Multinomial outcome logit models

B.D. in Business Administration and Economics
Course in Quantitative Methods III

Rosario Barone
University of Rome "Tor Vergata"

rosario.barone@uniroma2.it

Let Y be a categorical response with J categories. *Multicategory logit models* for nominal response variables simultaneously describe log odds for all $\binom{J}{2}$ pairs of categories.

Let $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$ with fixed $\mathbf{x}$, $\sum_j \pi_j(\mathbf{x}) = 1$.

For observations at that setting, we treat the counts at the J categories of Y as multinomial with probabilities $\{\pi_1(\mathbf{x}) \ldots, \pi_J(\mathbf{x})\}$

Alligator Food Choice Example

TABLE 7.1 Primary Food Choice of Alligators

Lake	Gender	Size (m)	Primary Food Choice: Fish	Invertebrate	Reptile	Bird	Other
Hancock	Male	≤ 2.3	7	1	0	0	5
		> 2.3	4	0	0	1	2
	Female	≤ 2.3	16	3	2	2	3
		> 2.3	3	0	1	2	3
Oklawaha	Male	≤ 2.3	2	2	0	0	1
		> 2.3	13	7	6	0	0
	Female	≤ 2.3	3	9	1	0	2
		> 2.3	0	1	0	1	0
Trafford	Male	≤ 2.3	3	7	1	0	1
		> 2.3	8	6	6	3	5
	Female	≤ 2.3	2	4	1	1	4
		> 2.3	0	1	0	0	0
George	Male	≤ 2.3	13	10	0	2	2
		> 2.3	9	0	0	1	2
	Female	≤ 2.3	3	9	1	0	1
		> 2.3	8	1	0	0	1

Source: Data courtesy of Clint Moore, from an unpublished manuscript by M. F. Delaney and C. T. Moore.

Baseline-Category Logits

Logit models pair each response category with a **baseline category**, often the last one or the most common one. The J-th baseline-category logit is

$$\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j' \mathbf{x}, \quad j = 1, \ldots, J-1$$

simultaneously describes the effects of $\mathbf{x}$ on the $J-1$ underlying logit models.

- the effects vary according to the response paired with the baseline.

- these $J-1$ equations determine parameters for logits with other pairs of response categories:

$$\log \frac{\pi_a(\mathbf{x})}{\pi_b(\mathbf{x})} = \log \frac{\pi_a(\mathbf{x})}{\pi_J(\mathbf{x})} - \log \frac{\pi_b(\mathbf{x})}{\pi_J(\mathbf{x})}$$

Estimating response probabilities

The equation that expresses multinomial logit models directly in terms of response probabilities $\{\pi_j(\mathbf{x})\}$ is

$$\pi_j(\mathbf{x}) = \frac{\exp\left(\alpha_j + \boldsymbol{\beta}_j \mathbf{x}\right)}{1 + \sum_{h=0}^{J-1} \exp\left(\alpha_h + \boldsymbol{\beta}_h \mathbf{x}\right)}$$

with $\alpha_J = 0$ and $\boldsymbol{\beta}_J = 0$

Note that $\sum_j \pi_j(\mathbf{x}) = 1$ and that for $J = 2$ the response variable is binomial and the model becomes a binary logit model.

Maximum Likelihood Inference

To estimate the unknown parameters, we need to maximize the likelihood subject to $\{\pi_j(\mathbf{x})\}$ simultaneously satisfying the $J-1$ equations that specify the model;

for $i = 1, \ldots, n$ let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})$ represent the multinomial trial for subject i, where $y_i = 1$ when the response is in category j and $y_i = 0$ otherwise: therefore, $\sum_j y_{ij} = 1$;

let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ denote the vector of covariates for the i-th observed individual ;

let $\boldsymbol{\beta} = (\beta_{j1}, \ldots, \beta_{jp})$ denote the parameters for the j-th logit model.

Maximum Likelihood Inference

By considering the n observations, the log-likelihood function is

$$\ell(\alpha, \boldsymbol{\beta}) = \sum_{i=1}^{n} \log \left[\prod_{j=1}^{J} \pi_j(\mathbf{x}_i)^{y_{ij}} \right]$$

By considering $\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$ and $y_{iJ} = 1 - \sum_{i=1}^{J-1}$, we rewrite the contribution to the log likelihood of the i-th subject as

$$\log \left[\prod_{j=1}^{J} \pi_j(\mathbf{x}_i)^{y_{ij}} \right] =$$

$$\sum_{j=1}^{J-1} y_{ij} \log \pi_j(\mathbf{x}_i) + \left(1 - \sum_{j=1}^{J-1} y_{ij} \right) \log \left(1 - \sum_{j=1}^{J-1} \pi_j(\mathbf{x}_i) \right)$$

Maximum Likelihood Inference

$$\log\left[\prod_{j=1}^{J}\pi_j(\mathbf{x}_i)^{y_{ij}}\right] =$$

$$\sum_{j=1}^{J-1} y_{ij}\log\frac{\pi_j(\mathbf{x}_i)}{1-\sum_{j=1}^{J-1}\pi_j(\mathbf{x}_i)} + \log\left(1-\sum_{j=1}^{J-1}\pi_j(\mathbf{x}_i)\right)$$

By considering the n individuals, we get

$$\ell(\alpha,\boldsymbol{\beta}) =$$

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{J-1} y_{ij}(\alpha_j+\boldsymbol{\beta}_j'\mathbf{x}) - \log\left[1+\sum_{j=1}^{J-1}\exp(\alpha_j+\boldsymbol{\beta}_j'\mathbf{x}_i)\right]\right)$$

Maximum Likelihood Inference

$$\ell(\alpha, \boldsymbol{\beta}) = \sum_{j=1}^{J-1} \left[\alpha_j \left(\sum_{i=1}^{n} y_{ij} \right) + \sum_{k=1}^{p} \beta_{jk} \left(\sum_{i=1}^{n} y_{ij} x_{ik} \right) \right] - \sum_{i=1}^{n} \log \left[1 + \sum_{j=1}^{J-1} \exp\left(\alpha_j + \boldsymbol{\beta}_j' \mathbf{x}_i \right) \right]$$

where:

- $\sum_{i=1}^{n} y_{ij} x_{ik}$ is the sufficient statistic for β_{jk} with $j = 1, \ldots, J-1$ and $k = 1, \ldots, p$;
- $\sum_{i=1}^{n} y_{ij}$ is the sufficient statistic for α_j with $j = 1, \ldots, J-1$;

Maximum Likelihood Inference

- The log-likelihood is concave, and the Newton-Raphson method yields the ML parameter estimates.
- The estimators have large-sample normal distributions.
- Their asymptotic standard errors are square roots of diagonal elements of the inverse information matrix.

Alternative approach

An alternative fitting approach fits binary logit models separately for the $J-1$ pairings of responses:

- for $j=1$ alone, using observation 1 and J for estimating α_1 and $\boldsymbol{\beta}_1$;
- for $j=2$ alone, using observation 2 and J for estimating α_2 and $\boldsymbol{\beta}_2$;
- ...
- for $j=J-1$ alone, using observation $J-1$ and J for estimating α_{J-1} and $\boldsymbol{\beta}_{J-1}$;

Alternative approach

$$\log \frac{\pi_j(\mathbf{x})/(\pi_j(\mathbf{x})+\pi_J(\mathbf{x}))}{\pi_J(\mathbf{x})/(\pi_j(\mathbf{x})+\pi_J(\mathbf{x}))}$$

The separate-fitting estimates differ from the ML estimates for simultaneous fitting of the $J-1$ logits. They are less efficient.

They are less efficient, tending to have larger standard errors. However, Begg and Gray (1984). showed that the efficiency loss is minor when the response category having highest prevalence is the baseline.

MC Logit Model as Multivariate GLM

A GLM having univariate response variable in the natural exponential family is

$$g(\boldsymbol{\mu_i}) = \mathbf{x}_i{}'\boldsymbol{\beta}$$

with:

- link function g;
- expected response $\mu_i = E(Y_i)$;
- vector of regressors $\mathbf{x}_i$ of length p;
- vector of parameters $\boldsymbol{\beta}$ of length p.

MC Logit Model as Multivariate GLM

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots)'$ be a vector response for subject i. We define the multivariate GLM as

$$\mathbf{g}(\boldsymbol{\mu}_i) = \mathbf{X}_i \boldsymbol{\beta},$$

where:

- $\boldsymbol{\mu}_i = E(\mathbf{Y}_i)$;
- $\mathbf{g}$ is a vector of link functions;

- $\mathbf{X}_i$ is the model matrix, such that its row h contains the explanatory variables corresponding to the response y_{ih}:

$$\mathbf{X}_i = \begin{bmatrix} 1 & \mathbf{x}_i' & 0 & 0 & \dots & 0 \\ 0 & 1 & \mathbf{x}_i' & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 & \mathbf{x}_i' \end{bmatrix};$$

- $\boldsymbol{\beta}' = (\alpha_1, \boldsymbol{\beta}_1, \dots, \alpha_{J-1}, \boldsymbol{\beta}_{J-1})$.

Alligator Food Choice Example

Example from "Categorical Data Analysis", Agresti, 2002.
Data: 219 alligators captured in four Florida lakes.

- Nominal response variable: primary food type (five categories: fish, invertebrate, reptile, bird, other)
- Covariates:
  - Lake of capture (L)
  - Gender (G)
  - Size (S)

Baseline-category logit models can investigate the effects of L, G, and S on primary food type.

Alligator Food Choice Example

TABLE 7.1 Primary Food Choice of Alligators

			Primary Food Choice				
Lake	Gender	Size (m)	Fish	Invertebrate	Reptile	Bird	Other
Hancock	Male	≤ 2.3	7	1	0	0	5
		> 2.3	4	0	0	1	2
	Female	≤ 2.3	16	3	2	2	3
		> 2.3	3	0	1	2	3
Oklawaha	Male	≤ 2.3	2	2	0	0	1
		> 2.3	13	7	6	0	0
	Female	≤ 2.3	3	9	1	0	2
		> 2.3	0	1	0	1	0
Trafford	Male	≤ 2.3	3	7	1	0	1
		> 2.3	8	6	6	3	5
	Female	≤ 2.3	2	4	1	1	4
		> 2.3	0	1	0	0	0
George	Male	≤ 2.3	13	10	0	2	2
		> 2.3	9	0	0	1	2
	Female	≤ 2.3	3	9	1	0	1
		> 2.3	8	1	0	0	1

Source: Data courtesy of Clint Moore, from an unpublished manuscript by M. F. Delaney and C. T. Moore.

G^2 vs χ^2 statistics

Used to determine whether there is a statistically significant difference between the expected frequencies E_i and the observed frequencies O_i in one or more categories of a contingency table.

$$G = 2\sum_i O_i \cdot \ln\left(\frac{O_i}{E_i}\right)$$

The total observed count should be equal to the total expected count:

$$\sum_i O_i = \sum_i E_i = N$$

G^2- statistic

- Likelihood-ratio or maximum likelihood statistical significance;
- use it when the sample size is large;
- more reliable for comparing models than for testing fit: the smaller is the value of the statistic, the closer the observed frequencies are to the expected frequencies.

Alligator Food Choice Example

TABLE 7.2 Goodness of Fit of Baseline-Category Logit Models for Table 7.1

Model[a]	G^2	X^2	df
( )	116.8	106.5	60
(G)	114.7	101.2	56
(S)	101.6	86.9	56
(L)	73.6	79.6	48
$(L+S)$	52.5	58.0	44
$(G+L+S)$	50.3	52.6	40
Collapsed over G			
( )	81.4	73.1	28
(S)	66.2	54.3	24
(L)	38.2	32.7	16
$(L+S)$	17.1	15.0	12

[a] G, gender; S, size; L, lake of capture. See the text for details.

Alligator Food Choice Example

TABLE 7.3 Observed and Fitted Values for Study of Alligator's Primary Food Choice

Lake	Size of alligator (meters)	Primary Food Choice: Fish	Invertebrate	Reptile	Bird	Other
Hancock	≤ 2.3	23 (20.9)	4 (3.6)	2 (1.9)	2 (2.7)	8 (9.9)
	> 2.3	7 (9.1)	0 (0.4)	1 (1.1)	3 (2.3)	5 (3.1)
Oklawaha	≤ 2.3	5 (5.2)	11 (12.0)	1 (1.5)	0 (0.2)	3 (1.1)
	> 2.3	13 (12.8)	8 (7.0)	6 (5.5)	1 (0.8)	0 (1.9)
Trafford	≤ 2.3	5 (4.4)	11 (12.4)	2 (2.1)	1 (0.9)	5 (4.2)
	> 2.3	89 (8.6)	7 (5.6)	6 (5.9)	3 (3.1)	5 (5.8)
George	≤ 2.3	16 (18.5)	19 (16.9)	1 (0.5)	2 (1.2)	3 (3.8)
	> 2.3	17 (14.5)	1 (3.1)	0 (0.5)	1 (1.8)	3 (2.2)

Alligator Food Choice Example

TABLE 7.4 Estimated Parameters in Logit Model for Alligator Food Choice, Based on Dummy Variable for First Size Category and Each Lake Except Lake George[a]

			Lake		
Logit[b]	Intercept	Size $\leq$ 2.3	Hancock	Oklawaha	Trafford
$\log(\pi_I/\pi_F)$	−1.55	1.46 (0.40)	−1.66 (0.61)	0.94 (0.47)	1.12 (0.49)
$\log(\pi_R/\pi_F)$	−3.31	−0.35 (0.58)	1.24 (1.19)	2.46 (1.12)	2.94 (1.12)
$\log(\pi_B/\pi_F)$	−2.09	−0.63 (0.64)	0.70 (0.78)	−0.65 (1.20)	1.09 (0.84)
$\log(\pi_O/\pi_F)$	−1.90	0.33 (0.45)	0.83 (0.56)	0.01 (0.78)	1.52 (0.62)

[a] SE values in parentheses.
I, invertebrate; *R*, reptile; *B*, bird; *O*, other; *F*, fish.

Effects of lake (L) and size (S) on the odds that alligators select other primary food types instead of fish.

Multinomial outcome logit models in R

- `library(nnet)`

- `multinom(formula, data, weights, subset, na.action, contrasts = NULL, Hess = FALSE, summ = 0, censored = FALSE, model = FALSE, ...)`