BIOS 6110
Applied Categorical Data Analysis

Instructor: Yuan Huang, Ph.D.

Department of Biostatistics

Fall 2017

Part VIII

Multicategory Logit Models

Introduction

Previously, we have learnt logistic models for binary/dichotomous response. For example, $Y = 1$ for event and $Y = 0$ for no event.

Specifically, we are modeling

$$\text{logit}(P(Y=1)) = \log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \log\left(\frac{P(Y=1)}{P(Y=0)}\right)$$

The predictors can be categorical or continuous.

When predictors are all categorical, we discuss a special case with XYZ three-way contingency table for $Y = 2$ level, $Z = K$ level.

- For homogeneous association (main effect model), dependence can be measured by the logistic model or CMH statistics

Introduction (cont.)

In practice, there are many cases where the response has more than two categories (polychotomous).

Those categories can be nominal or ordinal.

- Nominal: Primary food choice (Fish, Invertebrates, and Others)
- Ordinal: Political Ideology (Very liberal, Slightly liberal, Moderate, Slightly conservative, and Very conservative)

Overview

When Y is nominal,

- Baseline-Category Logits Model
  - One single continuous predictor
  - Categorical predictors

When Y is ordinal,

- Baseline-Category Logits Model
- Cumulative Logit Model
  - The proportional Odds Property
  - Latent variable motivation
- Adjacent-Categories Logits
- Continuation-Ratio Logits

Learning Objectives

1. Understand model applicability
2. Interpret parameters and conduct hypothesis testing
3. Estimate response probabilities

Reading: Agresti (2002), Section 6.1-6.3

Note,

- Understand examples for different models.
- Model checking is very similar to logistic cases and therefore is not addressed separately in our textbook.
- For example, goodness of fit only applies to grouped data.
- When Y has more than two categories, we could provide dependence measure to XYZ contingency table where Y has more than 2 levels. Alternatively, we can use generalized CMH statistics. This is covered in book section 6.4 and we will skip it in class.

Example: Alligator Data

In a study by the Florida Game and Fresh Water Fish Commission on the foods that alligators in the wild choose to eat, 59 alligators in Lake George, Florida, were sampled and the primary food type found in the alligators stomach was recorded along with the alligator length.

	Primary Food Choice (Y)		
Alligator Length (X)	Fish (F)	Invertebrates (I)	Other (O)
1.24	0	1	0
1.30	0	1	0
1.30	0	1	0
1.32	1	0	0
1.32	1	0	0
⋮	⋮	⋮	⋮
3.66	1	0	0
3.68	0	0	1
3.71	1	0	0
3.89	1	0	0

Note that alligators of the same length each has its own row, i.e., ungrouped.

Multinomial Distribution

Y has J categories, where $J \geq 2$.

- Joint distribution $\{\pi_1, \pi_2, \dots, \pi_J\}$, where

$$\pi_j = \text{the response probability for category } j, \quad j = 1, \dots, J$$

The π's satisfies $\sum_j \pi_j = 1$.

- $n_1 =$ number of observations having outcomes in category 1
$n_2 =$ number of observations having outcomes in category 2, ...
where the total sample $n = \sum_j n_j$.

$\{n_1, n_2, \dots, n_J\}$ has a multinomial distribution:

$$\Pr(\# \text{ of } Y \text{ in category } 1 = n_1, \cdots, \# \text{ of } Y \text{ in category } J = n_J)$$
$$= \frac{n!}{n_1! n_2! \cdots n_J!} \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_J^{n_J}$$

- Special case, for ungrouped data, $n = 1$ for each observation.

Baseline-Category Logits

Form logits: For J outcomes, there are totally $J(J-1)/2$ logits (log odds) that can be formulated, only $J-1$ are non-redundant.

Logit models for nominal response variables pair each category with a baseline category. When the last category (J) is the baseline, the baseline-category logits are

$$\log\left(\frac{\pi_j}{\pi_J}\right), \;\; j = 1, \ldots, J-1$$

The baseline-category logit model with a predictor x is

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \;\; j = 1, \ldots, J-1$$

Separate set of parameters (α_j, β_j) for each logit.
e^{β_j} is the multiplicative effect of a 1-unit increase in x on the conditional odds of response j given that the response is either j or J. That is, of the odds of j versus the baseline J.

Baseline-Category Logits (cont.)

There are $J-1$ equations. These equations will be fit simultaneously, resulting in smaller standard errors of parameter estimates than when fitting them separately. For instance

- For $J=3$, the 2 equations are

$$\log\left(\frac{\pi_1}{\pi_3}\right) = \alpha_1 + \beta_1 x,$$

  and

$$\log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

- For $J=2$, the model equation is

$$\log\left(\frac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x$$

  which is the simple logistic regression since $\pi_1 + \pi_2 = 1$

Baseline-Category Logits (cont.)

The $J-1$ model equations jointly determine equations for all other pairs of categories. Therefore, the baseline category is arbitrary and does not affect model fit.

For example, for an arbitrary pair of categories a and b,

$$\begin{aligned}\log\left(\frac{\pi_a}{\pi_b}\right) &= \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right)\\ &= \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right)\\ &= (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x)\\ &= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x\end{aligned}$$

The probability $\pi_j = \Pr(Y \text{ is in category } j)$ is given by

$$\pi_j = \frac{\exp\{\alpha_j + \beta_j\, x\}}{\sum_h \exp\{\alpha_h + \beta_h\, x\}}, \quad j = 1, \ldots, J$$

with $\alpha_J = \beta_J = 0$ assuming J is the reference category.

SAS output: Alligator Data

Analysis of Maximum Likelihood Estimates

Parameter	food	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	F	1	1.6177	1.3073	1.5314	0.2159
Intercept	I	1	5.6974	1.7938	10.0881	0.0015
length	F	1	-0.1101	0.5171	0.0453	0.8314
length	I	1	-2.4654	0.8997	7.5101	0.0061

The ML prediction equations are

$$\log(\hat{\pi}_1/\hat{\pi}_3) = 1.618 - 0.110\,x$$

and

$$\log(\hat{\pi}_2/\hat{\pi}_3) = 5.697 - 2.465\,x$$

The estimated log odds that the response is "fish" rather than "invertebrate" equals

$$\begin{aligned}\log(\hat{\pi}_1/\hat{\pi}_2) &= (1.618 - 5.697) + [-0.110 - (-2.465)]\,x \\ &= -4.08 + 2.355\,x\end{aligned}$$

- The estimates for a particular equation are interpreted as in binary logistic regression, conditional on the event that the outcome was one of those two categories.

  For instance, given that the primary food type is fish or invertebrate, for alligators of length $x + 1$ meters, the estimated odds that primary food type is "fish" rather than "invertebrate" equal $\exp(2.355) = 10.5$ times the estimated odds at length x meters

- The hypothesis that primary food choice is independent of alligator length is

$$H_0 : \beta_1 = \beta_2 = 0$$

  The likelihood-ratio test statistic equals 16.8, with df $= 2$. The P-value of 0.0002 provides strong evidence of a length effect.

- Estimating Response Probabilities

  Given length x, the estimated probabilities of the outcomes (Fish, Invertebrate, Other) are

$$
\begin{aligned}
\hat{\pi}_1 &= \frac{\exp\{1.62 - 0.11\,x\}}{1 + \exp\{1.62 - 0.11\,x\} + \exp\{5.70 - 2.47\,x\}} \\
\hat{\pi}_2 &= \frac{\exp\{5.70 - 2.47\,x\}}{1 + \exp\{1.62 - 0.11\,x\} + \exp\{5.70 - 2.47\,x\}} \\
\hat{\pi}_3 &= \frac{1}{1 + \exp\{1.62 - 0.11\,x\} + \exp\{5.70 - 2.47\,x\}} \\
&= 1 - \hat{\pi}_1 - \hat{\pi}_2
\end{aligned}
$$

  The "1" term in each denominator is generated by $\exp\{\hat{\alpha}_3 + \hat{\beta}_3\,x\} = 1$ because $\hat{\alpha}_3 = \hat{\beta}_3 = 0$ with the baseline category.

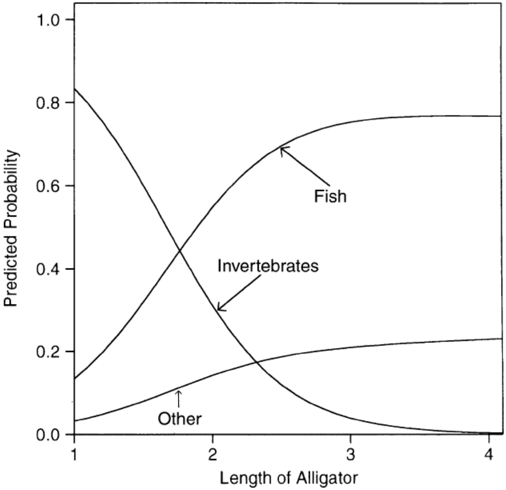

Figure 6.1. Estimated probabilities for primary food choice.

Categorical Predictors

When explanatory variables are entirely categorical, a contingency table can summarize the data. If the data are not sparse, one can test model goodness of fit using the X^2 or G^2 statistics

Example: Afterlife Data

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	371	49	74
	Male	250	45	71
Black	Female	64	9	15
	Male	25	5	13

Y = belief in life after death (1-Yes, 2-Undecided, 3-No)

x_1 = gender: 1 for female; 0 for male

x_2 = race: 1 for whites; 0 for blacks

Use "No" as the baseline category for Y, the multicategory logit model is

$$\log(\pi_j/\pi_3) = \alpha_j + \beta_j^G x_1 + \beta_j^R x_2, \quad j = 1, 2$$

where G and R superscripts identify the gender and race parameters

- β_1^G is the conditional log odds ratio between gender and response categories 1 and 3 (Yes and No), given race.

SAS output: Alligator Data

Parameter		belief	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		undecide	1	-0.7582	0.3614	4.4031	0.0359
Intercept		yes	1	0.8828	0.2426	13.2390	0.0003
gender	female	undecide	1	0.1051	0.2465	0.1817	0.6699
gender	female	yes	1	0.4186	0.1713	5.9737	0.0145
race	white	undecide	1	0.2712	0.3541	0.5863	0.4438
race	white	yes	1	0.3420	0.2370	2.0814	0.1491

Therefore,

$$\begin{aligned}\log(\pi_1/\pi_3) &= 0.883 + 0.419x_1 + 0.342x_2 \\ \log(\pi_2/\pi_3) &= -0.758 + 0.105x_1 + 0.271x_2\end{aligned}$$

- For females the estimated odds of response "yes" versus "no" on afterlife is $\exp(\hat{\beta}_1^G) = \exp(0.419) = 1.5$ times those for males, controlling for race.
- For whites, the estimated odds of response "yes" versus "no" on afterlife is $\exp(\beta_1^R) = \exp(0.342) = 1.4$ times those for blacks, controlling for gender.

Based on the prediction models,

$$\begin{aligned}\log(\pi_1/\pi_3) &= 0.883 + 0.419x_1 + 0.342x_2 \\ \log(\pi_2/\pi_3) &= -0.758 + 0.105x_1 + 0.271x_2,\end{aligned}$$

we can obtain estimated probabilities.

To illustrate, for white females ($x_1 = 1, x_2 = 1$), the estimated probability of response 1 ("yes") on afterlife is

$$\frac{\exp\{0.883 + 0.419(1) + 0.342(1)\}}{1 + \exp\{0.883 + 0.419(1) + 0.342(1)\} + \exp\{-0.758 + 0.105(1) + 0.271(1)\}}$$

$= 0.76$

All these probabilities are shown below:

Estimated Probabilities for Belief in Afterlife

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	0.76	0.10	0.14
	Male	0.68	0.12	0.20
Black	Female	0.71	0.10	0.19
	Male	0.62	0.12	0.26

Cumulative Logit Models for Ordinal Responses

Question: if there is a natural ordering in the response, how to use such information in the modeling?

A cumulative probability for Y is defined as

$$\Pr(Y \leq j) = \pi_1 + \cdots + \pi_j, j = 1, \ldots, J$$

The cumulative probabilities are non-decreasing:

$$\Pr(Y \leq 1) \leq \Pr(Y \leq 2) \leq \cdots \leq \Pr(Y \leq J) = 1$$

The logits of the cumulative probabilities are

$$\begin{aligned}
\mathrm{logit}[\Pr(Y \le j)] &= \log\left[\frac{\Pr(Y \le j)}{1-\Pr(Y \le j)}\right] \\
&= \log\left[\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right], \quad j = 1, \ldots, J-1
\end{aligned}$$

These are called cumulative logits. Note that there are only $J-1$ logits to be model, because the last one, $\Pr(Y \le J)$, is always 1.

- For $J = 3$, for example, 2 equations are used to define the model:

$$\begin{aligned}
\mathrm{logit}[\Pr(Y \le 1)] &= \log[\pi_1/(\pi_2 + \pi_3)] \\
\mathrm{logit}[\Pr(Y \le 2)] &= \log[(\pi_1 + \pi_2)/\pi_3]
\end{aligned}$$

Example: Political Ideology and Party Affiliation

Political Ideology by Gender and Political Party

		Political Ideology				
Gender	Political Party	Very Liberal	Slightly Liberal	Moderate	Slightly Conservative	Very Conservative
Female	Democratic	44	47	118	23	32
	Republican	18	28	86	39	48
Male	Democratic	36	34	53	18	23
	Republican	12	18	62	45	51

In this example, there are $J = 5$ response categories. Let x be an indicator variable for political party, with $x = 1$ for Democrats and $x = 0$ for Republicans.

The model we want to fit is

$$\text{logit}[\Pr(Y \leq j)] = \alpha_j + \beta x, \quad j = 1, 2, 3, 4$$

Cumulative Logit Models with Proportional Odds Property

For an explanatory variable x, the cumulative logit model is

$$\text{logit}[\Pr(Y \leq j)] = \alpha_j + \beta x, \quad j = 1, \ldots, J-1$$

Note that β does not have a j subscript: This model assumes that the effect x is identical for all $J-1$ cumulative logits.

How many parameters does this model have?

Because

$$\Pr(Y \leq j) = \frac{\exp\{\alpha_j + \beta x\}}{1 + \exp\{\alpha_j + \beta x\}},$$

we can plot the estimated cumulative probabilities.

$\rightarrow$ At any fixed x value, the curves have the same ordering as the cumulative probabilities.

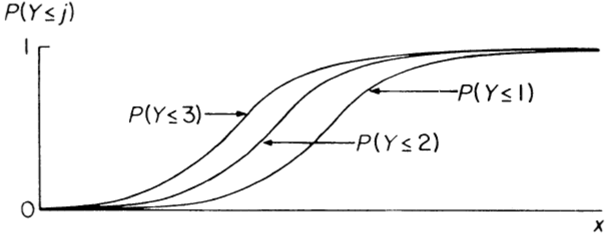

Figure 6.2. Depiction of cumulative probabilities in proportional odds model.

Depiction of cumulative probabilities in proportional odds model for $\beta > 0$

The Proportional Odds Property

For two values x_1 and x_2 of x, the odds ratio comparing the cumulative probabilities is

$$\begin{aligned}
&\frac{\Pr(Y \le j | X = x_2) / \Pr(Y > j | X = x_2)}{\Pr(Y \le j | X = x_1) / \Pr(Y > j | X = x_1)} \\
&= \frac{\exp\{\alpha_j + \beta x_2\}}{\exp\{\alpha_j + \beta x_1\}} \\
&= \exp\{\beta(x_2 - x_1)\}
\end{aligned}$$

That is, the log of this odds ratio is $\beta(x_2 - x_1)$, which is proportional to the difference $x_2 - x_1$. This is true for any j.

Therefore, we also call this model as proportional odds model.

SAS Output: Political Ideology and Party Affiliation

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 1	1	-2.4690	0.1318	350.8122	<.0001
Intercept 2	1	-1.4745	0.1091	182.7151	<.0001
Intercept 3	1	0.2371	0.0948	6.2497	0.0124
Intercept 4	1	1.0695	0.1046	104.6082	<.0001
party	1	0.9745	0.1291	57.0182	<.0001

- The estimated effect of political party is $\hat{\beta} = 0.975$ ($SE = 0.129$)
- For any fixed j, the estimated odds in the liberal direction rather than the conservative direction (i.e., $Y \leq j$ rather than $Y > j$) for Democrats (party=1) vs Republicans (party=0) is $\exp(0.975) = 2.65$. That is, Democrats tend to be more liberal than Republicans.

- The Wald statistic for testing $H_0 : \beta = 0$ is 57.0182 with $df = 1$. There is an strong indication of association ($p < 0.0001$).
- A 95% confidence interval for β is $0.975 \pm 1.96(0.129) = (0.72, 1.23)$. The confidence interval for the odds ratio of cumulative probabilities equals $(\exp(0.72), \exp(1.23)) = (2.1, 3.4)$. The odds of being at the liberal end of the political ideology scale is at least twice as high for Democrats as for Republicans. The effect is practically significant as well as statistically significant.

- Estimating cumulative probabilities:
  Using the formula that

$$\Pr(Y \leq j) = \frac{\exp\{\alpha_j + \beta x\}}{1 + \exp\{\alpha_j + \beta x\}}$$

the first estimated cumulative probability for Democrats ($x = 1$) is

$$\Pr(Y \leq 1) = \frac{\exp\{-2.469 + 0.975(1)\}}{1 + \exp\{-2.469 + 0.975(1)\}} = 0.18$$

Similarly,

$$\Pr(Y \leq 2) = 0.38, \Pr(Y \leq 3) = 0.77, \Pr(Y \leq 4) = 0.89$$

The estimated probability that a Democrat is moderate (category 3) is

$$\Pr(Y = 3) = \Pr(Y \leq 3) - \Pr(Y \leq 2) = 0.77 - 0.38 = 0.39$$

Checking Model Fit

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
3.9106	3	0.2713

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	3.6877	3	1.2292	0.2972
Pearson	3.6629	3	1.2210	0.3002

- The score test of the proportional odds assumption. Under H_0, the effect is β for all four cumulative logit models. Under H_1, $\beta_j, j \in 1, \ldots, 4$ will be estimated. This statistic equals 3.9 with $df = 3$ and $p = 0.2714$, indicating no evidence of lack of fit.
- The Pearson X^2 and deviance G^2 statistics can be applied to assess the goodness of fit for grouped data. When nearly all expected cell counts are at least about 5, these test statistics have approximate chi-squared distributions. Here $X^2 = 3.7$ and $G^2 = 3.7$, and p-values are about 0.3. The model fits adequately.

Latent Variable Motivation

An unobserved variable assumed to underlie what we actually observe is called a latent variable. Let Y^* denote a latent variable with distribution function $G(y^* - \eta)$. Suppose

$$-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_{J-1} < \alpha_J = \infty$$

are cutpoints of the continuous scale for Y^* such that the observed response Y satisfies

$$Y = j \text{ if } \alpha_{j-1} < Y^* \leq \alpha_j$$

In other words, we observe Y in category j when the latent variable falls in the j-th interval of values. Assume $\eta(x) = \beta x$ is the mean of Y^*. Then

$$\Pr(Y \leq j|x) = \Pr(Y^* \leq \alpha_j|x) = G(\alpha_j - \beta x)$$

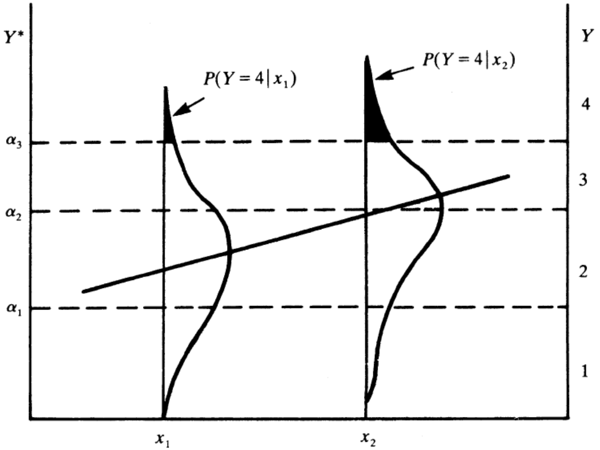

Figure 6.4. Ordinal measurement, and underlying regression model for a latent variable.

Invariance to Choice of Response Categories

In latent variable motivation, the same parameters occur for the effects regardless of how the cutpoints $\{\alpha_j\}$ discretize the real line to form the scale for Y. i.e., the effect parameters are invariant to the choice of categories for Y. This nice feature of the model makes it possible to compare estimates from studies using different response scales.

- If a continuous variable measuring political ideology has a linear regression with some predictor variables, then the same effect parameters apply to a discrete version of political ideology with categories (liberal, moderate, conservative).
- If one models political ideology using (very liberal, slightly liberal, moderate, slightly conservative, very conservative) and the other uses (liberal, moderate, conservative), the parameters for the effect of a predictor are roughly the same.
- For instance, we combine the two liberal categories and combine the two conservative categories so that there are 3 categories for the response. The estimated party affiliation effect changes from 0.975 (SE = 0.129) to 1.006 (SE = 0.132). Interpretations are unchanged.

Adjacent-Categories Logits

The adjacent-categories logits are

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right), \quad j = 1, \ldots, J-1$$

For example, with $J = 3$, these logits are $\log(\pi_2/\pi_1)$ and $\log(\pi_3/\pi_2)$.

With a predictor x, the adjacent-categories logit model has form

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_j x, \quad j = 1, \ldots, J-1$$

The adjacent-categories logits, like the baseline-category logits, determine the logits for all pairs of response categories.

Adjacent-Categories Logits (cont.)

Without further structure, the adjacent-categories logit model

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_j x, \quad j = 1, \dots, J-1,$$

it is just a re-parameterization of the baseline logit model. In other words, we can obtain the parameter estimate from one model to the other model.

For example, if we have the estimated proportional odds model with the last category as the baseline when $J = 3$ as follows.

$$\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_3}\right) = a_1 + b_1 x, \quad \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right) = a_2 + b_2 x$$

Then the adjacent-categories logits model can be obtained accordingly as

$$\log\left(\frac{\hat{\pi}_2}{\hat{\pi}_1}\right) = (a_2 - a_1) + (b_2 - b_1)x, \quad \log\left(\frac{\hat{\pi}_3}{\hat{\pi}_2}\right) = -a_2 - b_2 x$$

Adjacent-Categories Logits - A Simpler Version

A simpler version of the model is

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta x, \quad j = 1, \ldots, J-1$$

That is, the effects $\{\beta_j = \beta\}$ of x on the odds of making the higher instead of the lower response are identical for each pair of adjacent response categories

Likelihood ratio test can be applied to determine the $H_0 : \beta_1 = \beta_2 = \ldots = \beta_{J-1}$.

Adjacent-Categories Logits - Estimate Probabilities

We can also estimate the probabilities from the models. Still take the $J = 3$ case as an example,

$$\log\left(\frac{\hat{\pi}_2}{\hat{\pi}_1}\right) = a_1 + b_1 x, \quad \log\left(\frac{\hat{\pi}_3}{\hat{\pi}_2}\right) = a_2 + b_2 x$$

Then we have

$$\hat{\pi}_2 = e^{a_1 + b_1 x}\hat{\pi}_1 \quad \hat{\pi}_3 = e^{a_2 + b_2 x}\hat{\pi}_2$$

Assume $\hat{\pi}_1 = \phi$, then we have

$$\begin{aligned}
\hat{\pi}_1 &= \phi \\
\hat{\pi}_2 &= e^{a_1 + b_1 x}\phi \\
\hat{\pi}_3 &= e^{a_2 + b_2 x}e^{a_1 + b_1 x}\phi = e^{(a_1 + a_2) + (b_1 + b_2)x}\phi
\end{aligned}$$

By $\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 = 1$, we can solve for

$$\phi = \frac{1}{1 + e^{a_1 + b_1 x} + e^{(a_1 + a_2) + (b_1 + b_2)x}}$$

Continuation-Ratio Logits

Continuation-Ratio Logits model uses another approach to forms logits for ordered response categories in a sequential manner.

For J categories, those logits are

$$\log\left(\frac{\pi_1}{\pi_2}\right)$$
$$\log\left(\frac{\pi_1+\pi_2}{\pi_3}\right)$$
$$\ldots$$
$$\log\left(\frac{\pi_1+\pi_2+\ldots+\pi_{J-1}}{\pi_J}\right)$$

Each logit can be understood as a binary response that contrasts each category with a grouping of categories from lower levels of the response.

The predictor effect can be interpreted accordingly to the logit. e.g., in the 2nd equation, the interpretation for predictor X's effect is: (controlling the rest predictors), for one unit change in predictor X, the estimated odds that it is category 1 or 2 rather than 3 changes by a multiplicative factor of e^{β_x}.

We will not cover further details in this model. However, the estimation of probabilities, the Wald test, and Wald CI can be similarly constructed.

Overdispersion

Overdispersion means that the actual covariance matrix of y_i exceeds that specified by the multinomial model $V(y_i) = n_i[\text{Diag}(\pi_i) - \pi_i\pi_i^T]$,

We may consider overdispersion for grouped data when the model has already contained all covariates that worth considering and the overall Pearson X^2 is substantially larger than its df.

In this situation, we may use the quasi-likelihood approach and introduce a scale parameter σ^2 so that $V(y_i) = n_i[\text{Diag}(\pi_i) - \pi_i\pi_i^T]\sigma^2$.

A usual estimate for σ^2 is

$$\hat{\sigma}^2 = \frac{X^2}{df}$$

Once we have $\hat{\sigma}^2$, we can use $scale = \hat{\sigma}$ in the model statement in SAS.

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	3.6877	3	1.2292	0.2972
Pearson	3.6629	3	1.2210	0.3002

$\rightarrow$

$$\hat{\sigma}^2 = 1.221$$
$$\hat{\sigma} = 1.105$$

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square
Intercept 1	1	-2.4690	0.1318	350.8122
Intercept 2	1	-1.4745	0.1091	182.7151
Intercept 3	1	0.2371	0.0948	6.2497
Intercept 4	1	1.0695	0.1046	104.6082
party	1	0.9745	0.1291	57.0182

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
party	2.650	2.058	3.412

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square
Intercept 1	1	-2.4690	0.1457	287.3096
Intercept 2	1	-1.4745	0.1205	149.6408
Intercept 3	1	0.2371	0.1048	5.1184
Intercept 4	1	1.0695	0.1156	85.6725
party	1	0.9745	0.1426	46.6970

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
party	2.650	2.004	3.504

- Upper table: original fit
- Lower table: adjusted fit for overdispersion
- Both table has same estimated coefficients
- Adjusted fit has larger standard
- Adjusted fit has wider CI
- Recall that using this approach w not change the parameter estima but the new estimated standard is $\hat{\sigma}$ times of the original and he the new Wald CI is wider.