Multilevel Logistic Regression for Polytomous Data and Rankings

ANDERS SKRONDAL
Division of Epidemiology
Norwegian Institute of Public Health

Joint work with
SOPHIA RABE-HESKETH

EFRON-SEMINAR
September 3, 2002

Outline

1. Introduction to Application: British Election Panel
2. Logistic Models as Random Utility Models
3. Independence from Irrelevant Alternatives (IIA)
4. Logistic Models with Observed Heterogeneity
5. Logistic Models with Shared Unobserved Heterogeneity
6. Multilevel Logistic Regression
7. Applying Multilevel Logistic Regression: British Election Panel
8. Generalized Linear Latent and Mixed Models (GLLAMM): Framework, Estimation and Prediction

British Election Panel: Outcome

- Alternatives: Conservative, Labour, Liberal (exclude minor parties)
- Polytomous / First choice: Party voted for
- Rankings: Parties not explicitly ranked – rankings derived
  A. Party voted for given rank 1
  B. Parties not voted for ranked into second and third place
  using rating scales:

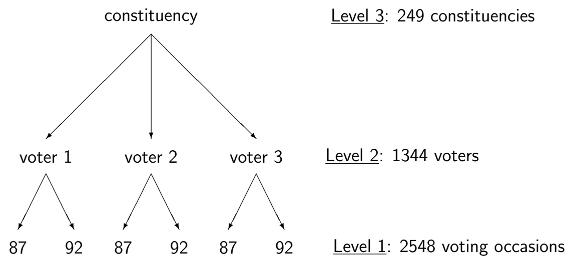
$$\begin{array}{ccc} 1 & & 5 \\ \text{strongly against} & \longrightarrow & \text{strongly in favour} \end{array}$$

British Election Panel: Covariates

- Election-specific covariates:
  - **[1987]**
  - **[1992]**
- Voter-specific covariates:
  - **[Male]**
  - **[Age]**
    Age in 10 year units
  - **[Manual]**
    Father a manual worker
- Election and voter-specific covariate:
  - **[Inflation]**
    Since last election, prices fallen a lot → increased a lot
- Election, voter and alternative-specific covariate:
  - **[LRdist]**
    Absolute distance between voters's position on left-right political continuum and party's mean position

British Election Panel: Three-Level Data

1608 people participated in survey of voting in 1987 and 1992 elections. Excluded voting occasions with voting on minor parties and missing covariates



Requirements of Methodology

Methodology should handle:

- Rankings as well as first choices
  (rankings beneficial for efficiency & identification)
- Multilevel data
  (dependence induced at several levels)
- Different types of covariates
  (including alternative specific)
- Varying alternative sets
- Ties
- Responses 'Missing at Random' (MAR)

and be

- Implemented in publicly available software

Random Utility Models

- Utility formulation useful:
  - Insight into logistic regression models
    (e.g. specification, identification)
  - Facilitates extension of conventional logistic regression for first choice and rankings to MULTILEVEL designs
- Unobserved 'utility' U_i^a associated with each alternative $a=1,...,A$ for unit $i=1,...,N$
- Random utility models composed as

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

  - V_i^a is fixed linear predictor representing observed heterogeneity
  - ϵ_i^a is random term representing unobserved heterogeneity
    (independent over i and a)

First Choice as Utility Maximization

- Alternative f is chosen if

$$U_i^f > U_i^g \text{ for all } g \neq f$$

- ϵ_i^a independent (over i and a) Gumbel or extreme value distributed of type I:

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

- McFadden (1973), Yellott (1977):

ϵ_i^a independent Gumbel

$$\Updownarrow$$

$$\Pr(f_i) \;=\; \frac{\exp(V_i^f)}{\sum_{a=1}^{A}\exp(V_i^a)}$$

[Conventional multinomial logit]

Ranking as Utility Ordering

- r_i^s is alternative with rank s for unit i.
  Ranking defined as $R_i = (r_i^1, r_i^2, \cdots, r_i^A)$, e.g. $(2,1,3)$
- R_i is obtained if
  $$U_i^{r_i^1} > U_i^{r_i^2} > \cdots > U_i^{r_i^A}$$
- Luce & Suppes (1965); Beggs, Cardell & Hausman (1981):

$$\epsilon_i^a \text{ independent Gumbel}$$

$$\Downarrow$$

$$\Pr(R_i) \;=\; \frac{\exp(V_i^{r_i^1})}{\sum_{s=1}^{A}\exp(V_i^{r_i^s})} \times \frac{\exp(V_i^{r_i^2})}{\sum_{s=2}^{A}\exp(V_i^{r_i^s})} \times \cdots \times \frac{\exp(V_i^{r_i^A})}{\sum_{s=A-1}^{A}\exp(V_i^{r_i^s})}$$

[Exploded logit]

- At each 'stage', a first choice is made among the remaining alternatives.
- Duality with partial likelihood contribution of stratum in Cox regression ('surviving' alternatives as risk sets and choices as failures)
  $\Longrightarrow$ Survival software applicable
- No 'explosion' for normally distributed utilities!

Identification

- Probability of choosing alternative 1 among alternatives 1, 2 and 3 can be expressed in terms of utility differences
  $\Pr(U^1 - U^2 > 0 \,\cap\, U^1 - U^3 > 0)$
- Probability of ranking alternatives as 1 2 3 similarly becomes
  $\Pr(U^1 - U^2 > 0 \,\cap\, U^2 - U^3 > 0)$
- First choice and ranking probabilities (and likelihoods) depend only on utility differences
  - Location of V_i^a is arbitrary
  - Scale of V_i^a not arbitrary since variance of ϵ_i^a fixed (at $\pi^2/6$),

$$\frac{\exp(V_i^1)}{\sum_a \exp(V_i^a)} = \frac{\exp(V_i^1 + c_i)}{\sum_a \exp(V_i^a + c_i)} \neq \frac{\exp(s_i \times V_i^1)}{\sum_a \exp(s_i \times V_i^a)}$$

Independence from Irrelevant Alternatives (IIA)

- Multinomial logit: Odds of alternative a versus b becomes
  $$\frac{\Pr(a)}{\Pr(b)} \;=\; \exp(V_i^a - V_i^b)$$
- Odds independent of properties of other alternatives
- Luce (1959) calls this 'Independence from Irrelevant Alternatives'

The Party Merger Problem

1. Initially three political parties: Lab1, Lab2 and Cons. Lab parties are indistinguishable and have the same linear predictor V_i^{Lab}, whereas Cons party has linear predictor V_i^{Cons}
   $$\Pr(\text{Lab1 or Lab2} \mid \text{Cons, Lab1, Lab2}) \;=\; \frac{2\exp(V_i^{\text{Lab}})}{2\exp(V_i^{\text{Lab}}) + \exp(V_i^{\text{Cons}})}$$
2. Lab1 and Lab2 merge to form a single Lab party
   $$\Pr(\text{Lab} \mid \text{Cons, Lab}) \;=\; \frac{\exp(V_i^{\text{Lab}})}{\exp(V_i^{\text{Lab}}) + \exp(V_i^{\text{Cons}})}$$
3. Follows that
   $$\Pr(\text{Lab} \mid \text{Cons, Lab}) \;<\; \Pr(\text{Lab1 or Lab2} \mid \text{Cons, Lab1, Lab2})$$
   Merger reduces the probability of voting Lab and increases the probability of voting Cons which is contraintuitive!
   Would expect no change in probability of voting Lab.

Heterogeneity and IIA

Numerical Examples of Probability Voting Lab

	Marginal Probability	
	Before merger	After merger
Heterogeneity	(Lab1 or Lab2)	(Lab)
none		
$V_i^{\text{Lab}} - V_i^{\text{Cons}} = 0$	0.67	0.50
observed		
men: $V_i^{\text{Lab}} - V_i^{\text{Cons}} = -1.2$		
women: $V_i^{\text{Lab}} - V_i^{\text{Cons}} = 2.8$	0.67	0.59
observed & shared unobserved		
men: $V_i^{\text{Lab}} - V_i^{\text{Cons}} = -0.8 + 4\delta_i$		
women: $V_i^{\text{Lab}} - V_i^{\text{Cons}} = 3.2 + 4\delta_i$		
$\delta_i \sim N(0,1)$	0.67	0.63

Observed Heterogeneity

Linear predictor for unit i and alternative a:

$$V_i^a \;=\; m^a + \mathbf{g}^{a\prime}\mathbf{x}_i + \mathbf{b}'\mathbf{x}_i^a$$

- Covariates and Parameters:
  - m^a alternative specific constants
  - $\mathbf{x}_i$ varies over units (but not alternatives) and has fixed effects $\mathbf{g}^{a\prime}$ varying over alternatives
    Examples: [Age] and [Male] for voter
  - $\mathbf{x}_i^a$ varies over alternatives (and possibly units) and has fixed effects $\mathbf{b}'$ not varying over alternatives
    Example: [LRdist] between different parties and voter
- Identification: Alternative 1 reference alternative,
  set $m^1 = 0$ and $g_k^1 = 0$ for all k.
- Common special cases:

$$V_i^a \;=\; m^a + \mathbf{g}^{a\prime}\mathbf{x}_i \quad \text{[statistics/biostatistics]}$$
$$V_i^a \;=\; m^a + \mathbf{b}'\mathbf{x}_i^a \quad \text{[econometrics/psychometrics]}$$

Shared Unobserved Heterogeneity

- Unit-specific unobserved heterogeneity <u>shared</u> between alternatives

  ⇓

  Utilities for alternatives dependent within units
- Use latent variables $\delta_i^{a(1)}$ to obtain flexible yet parsimonious covariance structure for utilities:
  I. Random Coefficient Models
  II. Factor Models

Covariance Structure I: Random Coefficient Models

- For alternative-specific covariates, $\mathbf{z}_i^a$, we consider *random coefficients* $\boldsymbol{\beta}_i$ representing unit-specific effects of the covariates

$$U_i^a \;=\; V_i^a + \delta_i^{a(1)} + \epsilon_i^a$$
$$\delta_i^{a(1)} = \boldsymbol{\beta}_i'\mathbf{z}_i^a$$

where

$$\boldsymbol{\beta}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}_\beta)$$

- Example: β_{li} is the voter-specific effect of political distance when $z_{li}^a =$ [LRdist]

Covariance Structure II: Factor Models

- One-factor model:

$$U_i^a \;=\; V_i^a + \delta_i^{a(1)} + \epsilon_i^a$$
$$\delta_i^{a(1)} = \lambda^a \eta_i$$

where

$$\eta_i \sim \mathsf{N}(0, \psi_\eta)$$

η_i is a *common factor*, λ^a are *factor loadings* and ϵ_i^a are *unique factors* (independent Gumbel as before)

- Two interpretations of factor models:
  1. λ^a alternative-specific effect of unobserved unit-specific variable η_i
  2. λ^a an unobserved attribute of alternative a and η_i random effect
- Identification: Likelihood depends only on utility differences,

$$V_i^a - V_i^b + (\lambda^a - \lambda^b)\eta_i + \epsilon_i^a - \epsilon_i^b$$

one loading must be fixed, e.g. $\lambda^1 = 0$, and scale of factor must also be fixed, e.g. $\lambda^2 = 1$

- Fragile identification for first choices unless alternative-specific covariates included
- Can be extended to multidimensional factors

Multilevel Designs and Latent Variables

- Three-level application:
  - Constituencies (level 3) indexed k
  - Voters (level 2) indexed j
  - Elections (level 1) indexed i
- Latent variables introduced at each level to represent unobserved heterogeneity at that level (induces dependence at all lower levels):
  - Latent variables at election level
    $\Longrightarrow$ Cross-sectional dependence between utilities within voter j at given election i
  - Latent variables at voter level
    $\Longrightarrow$ Longitudinal dependence between utilities within voter j over elections i
  - Latent variables at constituency level
    $\Longrightarrow$ Dependence between utilities between voters j within constituency k

Multilevel Logistic Regression

- The general three-level model

$$U_{ijk}^a \;=\; V_{ijk}^a + \delta_{ijk}^{(1)} + \delta_{ijk}^{(2)} + \delta_{ijk}^{(3)} + \epsilon_{ijk}^a$$

- Election level latent variables $\delta_{ijk}^{(1)}$ are composed as

$$\delta_{ijk}^{(1)} \;=\; \boldsymbol{\beta}_{ijk}^{(1)\prime}\mathbf{z}_{ijk}^{a(1)} + \boldsymbol{\lambda}^{a(1)\prime}\boldsymbol{\eta}_{ijk}^{(1)}$$

- Voter level latent variables $\delta_{ijk}^{(2)}$ are composed as

$$\delta_{ijk}^{(2)} \;=\; \boldsymbol{\beta}_{jk}^{(2)\prime}\mathbf{z}_{ijk}^{a(2)} + \boldsymbol{\gamma}_{jk}^{a(2)\prime}\mathbf{z}_{ijk} + \boldsymbol{\lambda}^{a(2)\prime}\boldsymbol{\eta}_{jk}^{(2)}$$

  - Random Coefficients I: $\boldsymbol{\beta}_{jk}^{(2)}$ are voter level random coefficients for *alternative*-specific covariates $\mathbf{z}_{ijk}^{a(2)}$
    (EX: effect of [LRdist] on party preference varies between voters)
  - Random Coefficients II: $\boldsymbol{\gamma}_{jk}^{a(2)}$ are voter level alternative specific random coefficients for *election*-specific covariates $\mathbf{z}_{ijk}$
    (EX: effects of [1987] and [1992] on party preference vary between voters)
  - Factors: $\boldsymbol{\lambda}^{a(2)\prime}\boldsymbol{\eta}_k^{(2)}$ induces dependence between different elections for a voter
- Constituency level latent variables $\delta_{ijk}^{(3)}$ are composed as

$$\delta_{ijk}^{(3)} \;=\; \boldsymbol{\beta}_k^{(3)\prime}\mathbf{z}_{ijk}^{a(3)} + \boldsymbol{\gamma}_k^{a(3)\prime}\mathbf{z}_{ijk} + \boldsymbol{\lambda}^{a(3)\prime}\boldsymbol{\eta}_k^{(3)}$$

British Election Panel: Retained Model and Estimates

- Latent variables at voter and constituency levels
- Correlated alternative specific random intercepts

	'Corr. Random Intercepts'		'Independence'	
	Lab vs. Cons	Lib vs. Cons	Lab vs. Cons	Lib vs. Cons
	Est. (SE)	Est. (SE)	Est. (SE)	Est. (SE)
FIXED PART:				
g_1^a [1987]	0.70 (0.51)	0.71 (0.35)	0.38 (0.20)	0.12 (0.17)
g_2^a [1992]	1.24 (0.53)	0.75 (0.37)	0.51 (0.20)	0.13 (0.18)
g_3^a [Male]	-0.92 (0.31)	-0.67 (0.20)	-0.79 (0.11)	-0.53 (0.09)
g_4^a [Age]	-0.74 (0.10)	-0.37 (0.04)	-0.37 (0.04)	-0.18 (0.03)
g_5^a [Manual]	1.63 (0.35)	0.12 (0.21)	0.65 (0.11)	-0.05 (0.10)
g_6^a [Inflation]	1.27 (0.18)	0.72 (0.13)	0.87 (0.09)	0.18 (0.03)
b [LRdist]	-0.78 (0.04)		-0.62 (0.02)	
RANDOM PART:				
Voter Level				
$\psi_{\gamma^a}^{(2)}$	15.85 (2.02)	5.73 (0.85)		
$\psi_{\gamma^2,\gamma^3}^{(2)}$	8.20 (1.09)			
Const. Level				
$\psi_{\gamma^a}^{(3)}$	5.15 (1.07)	0.76 (0.28)		
$\psi_{\gamma^2,\gamma^3}^{(3)}$	1.39 (0.47)			
log$L$	-2601.33		-2963.68	

The GLLAMM Framework

Generalized Linear Latent and Mixed Models (GLLAMM):

1. RESPONSE MODEL: Generalised linear model conditional on latent variables
   - Linear predictor:
     - observed covariates
     - multilevel latent variables (factors and/or random coefficients)
   - Links and distributions:
     as for GLM's plus ordinal and polytomous responses and rankings
2. STRUCTURAL MODEL: Equations for the latent variables
   - Regressions of latent variables on observed covariates
   - Regressions of latent variables on other latent variables (possibly at higher levels)
3. DISTRIBUTION OF LATENT VARIABLES (DISTURBANCES)
   - Multivariate normal
   - Discrete with unspecified distribution

`gllamm`: Stata program for estimation and prediction

- To obtain the likelihood of GLLAMM's, the latent variables must be integrated out
  - Sequentially integrate over latent variables, starting with the lowest level using a recursive algorithm
  - Use Gauss-Hermite quadrature to replace integrals by sums
  - Scale and translate quadrature locations to match the peak of the integrand using **adaptive quadrature**
- Maximum likelihood estimates obtained using Newton-Raphson
- Empirical Bayes (EB) predictions of latent variables and EB standard errors obtained using adaptive quadrature

Some links and references

- Skrondal, A. & Rabe-Hesketh, S. (2002). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, in press.
- GLLAMM framework:
  - Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002a). Generalized multilevel structural equation modelling. *Psychometrika*, in press.
  - Skrondal, A. & Rabe-Hesketh, S. (2003). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/ CRC.
- `gllamm` software:
  - `gllamm` and manual can be downloaded from `http://www.iop.kcl.ac.uk/iop/departments/biocomp/programs/gllamm.html`
  - Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002b). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal, 2*, 1–21.