C on C and X

Tihomir Asparouhov and Bengt Muthén

July 26, 2011

1 Introduction

In this note we describe a causal recursive system of logit models for latent categorical variables implemented in Mplus, that is used for latent transition modeling. Suppose that we have s latent categorical variables C_1, C_2, ... , C_s. Suppose that we have a set of covariates $X = (1, X_0)$ available in the model. We assume that the measurement model for C_i is defined by a set of equations describing the conditional distributions $[Y_i|C_i, X]$, where Y_i are observed continuous or categorical variables. We will focus on the relationship between the C_i. Since non-recursive systems are not allowed in our system of logit models we can assume a certain ordering for the C_i which defines the possible casual relationships. Any C_i can be influenced by all preceding C_1 , ..., C_{i-1} and can influence any of the following C_{i+1}, ..., C_s. The set of logit models that we consider is the following.

$$P(C_1 = i_1|X) = \frac{Exp(u(i_1)X)}{\sum_i Exp(u(i)X)}$$

$$P(C_2 = i_2|X, C_1 = i_1) = \frac{Exp(u(i_1, i_2)X)}{\sum_i Exp(u(i_1, i)X)}$$

$$P(C_3 = i_3|X, C_1 = i_1, C_2 = i_2) = \frac{Exp(u(i_1, i_2, i_3)X)}{\sum_i Exp(u(i_1, i_2, i)X)}$$

$$...$$

$$P(C_s = i_s|X, C_1 = i_1, C_2 = i_2, ..., C_{s-1} = i_{s-1}) = \frac{Exp(u(i_1, i_2, ..., i_s)X)}{\sum_i Exp(u(i_1, i_2, ..., i)X)}.$$

The parameters U in the above equation are not the parameters that we are interested in, rather we will estimate the parameters of a set of loglinear models that produce the above conditional logit models. For example the second equation can be viewed as the conditional logit model obtained from this loglinear model

$$P(C_1 = i_1, C_2 = i_2|X) = \frac{Exp((\mu + w_1(i1) + w_2(i2) + w_{12}(i_1, i_2))X)}{\sum_{i_1,i_2} Exp((\mu + w_1(i1) + w_2(i2) + w_{12}(i_1, i_2))X)}$$

with the restriction that $w(i1) = 0$ when i_1 is the last category of C_1, $w_2(i2) = 0$ when i_2 is the last category of C_2, and $w_{12}(i1, i2) = 0$ when i_1 is the last category of C_1 or when i_2 is the last category of C_2. The conditional logit model obtained from the above model is equivalent to the second conditional logit model when $u(i_1, i_2) = w_2(i2) + w_{12}(i_1, i_2)$.

Thus we estimate the following w parameters which in composition produce the u parameters. The exact relationship between these is given by

$$u(i_1) = w_1(i_1)$$

$$u(i_1, i_2) = w_2(i_2) + w_{12}(i_1, i_2)$$

$$u(i_1, i_2, i_3) = w_3(i_3) + w_{13}(i_1, i_3) + w_{23}(i_2, i_3) + w_{123}(i_1, i_2, i_3)$$

$$...$$

$$u(i_1, i_2, ..., is) = w_s(i_s) + \sum_{j=1}^{s-1} w_{js}(i_j, i_s) + \sum_{j_1=1, j_2=2, j_2>j_1}^{s-1} w_{j_1 j_2 s}(i_{j_1}, i_{j_2}, i_s) +$$
$$\sum_{j_1=1, j_2=2, j_3=3, j_3>j_2>j_1} w_{j_1 j_2 j_3 s}(i_{j_1}, i_{j_2}, i_{j_3}, i_s) + ... + w_{123...s}(i_1, i_2, ..., i_s)$$

Note however that the parameters w_1 are obtained from the log-linear model for C_1, parameters w_2 and w_{12} are obtained from the log-linear model for C_1 and C_2. This log-linear model has also a parameter of the type of w_1 however that is not the parameter that we use, and in general this parameter will be different, because in general log-linear tables are not collapsible (see [A]). Thus, if we are estimating a saturated model the results are going to depend on the assumed order of the C variables. For example, let's consider the following two orders of variables $C_1, C_2, C_3, ...$ and $C_3, C_2, C_1, ...$ The parameter w_{13} will be obtained from the loglinear model of C_1, C_2 and C_3 for both orders and it will be the same for both orders. However, parameter w_{12} for the first order will be obtained from the loglinear model of C_1 and C_2, and for the second order it will be obtained from the loglinear model of C_1, C_2 and C_3, and thus, will be different in general.

The total number of w parameters is $[(n_1+1)...(n_s+1)-1](1+q)$ where n_i are the number of different categories for C_i and q is the number of covariates. We use the following identification restrictions $w_{j_1 j_2 ... j_k}(i_{j_1}, i_{j_2}, ..., i_{j_k}) = 0$ when for some r, $i_{j_r} = n_r$, i.e., the C_r category is the last category. Under such a restriction the number of free parameters is $(n_1...n_s - 1)(1+q)$, which represent a fully saturated model. Similar model for observed variables is considered in [1], Chapter 7.

2 Estimation

We use an EM algorithm to obtain the ML estimates where the $C = (C_1, ..., C_s)$ variables represent the missing variables. The total number of C categories is $k = n_1...n_s$, which we denote by $i = (i_1, ..., i_s)$. The complete data log-likelihood is

$$\sum_j \sum_i 1_{C_j=i} log([Y_j|C_j, X_j]) + 1_{C_j=i} log([C_j = i|X_j])$$

where j varies across individuals in the sample and therefore the expected complete data log-likelihood is

$$\sum_j \sum_i p_{ji} log([Y_j|C_j, X_j]) + p_{ji} log([C_j = i|X_j])$$

where p_{ji} is the posterior probability

$$p_{ji} = \frac{[Y_j|C_j, X_j][C_j = i|X_j]}{\sum_i [Y_j|C_j, X_j][C_j = i|X_j]}$$

The maximization of the measurement part of the expected complete data loglikelihood is obtained as follows. We focus on the conditional logit models part:

$$S = \sum_j \sum_i p_{ji} log([C_j = i|X_j]).$$

Since

$$[C_j = i|X_j] = [C_{j1} = i_1|X_j][C_{j2} = i_2|X_j, C_{j1} = i_1]...$$
$$[C_{js} = i_s|X_j, C_{j1} = i_1, ..., C_{j(s-1)} = i_{s-1}]$$

we get that

$$S = \sum_{j,i_1} P_j(i_1)[C_{j1} = i_1|X_j] + \sum_{i_1} \sum_{j,i_2} P_j(i_1, i_2)[C_{j2} = i_2|X_j, C_{j1} = i_1] + ... +$$

$$\sum_{i_1,i_2,...,i_{s-1}} \sum_{j,i_s} P_j(i_1,i_2,...,i_s)[C_{js}=i_s|X_j,C_{j1}=i_1,...,C_{j(s-1)}=i_{s-1}]$$

where $P_j(i_1,i_2,...,i_r)$ is the marginal posterior distribution for $C_1,...,C_r$, i.e.,

$$P_j(i_1,i_2,...,i_r) = \sum_{i=(i_1,i_2,...,i_r,*,...,*)} p_{ji}$$

We will maximize S with respect to the w parameters by Quasi-Newton or Newton-Ralphson optimization algorithm. For this purpose we need $\partial S/\partial w$ and $\partial^2 S/(\partial w)^2$. The first step to get these derivatives will be to get the derivatives with respect to the u parameters. Note that $\partial u/\partial w = D$ is a constant matrix and therefore

$$\frac{\partial S}{\partial w} = \frac{\partial S}{\partial u} D$$

and

$$\frac{\partial^2 S}{(\partial w)^2} = D^T \frac{\partial^2 S}{(\partial u)^2} D.$$

Thus we only need the derivatives with respect to the u parameters. Notice however that the above formula of S is simply a sum of weighted conditional multinomial logit models, i.e., S is simply a sum of terms of the following type

$$\sum_{j,i_r} P_j(i_1,i_2,...,i_r)[C_{jr}=i_r|X_j,C_{j1}=i_1,...,C_{j(r-1)}=i_{r-1}]$$

each one of which is the weighted loglikelihood of a multinomial logit model with respect to the u parameters and its derivatives are easy to compute and are well known. Given the first and the second derivatives of S the EM algorithm and the asymptotic covariance of the ML estimates are obtained as in [MSS].

3 Refrences

[A] Agresti, A.1996. An Introduction to Categorical Data Analysis. John Wiley & Sons, Inc. New York, New York, USA