Multinomial GLMs for Multinomial Response with An Example of Brain Injury Recovery Stages

Frances Lin

Fall 2020

Abstract

We expand beyond binary (or binomial) response to focus on polychotomous (or multinomial) response. We first differentiate ordinal response from nominal response. Then, we briefly review multinomial distribution and latent variable. Next, we define the models and discuss model assumptions and estimation. Finally, we include an example of traumatic brain injury outcomes to illustrate how the proportional-odds cumulative logit model and the baseline-category logit model are used for estimation and prediction in practice.

Introduction

Recall from class that when response is binomial (i.e. response has = 2 categories), we fit models such as binomial regression with `logistic`, `probit`, and `cloglog` link. When response is multinomial (i.e. response has > 2 categories), binomial regression can be extended to multinomial regression.

However, of multinomial response, it is important to distinguish between nominal response (i.e. response that does not follow specific order) and ordinal response (i.e. response that has specific order). Some models can only be applied to ordinal response, whereas some models can be applied to both nominal and ordinal response. We discuss it in later section.

Multinomial Distribution and Latent Variable

Multinomial Distribution

Recall again that we fit binomial regression, assuming that the response $Y_i \overset{iid}{\sim} Binomial(n, \pi)$, where π is the probability of y_i falls into the first category and $1 - \pi$ is the probability of y_i falls into the other category.

Here, we fit multinomial regression, assuming that the response $Y_{ij} \overset{iid}{\sim} Multinomial(n, \pi_{ij})$, where π_{i1} is the probability of y_{ij} falls into the first ($j = 1$) category, π_{i2} is the probability of y_{ij} falls into the second ($j = 2$) category, etc.

Latent Variable

We can think instead that the response Y is a categorized version of an latent continuous variable Z, where

$$Y_{ij} = j \ if \ \alpha_{j-1} < Z_i < \alpha_j \ \ j = 1, 2, ..., c$$

$$\alpha_0 = -\infty, \ \alpha_c = \infty, \ \alpha_0 < \alpha_1 < \alpha_2 < ...\alpha_{c-1}$$

and

$$Z = X\beta$$

For example,

$Y = 1$, if $Z < \alpha_1$,

$Y = 2$, if $\alpha_1 < Z < \alpha_2$, etc.

Models, Model Assumptions, and Model Estimation

Recall that the logit model is as defines

$$logit(P(Y=1)) = \log(\frac{P(Y=1)}{1-P(Y=1)}) = \log(\frac{\pi}{1-\pi}) = X\beta$$

Ordinal Regression for Ordinal Response

The cumulative logit model is a commonly used ordinal regression specifically for ordinal response. It is as defines

$$logit(P(Y \le j)) = \log(\frac{P(Y \le j)}{1-P(Y \le j)}) = \log(\frac{\pi_j}{1-\pi_j}) = X\beta, j = 1, 2, 3, 4$$

If $j = 1$, for example,

$$\log(\frac{P(Y \le 1)}{1-P(Y \le 1)}) = \log(\frac{\pi_1}{1-\pi_1}) = \log(\frac{\pi_1}{\pi_2+\pi_3+\pi_4+\pi_5}) = X\beta, j = 1$$

Multinomial Regression for Nominal and Ordinal Response

The baseline-category logit model is a commonly used multinomial regression for nomial response, although it can also be applied to ordinal response. It is as defines

$$\log(\frac{\pi_j}{\pi_j^*}) = X\beta, j^* = reference\ category$$

If we specify $j = 1$ as the reference category, for example,

$$\log(\frac{\pi_j}{\pi_1}) = X\beta, j = 2, 3, 4, 5$$

Models such as the cumulative logit model is suitable only for ordinal response, whereas models such as the baseline-category logit model can be applied to both nominal and ordinal response. However, taking ordinality into account often result in simpler, optimal models.

An Example: The Glasgow Outcome Scale (GOS)

We fit the proportional-odds cumulative logit model on the `trama` dataset first and the same model on the same dataset with collapsing categories of response second, and for illustration purposes, we also fit the baseline-category logit model on the same dataset.

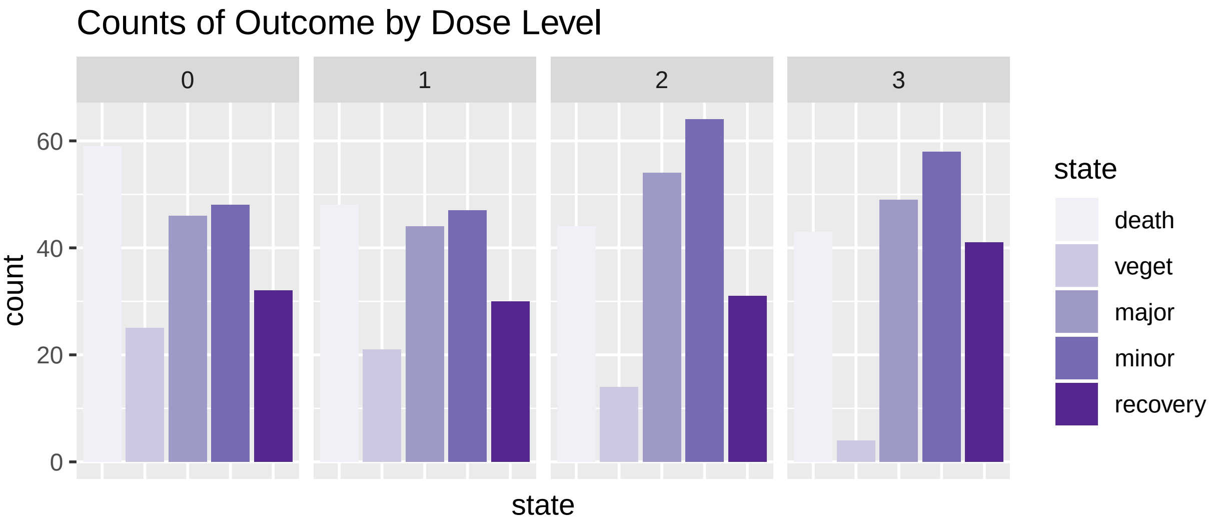
Data

Data were obtained from the 1977 paper by Chuang-Stein and Agresti. The Glasgow Outcome Scale (GOS) is a scale that is commonly used for traumatic brain injury assessment.

GOS is the response, and outcome is ordinal with ordered categories: *death, vegetative state, major disability, minor disability,* and *good recovery.* Intravenous dose is the covariate, and dose is categorical with four levels: *placebo, low dose, medium dose*, and *high dose.*

dose	death	veget	major	minor	recovery
placebo	59	25	46	48	32
low	48	21	44	47	30
med	44	14	54	64	31
high	43	4	49	58	41

Initial analysis shows that the *placebo* group (title = 0) has a higher relative proportions of patients whose outcomes fall into the *death* category, and the *high dose* group (title = 3) has a higher relative proportions of patients whose outcomes fall into the *good recovery* category.



The Cumulative Logit Model for Ordinal Response

First, we use the `plor` (proportional odds logistic regression) function from the `MASS` package to fit the cumulative logit model, as it is one of the most commonly used models for ordinal response.

Since we treat dose as categorical, dose is coded as *placebo* (dose = 0), *low dose* (dose = 1), *medium dose* (dose = 2), and *high dose* (dose = 3). The summary of the cumulative logit model is as follows:

	Value	Std. Error	t value	p value
factor(dose)1	0.1176	0.1791	0.6564	0.5116
factor(dose)2	0.3174	0.174	1.824	0.06816
factor(dose)3	0.5208	0.1794	2.903	0.003697
death|veget	-0.9188	0.1322	-6.949	3.685e-12
veget|major	-0.5183	0.1291	-4.015	5.936e-05
major|minor	0.4922	0.1298	3.792	0.0001491
minor|recovery	1.858	0.1462	12.71	5.391e-37

The Cumulative Link Model with Different Link Function

Fitting the cumulative link model with different link function (e.g. `probit`, `cloglog`), we see that the model with the `logit` link performs the best, but the difference is minimal.

	logit	probit	cloglog
AIC	2475	2477	2478

Model Interpretation

Exponentiating the coefficients and the intercepts to get the odds ratio (as supposed to log odds ratio), we conclude that

1. all dose levels lead to favorable outcomes, as a positive coefficient indicates a more favorable outcome
2. the odds of a patient being in the more favorable categories increases as dose level increases, as the increasing coefficients show
3. recall that our reference category is *placebo* (dose = 0), the estimated odds of outcome (y) for *low dose* (dose = 1) is 1.125 times the estimated odds for placebo (dose = 0), the estimated odds of outcome for *medium dose* (dose = 2) is 1.374 times the estimated odds for *placebo* (dose = 0), etc
4. for the reference category *placebo* (dose = 0), the estimated odds of outcome falls into category 1 (*death*) versus all other categories is 0.399, the estimated odds of outcome falls into category 1 (*death*) or category 2 (*vegetative state*) versus all other categories is 0.5956, etc.

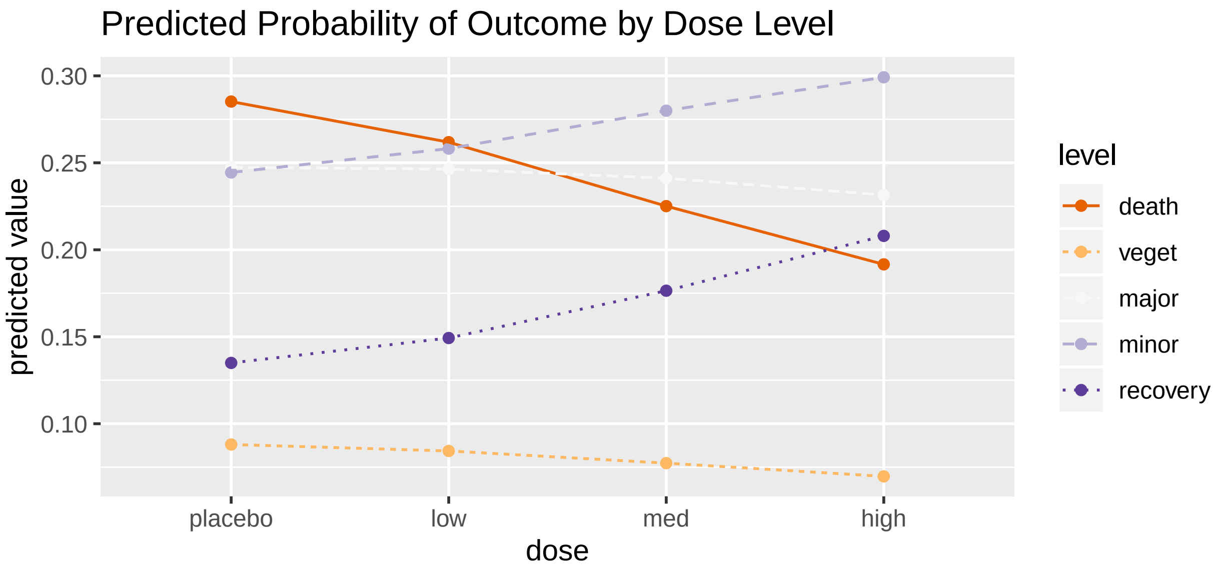
factor(dose)1	factor(dose)2	factor(dose)3
1.125	1.374	1.683

death|veget	veget|major	major|minor	minor|recovery
0.399	0.5956	1.636	6.41

Prediction (Predicted Probability)

Based on the fitted (or predicted) values, the *death* outcome does have a lower relative porportion in the *high dose* group and the *good recovery* outcome has a higer relative porportion in the *high dose* group.

dose	death	veget	major	minor	recovery
placebo	0.2852	0.08806	0.2474	0.2444	0.135
low	0.2619	0.08434	0.2464	0.2582	0.1493
med	0.2251	0.07735	0.2411	0.28	0.1765
high	0.1916	0.06973	0.2315	0.2992	0.208



Collapsing Categories

Next, we collapse five (*death, veget, major, minor, recovery*) to three categories (*unfavorable, major, favorable*) and see that AIC value drops significantly. However, we should note that it is advisable to check literature or domain knowledge before collapsing categories, as it may sometimes result in loss of information or aggregation bias.

	logit	logit_collapse
AIC	2475	1712

The Baseline-Category Logit Model for Nominal Response

For illustration purposes, we also use the `multinom` function from the `nnet` package to fit the baseline-category logit model. Other function such as `vglm` from the `VGAM` package can perform the same task too.

Exponentiating the coefficients and the intercepts again to get the risk ratio (as supposed to log risk ratio), we conclude that

1. the relative risk ratio switching from placebo (dose = 0) to low dose (dose = 1) is 1.032 for being in *veget* vs *death*, the relative risk ratio switching from dose = 0 to 2 is 0.7509 for being in *veget* vs *death*, etc
2. for the reference category *placebo* (dose = 0), the relative risk ratio for being in *veget* vs *death* is 0.4237, the relative risk ratio for being in *major* vs *death* is 0.7797, etc. The summary of the baseline-category logit model is as follows:

```
## # weights:  25 (16 variable)
## initial  value 1290.769206
## iter  10 value 1222.685775
## iter  20 value 1221.583400
## final  value 1221.583198
## converged
```

	(Intercept)	factor(dose)1	factor(dose)2	factor(dose)3
veget	0.4237	1.032	0.7509	0.2195
major	0.7797	1.176	1.574	1.462
minor	0.8136	1.204	1.788	1.658
recovery	0.5424	1.152	1.299	1.758

The p-value of summary of the baseline-category logit model is as follows:

	(Intercept)	factor(dose)1	factor(dose)2	factor(dose)3
veget	0.0003206	0.9281	0.4611	0.008324
major	0.2058	0.5725	0.1086	0.186
minor	0.2885	0.5122	0.03523	0.07077
recovery	0.005326	0.6577	0.4155	0.06842

We see no AIC advantage of fitting the multinomial regression model (e.g. the baseline-category logit model) as supposed to fitting the ordinal regression model (e.g. the cumulative link model).

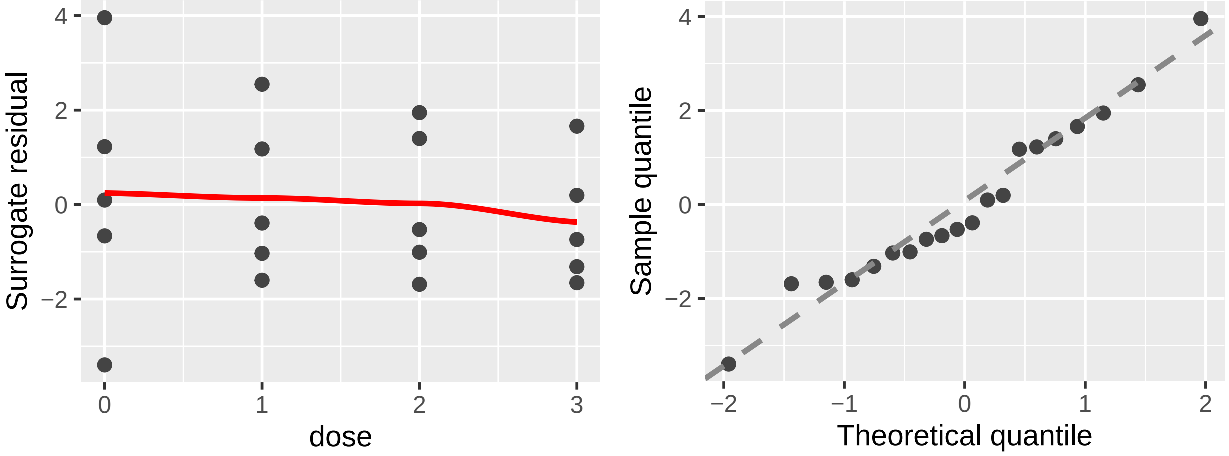
	logit	logit_multi
AIC	2475	2475

Discussion

With higher relative AIC values for all models, we have reason to suspect issues with model fit. Major issues with model fit are perhaps not having enough data points and only including dose as the only covariate.

Second, for the cumulative logit model, all except p-value for the coefficient that captures the difference of estimated odds between *low dose* (dose = 1) vs *placebo* (dose = 0) and between *medium dose* (dose = 2) vs *placebo* (dose = 0) are significant. In contrast, for the baseline-category logit model, more than one p-value for the coefficient and the slope are not significant.

Third, additional constraint for model for ordinal response makes defining residuals challenging. An attempt to perform residual analysis using the `resids` function from the `sure` (SUrrogate REsiduals) package is included:



However, since we have not had much discussion about residual analysis for ordinal regression, we will not go in-depth here but instead needing to plot Residuals vs Fitted plot if time permits.

Lastly, this analysis is meant for a demonstration for fitting the aforementioned models, so further analysis should attempt to include additional covariates such as age, gender, and other health condition.

Reference

Agresti, A. (2010). Analysis of ordinal categorical data (Vol. 656). John Wiley & Sons.

Chuang-Stein, C., & Agresti, A. (1997). Tutorial in Biostatistics A review of tests for detecting a monotone dose–response relationship with ordinal response data. Statistics in Medicine, 16(22), 2599-2618.

Greenwell, B. M., McCarthy, A. J., Boehmke, B. C., & Liu, D. (2018). Residuals and Diagnostics for Binary and Ordinal Regression Models: An Introduction to the sure Package. R J., 10(1), 381.

Reading

Penn State's Polytomous (Multinomial) Logistic Regression

Penn State's Proportional-Odds Cumulative Logit Model

UCLA's Multinomial Ligistic Regression | R Data Analysis Example

UCLA's Ordinal Logistic Regression | R Data Analysis Example

R Code

```r
knitr::opts_chunk$set(echo = FALSE, message=FALSE, warning=FALSE)
library(tidyverse)
library(ggplot2)
library(MASS)
library(pander)
library(reshape2)
library(VGAM)
library(nnet)
library(sure)
# Enter data as a flat table
# the 1977 paper by Chuang-Stein and Agresti
# also p. 207 of the Analysis of ordinal categorical data
trama_flat <- tibble(
  dose = c("placebo", "low", "med", "high"),

  death = c(59, 48, 44, 43),
  veget = c(25, 21, 14, 4),
  major = c(46, 44, 54, 49),
  minor = c(48, 47, 64, 58),
  recovery = c(32, 30, 31, 41),
)
# View data
trama_flat %>% pander
# Prep data
# p. 207 of the Analysis of ordinal categorical data
trama <- tibble(
  state = rep(c("death",
                "veget",
                "major",
                "minor",
                "recovery"), 4),

  dose = c(rep(0, 5),   #
           rep(1, 5),
           rep(2, 5),
           rep(3, 5)),

  count = c(59, 25, 46, 48, 32,
            48, 21, 44, 47, 30,
            44, 14, 54, 64, 31,
            43,  4, 49, 58, 41)
)
# Change state to a factor with specified order
trama <- trama %>%
  mutate(
    state = factor(state, levels = c("death",
                                     "veget",
                                     "major",
                                     "minor",
                                     "recovery"))
  )
```

```r
# Plot data
plot <- ggplot(data = trama, aes(x = state, y = count, fill = state)) +
  geom_bar(stat="identity") +
  scale_fill_brewer(palette="Purples") + # set color
  facet_wrap(~ dose, ncol = 4) +
  theme(axis.ticks.x = element_blank(),  # remove x text and tick
        axis.text.x = element_blank()) +
  labs(title = "Counts of Outcome by Dose Level") +
  coord_fixed(ratio = 0.125) # shrink the plot
plot
# View data
#head(trama) %>% pander
# Fit the cumulative logit model
model_logit <- polr(state ~ factor(dose), weights = count, data = trama)

# Calculate p-val
# UCLA's Ordinal Logistic Regression | R Data Analysis Example
ctable <- summary(model_logit)$coeff
p_val <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
ctable <- cbind(ctable, "p value" = p_val)
ctable %>% pander
#summary(model_logit) %>% pander
# Fit the cumulative link model w/ probit
model_probit <- polr(state ~ factor(dose), weights = count, data = trama, method = "probit")
#AIC(model_probit)
# Fit the cumulative link model w/ cloglog
model_cloglog <- polr(state ~ factor(dose), weights = count, data = trama, method = "cloglog")
#AIC(model_cloglog)
# Obtain AIC
AIC <- cbind(AIC(model_logit), AIC(model_probit), AIC(model_cloglog))
AIC <- as.data.frame(AIC)
colnames(AIC) <- c("logit", "probit", "cloglog")
rownames(AIC) <- c("AIC")
AIC %>% pander
# Get odds by exp(beta)
exp(model_logit$coeff) %>% pander

# Get odds by exp(alpha_i)
exp(model_logit$zeta) %>% pander
# Fitted (or predicted) values
fitted_val <- model_logit$fitted[seq(1, 20, by = 5), ]
fitted_val <- as.data.frame(fitted_val)
fitted_val <- tibble(
  dose = c("placebo", "low", "med", "high"),
  death = fitted_val$death,
  veget = fitted_val$veget,
  major = fitted_val$major,
  minor = fitted_val$minor,
  recovery = fitted_val$recovery
)
fitted_val %>% pander

# Check to see if rowSums to 1
```

```
#rowSums(fitted) %>% pander #good
# Reshape data for plotting
fitted_melt <- melt(fitted_val, id.vars = "dose",
                    variable.name = "level",
                    value.name = "probability")
#fitted_melt
# Plot it
fitted_plot <- ggplot(fitted_melt, aes(x = dose, y = probability, group = level, color = level)) +
  geom_point() +
  geom_line(aes(linetype = level)) +
  scale_x_discrete(limits=c("placebo","low","med", "high")) + # reorder dose level
  labs(title = "Predicted Probability of Outcome by Dose Level", y = "predicted value") +
  coord_fixed(ratio = 8) + # shrink the plot
  scale_color_brewer(palette = "PuOr") # change color
fitted_plot
# CI for exp(beta)
#exp(confint(model_logit)) %>% pander
#fct_collapse()
trama_collapse <- tibble(
  state = rep(c("unfavorable",
                "major",
                "favorale"), 4),

  dose = c(rep(0, 3),   #
           rep(1, 3),
           rep(2, 3),
           rep(3, 3)),

  count = c(59 + 25, 46, 48 + 32,
            48 + 21, 44, 47 + 30,
            44 + 14, 54, 64 + 31,
            43 + 4,  49, 58 + 41)
)
# View data
#head(trama_collapse) %>% pander
# Change state to a factor with specified order
trama_collapse <- trama_collapse %>%
  mutate(
    state = factor(state, levels = c("unfavorable", "major", "favorale"))
  )
# # Fit the cumulative logit model
# model_collapse <- polr(state ~ factor(dose), weights = count, data = trama_collapse)
# pander(summary(model_collapse)) # error: response must have 3 or more levels
# Fit the cumulative logit model
model_collapse <- polr(state ~ factor(dose), weights = count, data = trama_collapse)
AIC <- cbind(AIC(model_logit), AIC(model_collapse))
AIC <- as.data.frame(AIC)
colnames(AIC) <- c("logit", "logit_collapse")
rownames(AIC) <- c("AIC")
AIC %>% pander
# # Fit the baseline-category logit model using VGAM
# #model_3 <- vglm(cbind(death, veget, major, minor, recovery) ~ dose, data = trama2, family = multinom
#
```

```r
# # Set death as baseline category
# model_3 <- vglm(cbind(veget, major, minor, recovery, death) ~ dose, data = trama2, family = multinomi
# #pander(summary(model_3))
# #pander(summary(model_3)$coef)
# summary(model_3) #AIC() won't work or is inaccurate
# Change this if want to make state other than death reference category
trama$state2 <- relevel(trama$state, ref = "death")

# Fit the baseline-category logit model using nnet
model_multi <- multinom(state2 ~ factor(dose), weights = count, data = trama)
#pander(summary(model_multi))
exp(summary(model_multi)$coeff) %>% pander
# Calculate p-val
# UCLA's Multinomial Logistic Regression | R Data Analysis Example
z_stat <- summary(model_multi)$coefficients / summary(model_multi)$standard.errors
p_val <- (1 - pnorm(abs(z_stat), 0, 1)) * 2
p_val %>% pander
AIC <- cbind(AIC(model_logit), AIC(model_multi))
AIC <- as.data.frame(AIC)
colnames(AIC) <- c("logit", "logit_multi")
rownames(AIC) <- c("AIC")
AIC %>% pander
# Obtain surrogate-based residuals
# Plot Residuals vs Covariate and Normal Q-Q plots for the logit model (model_1)
set.seed(1234)
sres <- resids(model_logit) # for reproducibility
p1 <- autoplot(sres, what = "covariate", x = trama$dose, xlab = "dose") +
  coord_fixed(ratio = 0.3)  # shrink the plot
p2 <- autoplot(sres, what = "qq", distribution = qnorm) +
  coord_fixed(ratio = 0.39)

grid.arrange(p1, p2, ncol = 2)
```