GENERALIZED LINEAR MIXED MODELS FOR NOMINAL DATA

Anders Skrondal
Norwegian Institute of Public Health

Sophia Rabe-Hesketh
King's College London

KEY WORDS: Multilevel models, generalized linear latent and mixed models, factor models, random coefficients, latent variables, polytomous data, rankings, discrete choice, permutations, nominal data, gllamm.

Introduction

Nominal variables include unordered polytomous variables and permutations. An unordered polytomous response is one among a set of categories whereas a permutation is an ordering of categories. The categories are nominal in the sense that they do not possess an inherent ordering shared by all units as is assumed for ordinal variables. Using decision terminology, we will refer to the categories as *alternatives*, unordered polytomous responses as *discrete choices* and permutations as *rankings*.

In this paper we analyze data from the British Election Panel Study 1987-1992. The central outcome is the voter's choice of party (Labour, Conservatives or Liberals). We will also utilize additional information enabling us to derive rankings of the parties.

The standard statistical model for discrete choices and rankings is logistic regression. In biostatistics it is often not recognized that these models can be derived from a random utility perspective. In econometrics and psychometrics this perspective naturally leads to the inclusion of alternative-specific covariates such as the distances between a voter and the parties on the left-right political continuum in addition to covariates not varying over alternatives such as the voter's age. Furthermore, it becomes clear that the alternative sets can vary between units as necessary if some parties do not have candidates in some constituencies, and how to handle rankings with tied alternatives.

However, the standard model must be extended for hierarchical or multilevel data like the election data where voting occasions (level 1) are nested in voters (level 2) who are nested in constituencies (level 3). We therefore include latent variables varying at different levels to represent unobserved heterogeneity and induce dependence among lower level units in the same higher level unit. The latent variables are of two types: random effects (intercepts and coefficients) and common factors.

Dependence structures previously considered have generally been more restrictive, typically including latent variables varying only at a single level and not combining random effects and factors. Furthermore, models treated in the multilevel literature have neither accommodated alternative-specific explanatory variables, nor alternative sets varying across units nor factor structures (cf. Goldstein, 2003; Raudenbush & Bryk, 2002) in models for discrete choices. Contributions on rankings appear to be missing in the multilevel literature.

The models considered here are special cases of the generalized linear latent and mixed model (GLLAMM) framework (Rabe-Hesketh, Skrondal & Pickles, 2004; Skrondal & Rabe-Hesketh, 2004) and subsume a wide range of models proposed in the psychometric, econometric, (bio)statistical and marketing literatures. Maximum likelihood estimation, empirical Bayes prediction and simulation can be performed using the gllamm software[1]. As far as we are aware, gllamm is the only program implementing adaptive quadrature for truly multilevel models (Rabe-Hesketh, Skrondal & Pickles, 2002).

[1] Downloadable from
`www.iop.kcl.ac.uk/iop/departments/biocomp/programs/gllamm.html`

Logistic Regression

Let $\mathcal{A}_i = \left\{ a_i^1, .., a_i^{A_i} \right\}$ be the set of A_i *alternatives* available for unit i and V_i^a a linear predictor for unit i and alternative a.

Discrete Choice

Letting c_i be the choice, a logistic model is usually specified as

$$\Pr(c_i \mid \mathcal{A}_i) = \frac{\exp(V_i^{c_i})}{\sum_{s=1}^{A_i} \exp(V_i^s)}, \qquad (1)$$

where V_i^a is a linear predictor.

An alternative specification of the logistic regression model, based on random utility models, is often used in econometrics. Here, the model is usually derived by introducing random utilities U_i^a for each unit i and alternative a. The utilities are modelled as

$$U_i^a = V_i^a + \epsilon_i^a,$$

where ϵ_i^a is a random term, assumed to be independently distributed across both units and alternatives with a Gumbel distribution

$$g(\epsilon_i^a) \;=\; \exp\left\{-\epsilon_i^a - \exp(-\epsilon_i^a)\right\}.$$

The probability of a choice can be viewed as *utility maximization* and expressed in terms of binary utility comparisons where the utility of the chosen alternative $U_i^{c_i}$ is larger than the utilities of all other alternatives. Remarkably, the logistic regression model (1) arises if and only if the random utilities are Gumbel distributed (e.g. McFadden, 1973).

Ranking

Let r_i^s be the alternative given rank s and $\boldsymbol{R}_i \equiv (r_i^1, r_i^2, \ldots, r_i^{A_i})$ the ranking of unit i.

The probability of a ranking can be construed as *utility ordering* and expressed in terms of binary utility comparisons, where the utility of the alternative ranked first is larger than the utility of that ranked second, which is larger than that ranked third and so on. Under the Gumbel specification of random utility this leads to the closed form "exploded logit" specification

$$\Pr(\boldsymbol{R}_i \mid \mathcal{A}_i) \;=\; \prod_{s=1}^{A_i-1} \frac{\exp(V_i^{r_i^s})}{\sum_{t=s}^{A_i} \exp(V_i^{r_i^t})}. \qquad (2)$$

The model is often denoted the "exploded logit" (Chapman & Staelin, 1982) since the ranking probability is written as a product of choice probabilities for successively remaining alternatives. That such an explosion results was proven by Luce & Suppes (1965).

Importantly, an analogous explosion is not obtained under normally distributed utilities. The Gumbel model is not reversible in the sense that successive choices starting with the worst alternative would lead to a different ranking probability. It is also worth noting that the ranking probability given above has the same form as the partial likelihood contribution for a stratum in Cox regression with "surviving" alternatives corresponding to risk sets and choices corresponding to failures. Exploiting this duality, alternatives given the same rank can be handled in the same way as tied survival times in Cox regression.

The task of providing a complete ranking of a full set of alternatives can be simplified by either presenting different subsets of alternatives to different individuals (using incomplete experimental designs) or by requiring only a top-ranking where for instance the three most preferred alternatives are ranked. Both situations are easily accommodated.

Modeling Heterogeneity

We now discuss the form of the linear predictor, including covariates to model observed heterogeneity and latent variables to model unobserved heterogeneity. The latent variables can vary at each of the three levels: voting occasion i (level 1), voter j (level 2) and constituency k (level 3). Latent variables at level 1 induce cross-sectional dependence between utilities of different parties at the same occasion for a voter. Latent variables at level 2 induce longitudinal dependence between utilities for different voting occasions for the same voter. Finally, latent variables at level 3 induce dependence between utilities for different voters within the same constituency. Note that latent variables at a given level induce dependence at all lower levels. We consider two types of latent variables, random coefficients and common factors.

Observed heterogeneity: Fixed Effects

We let the fixed effects be structured as

$$V^a_{ijk} \;=\; \boldsymbol{x}^{a\prime}_{ijk}\boldsymbol{b} + \boldsymbol{x}'_{ijk}\boldsymbol{g}^a,$$

where $\boldsymbol{x}^a_{ijk}$ is a covariate vector which varies over alternatives and may also vary over units and/or clusters whereas the vector $\boldsymbol{x}_{ijk}$ (often including a one for the alternative-specific intercepts) varies over units and/or clusters but not alternatives. The corresponding fixed coefficient vectors are $\boldsymbol{b}$ and $\boldsymbol{g}^a$, respectively. Note that the effects $\boldsymbol{b}$ are assumed to be the same on all utilities, so that this part of the model simply represents linear relationships between the utilities and alternative (and possibly unit and/or cluster)-specific covariates. For some alternative and unit and/or cluster-specific variables the effect may differ between alternatives. Such effects can be accommodated by including interactions between these variables and dummy variables for the alternatives in $\boldsymbol{x}^a_{ijk}$. In the election example, the covariate x_{ijk} could be the age of voter j at voting occasion i in constituency k with different effects g^a on the utilities for different parties. For identification, the first alternative typically serves as 'base alternative' with $g^1 = 0$. The party-specific covariate x^a_{ijk} could be a measure of the distance on the left-right political dimension between the ath party and the jth voter at occasion i in constituency j.

Random Coefficients

Covariates not varying over alternatives

For covariates $\boldsymbol{z}_{ijk}$ not varying over alternatives, we can have alternative-specific random effects $\boldsymbol{\gamma}^{a(2)}_{jk}$ and $\boldsymbol{\gamma}^{a(3)}_{k}$ at levels 2 and 3, giving a linear predictor of the form

$$V^a_{ijk} \;=\; \boldsymbol{z}'_{ijk}(\boldsymbol{g}^a + \boldsymbol{\gamma}^{a(2)}_{jk} + \boldsymbol{\gamma}^{a(3)}_{k}) + \ldots.$$

Again we treat the first alternative as a base-alternative with $\boldsymbol{\gamma}^{1(\ell)} = \boldsymbol{0}$. The supervectors $\boldsymbol{\gamma}^{(\ell)}$) containing the random coefficient vectors $\boldsymbol{\gamma}^{2(\ell)}$ to $\boldsymbol{\gamma}^{A(\ell)}$ at a given level ℓ have multivariate normal distributions,

$$\boldsymbol{\gamma}^{(\ell)} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Psi}^{(\ell)}_{\gamma})$$

and are independent across levels. In the election example, $\gamma^{a(2)}_{1jk}$ would be a voter and party-specific random intercept if $z_{1ijk} = 1$. Daniels & Gatsonis (1997) considered such a model for two-level data.

Alternative-specific covariates

For alternative-specific covariates $\boldsymbol{z}^a_{ijk}$, often called attributes of the alternatives, we could have random coefficients $\boldsymbol{\beta}^{(1)}_{ijk}$, $\boldsymbol{\beta}^{(2)}_{jk}$ $\boldsymbol{\beta}^{(3)}_{k}$ at levels 1, 2 and 3 giving a linear predictor of the form

$$V^a_{ijk} \;=\; \boldsymbol{z}^{a\prime}_{ijk}(\boldsymbol{\beta} + \boldsymbol{\beta}^{(1)}_{ijk} + \boldsymbol{\beta}^{(2)}_{jk} + \boldsymbol{\beta}^{(3)}_{k}) + \ldots$$

where

$$\boldsymbol{\beta}^{(\ell)} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Psi}^{(\ell)}_{\beta})$$

are independent across levels.

The coefficient of a given attribute z^a_{ijk} can be interpreted as the importance attached to that attribute, referred to as 'taste' by Hausman and Wise (1978). Allowing the coefficient to vary randomly therefore allows for heterogeneity in tastes. Note that we can have a random coefficient of z^a_{ijk} at level 1 since z^a_{ijk} varies within each level-1 unit so that the concept of a level-1 specific slope of z^a_{ijk} makes sense. This is in contrast to covariates $\boldsymbol{z}_{ijk}$ not varying over alternatives.

Common Factors

We can introduce common factors at each of the levels as follows:

$$V^a_{ijk} \;=\; \lambda^{a(1)}\eta^{(1)}_{ijk} + \lambda^{a(2)}\eta^{(2)}_{jk} + \lambda^{a(3)}\eta^{(3)}_{k} + \ldots$$

where

$$\eta^{(\ell)} \sim \mathrm{N}(0, \psi^{(\ell)}_{\eta})$$

$\eta^{(\ell)}$ is a factor at level ℓ, $\lambda^{a(\ell)}$ are the corresponding factor loadings and ϵ^a_{ijk} are unique factors (independent Gumbel as before) at the voting-occasion level. The factors are independent across levels. There are no unique factors at levels 2 and 3, but these could be represented by random intercepts $\gamma^{a(2)}_{1jk}$ and $\gamma^{a(3)}_{1k}$ with corresponding covariate z_{1ijk} set to 1.

There are two interpretations of random utility factor models. First, $\lambda^{a(\ell)}$ could be an alternative-specific effect of an unobserved variable $\eta^{(\ell)}$ at level ℓ. Second, $\lambda^{a(\ell)}$ could be an unobserved attribute of alternative a with random effect $\eta^{(\ell)}$. The model can easily be extended to multidimensional factors at each level. For single-level ranking data a probit factor model was discussed by Brady (1989).

Estimation and Prediction

All models considered in this paper can be estimated using the program gllamm (Rabe-Hesketh, Pickles & Skrondal, 2001). This program, written in Stata (StataCorp, 2003), implements maximum likelihood estimation and empirical Bayes prediction for many kinds of generalized linear mixed models with latent variables. Numerical integration by adaptive Gauss-Hermite quadrature (Rabe-Hesketh, Skrondal & Pickles, 2002) is used to integrate out the latent variables and obtain the marginal log-likelihood. This log-likelihood is maximized by Newton-Raphson using numerical first and second derivatives. Empirical Bayes predictions are posterior means of the latent variables given the observed responses with the parameter estimates plugged in. Both posterior means and standard deviations are obtained by numerical integration using adaptive quadrature.

Application

We now analyze data from the 1987-1992 panel of the British Election Study (Heath *et al.*, 1993). 1608 respondents participated in the panel. We excluded voting occasions with missing covariates and where the voters did not vote for candidates from the major parties. The resulting data comprised 2458 voting occasions, 1344 voters and 249 constituencies. The alternatives Conservative, Labour, and Liberal (Alliance) are labelled as $a = 1, 2, 3$, respectively. The voters were not explicitly asked to rank order the alternatives, but the first choice clearly corresponds to rank 1. The voters also rated the parties on a five point scale from "strongly against" to "strongly in favor". We used these ratings in assigning ranks to the remaining alternatives, ordering the parties in terms of their rating. Tied second and third choices were observed for 394 voting occasions yielding top-rankings.

The fixed part of all models considered includes the following election and/or voter specific covariates $\boldsymbol{x}_{ijk}$: [1987] and [1992] are dummy variables representing the elections, [Male] is a dummy for the voter being male, [Age] represents the age of the voters in 10 year units, [Manual] is a dummy for father of voter a manual worker and [Inflation] is a rating of perceived inflation since the last election on a five point scale. The fixed part also includes an election, voter and alternative specific covariate x^a_{ijk}; [LRdist]. This covariate represents the distance between a voter's position on the left-right political dimension and the mean position of the party voted for. The placements were constructed from four scales where respondents located themselves and each of the parties on a 11 point scale anchored by two contrasting statements (priority should be unemployment versus inflation, increase government services versus cut taxation, nationalization versus privatization, more effort to redistribute wealth versus less effort).

We consider three types of models for the random part: party-specific random intercepts $\gamma^{a(\ell)}$ at levels $\ell = 2, 3$ ($z_{ijk} = 1$), random slopes $\beta^{(\ell)}$ of political distance [LRdist] at levels $\ell = 1, 2, 3$ and common factors $\eta^{(\ell)}_{ijk}$ at levels $\ell = 1, 2, 3$. Skrondal and Rabe-Hesketh (2003) considered a number of combinations of these models. Models were estimated twice, comparing adaptive quadrature with 5 and 10 quadrature points per latent dimension to ensure reliable results. Here, we focus on their retained model based on rankings, which includes correlated alternative-specific random intercepts at the voter and constituency levels.

The estimates are given in Table 1. We see that the estimated effects of the election and/or voter specific covariates are in accordance with previous research on British elections. Being male and older increases the probability of voting Conservative, whereas a perceived high inflation since the last election harms the incumbent party (the Conservatives). The impact of social class is indicated by the higher probability of voting Labour among voters with a father who is/was a manual worker. Regarding our election, voter and *alternative* specific covariate [RLdist], the estimate also makes sense: the larger the political distance between voter and party, the less likely it is that the voter will vote for the party. The random intercept variances at the voter level are larger than at the constituency level consistent with a greater residual variability between voters within constituencies than between constituencies as would be expected. The variance of the random intercept for Labour, representing residual variability in the utility differences between Labour and Conservatives, is particularly large reflecting the presence

Table 1: Estimates for correlated alternative specific random intercepts model at voter and constituency levels

	Lab vs. Cons Est. (SE)	Lib vs. Cons Est. (SE)
FIXED PART:		
g_1^a [1987]	0.77 (0.56)	0.75 (0.37)
g_2^a [1992]	1.28 (0.59)	0.78 (0.39)
g_3^a [Male]	-0.99 (0.31)	-0.71 (0.20)
g_4^a [Age]	-0.74 (0.11)	-0.37 (0.07)
g_5^a [Manual]	1.57 (0.34)	0.10 (0.22)
g_6^a [Inflation]	1.31 (0.18)	0.74 (0.13)
b [LRdist]	-0.79 (0.04)	
RANDOM PART:		
Voter Level		
$\psi^{(2)}_{\gamma^a}$	16.13 (2.05)	6.03 (0.90)
$\psi^{(2)}_{\gamma^2,\gamma^3}$	8.53 (1.15)	
Constituency Level		
$\psi^{(3)}_{\gamma^a}$	4.91 (1.12)	0.60 (0.29)
$\psi^{(3)}_{\gamma^2,\gamma^3}$	1.21 (0.48)	
Log-likelihood	-2600.90	

of a mixture of people with strong residual (unexplained) support for the Labour or Conservative parties. There is a positive correlation between the random intercepts for the Labour and Liberal parties suggesting that those who prefer Labour to the Conservatives, after conditioning on the covariates, also tend to prefer the Liberal party to the Conservatives. This is consistent with the Liberal party being placed between the Labour and Conservative parties and suggests that the [LRdist] covariate has not fully captured this ordering.

Discussion

There appears to be unobserved heterogeneity at both voter and constituency levels, but not at the election level, in our application.

We have employed the gllamm software in the analyses reported in this article. An important merit of this software is the generality of the corresponding Generalized Linear Latent And Mixed (GLLAMM) model framework (e.g. Rabe-Hesketh, Skrondal & Pickles, 2004; Skrondal & Rabe-Hesketh, 2004), including the models considered here as special cases. This approach works well in a wide range of situations (Rabe-Hesketh, Skrondal & Pickles, 2002).

Although the framework for multilevel logistic regression presented here is general, it can be generalized further. For instance, latent variables can be regressed on covariates (e.g. Skrondal & Rabe-Hesketh, 2003) or higher order factors included to structure the covariance matrices of the latent variables. There could furthermore be covariates measured with error (e.g. Rabe-Hesketh, Pickles & Skrondal, 2003), several sets of discrete choices (e.g. Bock, 1972) or rankings, and pairwise comparisons (e.g. Böckenholt, 2001a). In fact, these extensions are all accommodated by the GLLAMM model framework and can be fitted in the gllamm software. We have not included finite mixtures for rankings like those proposed by Croon (1989) and Böckenholt (2001b) in this article. However, the Croon models are easily fitted in gllamm (Rabe-Hesketh, Pickles & Skrondal, 2001: Chapter 9). The gllamm program can also handle other response types including continuous, censored, ordinal, dichotomous, counts, discrete and continuous time durations and mixed responses.

References

Bock, R. D. 1972. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, **37**, 29–51.

Böckenholt, U. 2001a. Hierarchical modeling of paired comparison data. *Psychological Methods*, **6**, 49–66.

Böckenholt, U. 2001b. Mixed-effects analyses of rank-ordered data. *Psychometrika*, **66**, 45–62.

Brady, H. E. 1989. Factor and ideal point analysis for interpersonally incomparable data. *Psychometrika*, **54**, 181–202.

Chapman, R. G., & Staelin, R. 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, **14**, 288–301.

Croon, M. A. 1989. Latent class models for the analysis of rankings. *Pages 99–121 of:* De Soete, G., & Klauer, K. C. (eds), *New Developments in Psychological Choice Modeling.* Amsterdam: Elsevier.

Daniels, M. J., & Gatsonis, C. 1997. Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine*, **16**, 2311–2325.

Goldstein, H. 2003. *Multilevel Statistical Models.* London: Arnold.

Hausman, J. A., & Wise, D. A. 1978. A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, **46**, 403–426.

Heath, A., & et al. 1993. *British Election Panel Study, 1987-1992 [Computer file] SN: 2983.* Colchester, Essex: The Data Archive [Distributor].

Luce, R. D., & Suppes, P. 1965. Preference, Utility and Subjective Probability. *In:* Luce, R., Bush, R., & Galanter, E. (eds), *Handbook of Mathematical Psychology III.* New York: Wiley.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior. *Pages 105–142 of:* Zarembka, P. (ed), *Frontiers in Econometrics.* New York: Academic Press.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. 2001. *GLLAMM Manual.* Tech. rept. 2001/01. Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, **2**, 1–21.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. 2003. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling*, in press.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2004. Generalized multilevel structural equation modeling. *Psychometrika*, in press.

Raudenbush, S. W., & Bryk, A. S. 2002. *Hierarchical Linear Models.* Thousand Oaks: Sage.

Skrondal, A., & Rabe-Hesketh, S. 2003. Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, **68**, 267–287.

Skrondal, A., & Rabe-Hesketh, S. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/ CRC.

StataCorp. 2003 *Stata Statistical Software: Release 8.0.* College Station, TX: Stata Corporation.