556	-0.0067962	9.2514	0.0539279	-0.0838856

Std. Errors:

(Intercept)	trtmnt	trtmnt	gender	age	trtmnt:gender	trtmnt:age	gender:age
None	2.6135	3.2459	3.1641	0.0456481	1.4111	0.0584551	0.0538420
Some	3.2052	5.0468	2.3783	0.0546242	2.3845	0.0839400	0.0831563

Residual Deviance: 136.5842

AIC: 164.5842

xxxxxxxxxxxxxxxxxxxxxx

This output does not directly calculate the z-values or p-values. However, one can get these quantities using the table. One can also get fitted values etc.

xxxxxxxxxxxxxxxxxxxxxx

The final output of the stepwise procedure is

Coefficients:

	(Intercept)	trtmnt	gender	age	gender:age
None	4.9138	-2.1611	-1.8834	-0.0714897	0.0616876
Some	-0.85535	-1.0895	2.9181	0.0153738	-0.0611876

Residual Deviance: 138.6234

AIC: 158.6234

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Note that the stepwise procedure does give you the standard errors. However, you can run the function "multinom" with predictors trtmnt, gender, age, gender*age onec again to get the standard errors.
which are solved iteratively.

Multicategory (or multinomial) logistic regression with ordinal response.

Note that for the arthritis data, the response categories "none", "some" and "marked" have some ordering. Such a response is called an "ordinal response". There is a popular modeling option for this and it is called the "proportional odds model", Let us first write down what the model is. Let $\boldsymbol{X}_i^T = (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6})$, where X_{i1} 0-1 valued (treatment), X_{i2} is 0-1 valued (gender), X_{i3} is age, X_{i4} is treatment*gender, X_{i5} is treatment*age and X_{i6} is gender*age. Also number the categories "none", "some" and "marked" by 1, 2, 3. So $P(Y_i = 1)$ is the probability of reporting no improvement in arthritis , $P(Y \leq 2)$ is the probability of reporting at most some improvement. In general we may have J ordinal categories. Then the proportional odds model is, for any $i = 1, \dots, n$,

$$\frac{P(Y_i \le j)}{1 - P(Y_i \le j)} = \exp(\alpha_j + \boldsymbol{\beta}^T \mathbf{X}_i), \; j = 1, \dots, J-1, \text{i.e.,}$$

$$\log\left(\frac{P(Y_i \le j)}{1 - P(Y_i \le j)}\right) = \alpha_j + \boldsymbol{\beta}^T \mathbf{X}_i, \text{ ie, logit of } P(Y \le j) \text{ is } \alpha_j + \boldsymbol{\beta}^T \mathbf{X}_i.$$

Note that the for any given i, the odds $\frac{P(Y_i \le j)}{1-P(Y_i \le j)}$, $j = 1, \dots, J-1$, are proportional to each other and that is why it is called the proportional odds model. Note that here we have only one β vector and $J-1$ scalar parameters $\alpha_1, \dots, \alpha_{J-1}$. Thus the proportional odds model may require a lot fewer parameters than the full multinomial regression model (as described before). The multinomial model requires $2 \times 7 = 14$ parameters. Whereas, the ordinal logistic model here requires only 8 parameters. However, this model can be justified only if the response is ordinal.

[Mass package in R. Command: polr(y~trtmnt+gender+age+trtmnt*gender+trtmnt*age+gender*age)
Re-fitting to get Hessian
Call:
polr(formula = y ~trtmnt + gender + age + trtmnt * gender +
trtmnt * age + gender * age)

Coefficients:	Value	Std. Error	t value
trtment	2.56715	2.48168	1.0344
gender	0.49429	2.59563	0.1904
age	0.06371	0.03247	1.9622
trtmnt*gender	0.77082	1.28892	0.5980
trtmnt*age	−0.01770	0.04363	−0.4056
gender*age	−0.04281	0.04317	−0.9917
Intercepts:			
1|2	3.8630	1.8452	2.0936
2|3	4.7798	1.8719	2.5534

Residual Deviance: 143.1031
AIC: 159.1031
xxxxxxxx

Note that in the output, what are referred to as "t-values" are $\hat{\beta}_1/s(\hat{\beta}_1)$, $\hat{\beta}_2/s(\hat{\beta}_2)$ etc. One should use the normal table and not the t-table in order to find the p-values. Note that intercept terms seem to be significant.

One can carry out stepwise regression usning the 'step' command. In this case, it deleted all the interaction terms leading to the following final part of the output:

Coefficients:

trtmnet	gender	age
1.74528949	−1.25167969	0.03816199

Intercepts:

1|2	2|3
2.531932	3.430942

Residual Deviance: 145.4579

AIC: 155.4579

xxxxxx

Heuristic justification of logistic model for ordinal response.

Consider the "Arthritis" data. The response Y now takes three values 1 (none), 2 (some), 3 (marked). Let as assume that there is a latent (unobserved) continuous variable Y_i^c which guides the sense of improvement in pain. Patient reports: "none" if $Y_i^c \le T_1$, "some" when $T_1 < Y_i^c \le T_2$, and "marked" if $Y_i^c > T_2$. Thus the event $\{Y_i = 1\}$ is the same as $\{Y_i^c \le T_1\}$, the event $\{Y_i = 2\}$ is the same as $\{T_1 < Y_i^c \le T_2\}$ and the event $\{Y_i = 2\}$ is the same as $\{Y_i^c > T_2\}$.

Also assume that Y_i^c has linear relation with the independent variables, i.e.,

$$Y_i = \beta_0^* + \boldsymbol{\beta}^{*T}\mathbf{X}_i + \varepsilon_i,$$

where $\varepsilon_i = \sigma Z_i$ and Z_i's are iid standard logistic random variables. Thus

$$\begin{aligned}
P(Y_i \le 1) &= P(Y_i = 1) = P(Y_i^c \le T_1) \\
&= P(\beta_0^* + \boldsymbol{\beta}^{*T}\mathbf{X}_i + \varepsilon_i \le T_1) \\
&= P\left(Z_i \le \frac{T_1 - \beta_0^* - \boldsymbol{\beta}^{*T}\mathbf{X}_i}{\sigma}\right) \\
&= P(Z_i \le \alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i), \text{ setting } \alpha_1 = (T_1 - \beta_0^*)/\sigma,\ \boldsymbol{\beta} = -\boldsymbol{\beta}^*/\sigma, \\
&= \frac{\exp(\alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i)}{1 + \exp(\alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i)}, \text{ or} \\
\log\left(\frac{P(Y_i \le 1)}{1 - P(Y_i \le 1)}\right) &= \alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i.
\end{aligned}$$

Note that we do not carry out the calculations for $P(Y \le 3)$ since $P(Y_i \le 3) = 1$. Note that the above arguments specify the probabilities of $Y_i = 1$, $Y_i = 2$ and $Y_i = 3$, since

$$\begin{aligned}
\pi_{i1} &= P(Y_i = 1) = \frac{\exp(\alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i)}{1 + \exp(\alpha_1 + \beta^T X_i)}, \\
\pi_{i2} &= P(Y_i = 2) = P(Y_i \le 2) - P(Y_i \le 1) \\
&= \frac{\exp(\alpha_2 + \boldsymbol{\beta}^T\mathbf{X}_i)}{1 + \exp(\alpha_2 + \boldsymbol{\beta}^T\mathbf{X}_i)} - \frac{\exp(\alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i)}{1 + \exp(\alpha_1 + \boldsymbol{\beta}^T\mathbf{X}_i)}, \\
\pi_{i3} &= P(Y_i = 3) = P(Y_i \le 3) - P(Y_i \le 2) \\
&= \frac{\exp(\alpha_3 + \boldsymbol{\beta}^T\mathbf{X}_i)}{1 + \exp(\alpha_3 + \boldsymbol{\beta}^T\mathbf{X}_i)} - \frac{\exp(\alpha_2 + \boldsymbol{\beta}^T\mathbf{X}_i)}{1 + \exp(\alpha_2 + \boldsymbol{\beta}^T\mathbf{X}_i)}.
\end{aligned}$$

In general if Y has J ordered categories $1, \ldots, J$, we may follows the same arguments to get

$$\begin{aligned}
\frac{P(Y_i \le j)}{1 - P(Y_i \le j)} &= \exp(\alpha_j + \boldsymbol{\beta}^T\mathbf{X}_i),\ j = 1, \ldots, J-1, \text{i.e.,} \\
\log\left(\frac{P(Y_i \le j)}{1 - P(Y_i \le j)}\right) &= \alpha_j + \boldsymbol{\beta}^T\mathbf{X}_i, \text{ ie, logit of } P(Y \le j) \text{ is } \alpha_j + \boldsymbol{\beta}^T\mathbf{X}_i.
\end{aligned}$$

Appendix.

Likelihood function for multiple logistic regression.

Consider the arthritis data set where we have 6 independent variables of which three are interaction terms. Let $X_{i1}, i = 1, \ldots, n$, be the values of variable X_1, $X_{i2}, i = 1, \ldots, n$, be the vales of variable X_2 etc. The response is Y_i, which is 0-1 valued. If we denote $\pi_i = P(Y_i = 1)$, then the likelihood is

$$\begin{aligned} L &= \prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}, \text{ and} \\ \log L &= \sum_{i=1}^{n} [Y_i \log(\pi_i) + (1-Y_i)\log(1-\pi_i)]. \end{aligned}$$

When we fit $\pi_i' = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}$, where π_i' is the logit of π_i, then we have

$$\begin{aligned} \log L &= \sum_{i=1}^{n} [Y_i \log(\pi_i) + (1-Y_i)\log(1-\pi_i)] \\ &= \sum_{i=1}^{n} [Y_i \log(\pi_i/(1-\pi_i)) + \log(1-\pi_i)] \\ &= \sum_{i=1}^{n} [Y_i(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}) - \log(1 + \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6})]. \end{aligned}$$

In order to maximize $\log L$ with respect to $\beta_0, \beta_1, \ldots, \beta_6$ we differentiate $\log L$ with respect to $\beta_0, \beta_1, \ldots, \beta_6$ and equate the derivative to zero leading to the so-called "likelihood equations"

$$\begin{aligned} \sum_i \pi_i &= \sum_i Y_i, \\ \sum_i X_{i1}\pi_i &= \sum_i X_{i1} Y_i, \\ &\vdots \\ \sum_i X_{i6}\pi_i &= \sum_i X_{i6} Y_i, \end{aligned}$$

where $\pi_i = \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}) \,/\, [1 + \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6})]$. No explici expressions exist for the solutions. Iterative methods are needed to solve these equations and the solutions denoted by $\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_6$ are the maximum likelihood estimates of $\beta_0, \beta_1, \ldots, \beta_6$.

Thev likelihood equations can be compactly expressed. Let $\boldsymbol{X}$ be the $n \times 7$ matrix whose first column consits of 1's, the second column consists of the values of variable X_1, the third column consists of the values of variable X_2 etc. If $\boldsymbol{Y}$ is the $n \times 1$ vecor of Y-values and $\boldsymbol{\pi}$ is the vector $\pi_1, \ldots, \pi_n$ (noting that each π_i involves the betas), then the likelihood equations can be written in a matrix form

$$\boldsymbol{X}^T \boldsymbol{\pi} = \boldsymbol{X}^T \boldsymbol{Y},$$

where "T" denotes the matrix transpose.

Likelihood function for multicategory logistic case.

For the Arthritis data, each person reports one of the $J = 3$ responses. If we call these responses 1, 2, 3, then we may write the response of the i^{th} person as a vector (Y_{i1}, Y_{i2}, Y_{i3}). , where Y_{i1}, Y_{i2} and Y_{i3} are either 0 or 1. Also note that for each person only one of Y_{i1}, Y_{i2}, Y_{i3} is 1 and the rest are zeros, i.e. $Y_{i1}+Y_{i2}+Y_{i3} = 1$. If $P(Y_{i1} = 1) = \pi_{i1}$, $P(Y_{i2} = 1) = \pi_{i2}$ and $P(Y_{i3} = 1) = \pi_{i3}$, then we have $\pi_{i1} + \pi_{i2} + \pi_{i3} = 1$. We can now say that $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$ is distributed as multinomial$(1; \pi_{i1}, \pi_{i2}, \pi_{i3}0$. Thus the pdf of (Y_{i1}, Y_{i2}, Y_{i3}) is

$$\pi_{i1}^{Y_{i1}} \pi_{i2}^{Y_{i2}} \pi_{i3}^{Y_{i3}}.$$

So the joint probability of (Y_{i1}, Y_{i2}, Y_{i3}), $i = 1, \ldots, n$ is

$$\prod_{i=1}^{n} \left[\pi_{i1}^{Y_{i1}} \pi_{i2}^{Y_{i2}} \pi_{i3}^{Y_{i3}} \right].$$

If we assume the model specification earlier assuming category 3 as the base category, then the likelihood function turns out to be

$$L = \prod_{i=1}^{n} \left[\frac{\exp(Y_{i1}\beta_1^T X_i + Y_{i2}\beta_2^T X_i)}{1 + \exp(\beta_1^T X_i) + \exp(\beta_2^T X_i)} \right] = \frac{\exp\left(\sum Y_{i1}\beta_1^T X_i + \sum Y_{i2}\beta_2^T X_i \right)}{\prod_{i=1}^{n} \left[1 + \exp(\beta_1^T X_i) + \exp(\beta_2^T X_i) \right]}$$

The maximum likelihood estimators of β_1 and β_2 are obtained by maximizing the likelihood or, equivalently, by maximizing the log of the likelihood. Differentiating the logarithm of the likelihood with respect to β_1 and β_2 and equating the derivatives to zero lead to the following equations

$$\sum \frac{\exp(\beta_1^T X_i)}{1 + \exp(\beta_1^T X_i) + \exp(\beta_2^T X_i)} X_i = \sum Y_{i1} X_i,$$
$$\sum \frac{\exp(\beta_2^T X_i)}{1 + \exp(\beta_1^T X_i) + \exp(\beta_2^T X_i)} X_i = \sum Y_{i2} X_i.$$

Since $\pi_{i1} = \frac{\exp(\beta_1^T X_i)}{1+\exp(\beta_1^T X_i)+\exp(\beta_2^T X_i)}$ and $\pi_{i2} = \frac{\exp(\beta_2^T X_i)}{1+\exp(\beta_1^T X_i)+\exp(\beta_2^T X_i)}$, we may rewrite those equations as

$$\sum \pi_{i1} X_i = \sum Y_{i1} X_i, \;\; \sum \pi_{i2} X_i = \sum Y_{i2} X_i,$$

keeping in mind that π_{i1}'s and π_{i2}'s are functions of β_1 and β_2. These are solved by iterative methods as no explicit solutions are available.

If instead of 3 categories as we have discussed above, suppose that there are J categories of response, then we can write $(Y_{i1}, \ldots, Y_{iJ})$ is Multinomial$(1,(\pi_{i1}, \ldots, \pi_{iJ}))$. And the joint probability of $(Y_{i1}, \ldots, Y_{iJ})\,, i = 1, \ldots, J$, is

$$\prod_{i=1}^{n} \pi_{i1}^{Y_{i1}} \cdots \pi_{iJ}^{Y_{iJ}}.$$

When there are J categories of response, we have $J - 1$ beta vectors $\beta_1, \ldots, \beta_{J-1}$. The likelihood equations involved in estimating $\beta_1, \ldots, \beta_{J-1}$ are

$$\sum \pi_{i1} X_i = \sum Y_{i1} X_i, \ldots, \;\; \sum \pi_{i,J-1} X_i = \sum Y_{i,J-1} X_i$$

which are solved iteratively.

Estimation of parameter in ordinal logistic regression.

The method of estimation is via maximum likelihood. The likelihood function is

$$L = \prod_{i=1}^{n} \prod_{j=1}^{J} \pi_{ij}^{Y_{ij}},$$

where $\pi_{i1}, \ldots, \pi_{iJ}$ are functions of the parameters $\alpha_1, \ldots, \alpha_{J-1}$ and β. As usual the packages maximize $\log L$ using iterative methods and no closed-form of the solutions are available.