Estimation of multinomial logit model using the Begg & Gray approximation

Hana Ševčíková and Adrian Raftery

August 27, 2024

1 MNL Model Specification

Our observations correspond to N individuals each of whom makes one choice out of J alternatives. The dependent variable, Y_n, is the choice made by the n-th individual. The set of independent variables is divided into a set of variables that are individual-specific, say $X_n = (x_{n1}, \ldots, x_{nK_o})^T$, and a set of variables that are alternative-specific, say $W_{ni} = (w_{ni1}, \ldots, w_{niK_a})^T$, $i = 1, \ldots, J$. The probability of individual n choosing alternative i is given by the standard multinomial logit formula

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} \quad \text{where} \quad V_{ni} = \alpha_{i0} + \alpha_i X_n + \beta_i W_{ni} \tag{1}$$

Here, $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{iK_o})$ and $\beta_i = (\beta_{i1}, \ldots, \beta_{iK_a})$. We call $\alpha_{10}, \ldots, \alpha_{J0}$ the alternative-specific constants.
It is often of interest to constrain the coefficients to be the same over the given set of alternatives, i.e. $\alpha_{1k} = \alpha_{2k} \cdots = \alpha_{Jk}$ for a given k. The same applies to the β coefficients.

Base Alternative

In order to be able to use the Begg & Gray approximation [1], we need to set a base alternative and treat the remaining alternatives as differences to the base. Thus, if the base alternative is 1, $V_{n1} = 0$ for all n. Furthermore,

$$V_{ni} = \alpha_{i0} + \alpha_i X_n + \beta_i W'_{ni} \quad \text{where} \quad W'_{ni} = W_{ni} - W_{n1} \quad \text{for} \quad i = 2, \ldots J \tag{2}$$

2 Conversion

The conversion is done analogously to [2]. We decompose the original dataset into $D_0 = \{D_b, D_r\}$, where D_b denotes the set of individuals that chose the base alternative, and D_r denotes the set of the remaining individuals. For each

$i = 1, \ldots, J$, set N_i to be the number of individuals that chose alternative i. Then the converted dataset is constructed as follows:

1. Form J matrices $M_1, \ldots, M_J$ where each M_i has N_i rows and the columns consist of Y, X and $U = W'_{\cdot i}$.
2. Form $J-1$ blocks, $D_2, \ldots, D_J$, where D_i has $(N_1 + N_i)$ rows and is formed as follows:
   (a) Take the rows of M_1 and M_i.
   (b) Add columns:
      - $Y^* = \begin{cases} 0 & : \quad Y = 1 \\ 1 & : \quad \text{otherwise} \end{cases}$
      - $\{Z_2, \ldots, Z_J\}$, where $Z_j = \begin{cases} \mathbf{1} & : \quad j = i \\ \mathbf{0} & : \quad \text{otherwise} \end{cases}$
      - $\{Z_2 X, \ldots, Z_J X\}$
      - $\{Z_2 U, \ldots, Z_J U\}$
3. Combine the rows of $D_2, \ldots, D_J$.

The approximated binary logistic model is given by

$$\operatorname{logit}(P[Y^* = 1]) = \gamma_2 + \sum_{l=3}^{J} \gamma_l Z_l + \sum_{k=1}^{K_o} Q(\delta_{(\cdot)k}, X_k) + \sum_{k=1}^{K_a} Q(\theta_{(\cdot)k}, U_k) \qquad (3)$$

where

$$Q(\tau_{(\cdot)k}, S) = \begin{cases} \tau_k S & : \quad \text{if the coefficients of } S \text{ are contrained to be} \\ & \quad\;\; \text{the same for all alternatives, i.e. } \tau_{(\cdot)k} = \tau_k \\ \sum_{l=2}^{J} \tau_{lk} Z_l S & : \quad \text{otherwise} \end{cases}$$

Given estimated coefficients $\widehat{\gamma}$, $\widehat{\delta}$ and $\widehat{\theta}$, estimators of the coefficients of the original model in Equation (2) are given by:

$$\begin{aligned} &\widehat{\alpha}_{10} = 0, \;\; \widehat{\alpha}_{20} = \widehat{\gamma}_2, \;\; \widehat{\alpha}_{i0} = \widehat{\gamma}_i + \widehat{\gamma}_2, \quad \text{for } \; i = 3, \ldots, J \\ &\widehat{\alpha}_{ik} = \widehat{\delta}_{ik}, \quad \text{for } \; i = 2, \ldots, J, \;\; k = 1, \ldots, K_o \\ &\widehat{\beta}_{ik} = \widehat{\theta}_{ik}, \quad \text{for } \; i = 2, \ldots, J, \;\; k = 1, \ldots, K_a \end{aligned}$$

3 Example

Consider the following toy dataset with eight individuals, four alternatives and two independent variables, X and W:

id	Y	X	W_1	W_2	W_3	W_4
1	1	x_1	w_{11}	w_{12}	w_{13}	w_{14}
2	1	x_2	w_{21}	w_{22}	w_{23}	w_{24}
3	2	x_3	w_{31}	w_{32}	w_{33}	w_{34}
4	2	x_4	w_{41}	w_{42}	w_{43}	w_{44}
5	3	x_5	w_{51}	w_{52}	w_{53}	w_{54}
6	3	x_6	w_{61}	w_{62}	w_{63}	w_{64}
7	4	x_7	w_{71}	w_{72}	w_{73}	w_{74}
8	4	x_8	w_{81}	w_{82}	w_{83}	w_{84}

Setting the base alternative to 1, the converted dataset is of the form:

	Y	X	U	Y^*	Z_2	Z_3	Z_4	Z_2X	Z_3X	Z_4X	Z_2U	Z_3U	Z_4U
D_2	1	x_1	$w_{12}-w_{11}$	0	1	0	0	x_1	0	0	$w_{12}-w_{11}$	0	0
	1	x_2	$w_{22}-w_{21}$	0	1	0	0	x_2	0	0	$w_{22}-w_{21}$	0	0
	2	x_3	$w_{32}-w_{31}$	1	1	0	0	x_3	0	0	$w_{32}-w_{31}$	0	0
	2	x_4	$w_{42}-w_{41}$	1	1	0	0	x_4	0	0	$w_{42}-w_{41}$	0	0
D_3	1	x_1	$w_{13}-w_{11}$	0	0	1	0	0	x_1	0	0	$w_{13}-w_{11}$	0
	1	x_2	$w_{23}-w_{21}$	0	0	1	0	0	x_2	0	0	$w_{23}-w_{21}$	0
	3	x_5	$w_{53}-w_{51}$	1	0	1	0	0	x_5	0	0	$w_{53}-w_{51}$	0
	3	x_6	$w_{63}-w_{61}$	1	0	1	0	0	x_6	0	0	$w_{63}-w_{61}$	0
D_4	1	x_1	$w_{14}-w_{11}$	0	0	0	1	0	0	x_1	0	0	$w_{14}-w_{11}$
	1	x_2	$w_{24}-w_{21}$	0	0	0	1	0	0	x_2	0	0	$w_{24}-w_{21}$
	4	x_7	$w_{74}-w_{71}$	1	0	0	1	0	0	x_7	0	0	$w_{74}-w_{71}$
	4	x_8	$w_{84}-w_{81}$	1	0	0	1	0	0	x_8	0	0	$w_{84}-w_{81}$

An MNL model, specified in `mlogitBMA` by $Y \sim 1 \mid X + U$, is approximated using the logit model

$$Y^* \sim Z_3 + Z_4 + Z_2X + Z_3X + Z_4X + Z_2U + Z_3U + Z_4U$$

The MNL coefficients from Equation (2) correspond to:

$$\begin{aligned}
(\alpha_{20}, \alpha_2, \beta_2) &= (\text{Intercept},\ Z_2X\text{coef.},\ Z_2U\text{coef.}) \\
(\alpha_{30}, \alpha_3, \beta_3) &= (\text{Intercept} + Z_3\text{coef.},\ Z_3X\text{coef.},\ Z_3U\text{coef.}) \\
(\alpha_{40}, \alpha_4, \beta_4) &= (\text{Intercept} + Z_4\text{coef.},\ Z_4X\text{coef.},\ Z_4U\text{coef.})
\end{aligned}$$

If we constrain the coefficients to be the same for all alternatives, i.e. $\alpha = \alpha_2 = \alpha_3 = \alpha_4$ and $\beta = \beta_2 = \beta_3 = \beta_4$, which is specified in `mlogitBMA` by $Y \sim X + U$, the logit model

$$Y^* \sim Z_3 + Z_4 + X + U$$

is used as an approximation. In this case, α corresponds to the coefficient of X and β corresponds to the coefficient of U.

References

[1] Begg, C.B., Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika* **71**, 11–18.

[2] Yeung, K.Y., Bumgarner, R.E., Raftery, A.E. (2005) Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* **21** (10), 2394–2402.