I. Ordered categorical responses

- Small number of mutually exclusive ‘categories’, $y_i = 1, \ldots, S$
- Categories are ordered, so $y_i$ is ordinal
- Examples:
  - Severity of symptom (e.g. pain): none, moderate, severe
  - Frequency of behavior: never, occasionally, daily
  - Agreement (Likert scale): disagree strongly, disagree, agree, agree strongly
  - Educational attainment: high school, college degree, graduate or professional degree

Latent response models

- Latent (unobserved) continuous response $y_i^*$ underlies observed ordinal response $y_i$
- Threshold model determines observed response:
  $$y_i = \begin{cases} 
  1 & \text{if } y_i^* \leq \kappa_1 \\
  2 & \text{if } \kappa_1 < y_i^* \leq \kappa_2 \\
  \vdots & \vdots \\
  S & \text{if } \kappa_{S-1} < y_i^* 
  \end{cases}$$
- Latent response modeled as linear regression without intercept (for identification)
  $$y_i^* = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \epsilon_i$$
  $$= \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i$$
Latent response modeled as linear regression

Latent response modeled as linear regression

Cumulative versus individual response probabilities

Cumulative versus individual response probabilities

From latent response to generalized linear model formulations

From latent response to generalized linear model formulations

Logit link and odds ratios

Logit link and odds ratios
Two-level random intercept models

- Subjects \( i \) nested in clusters \( j \) (e.g. schools)
- Include random intercept \( \zeta_j \) for clusters

\[
y_{ij}^* = x_{ij}'\beta + \zeta_j + \epsilon_{ij}, \quad \zeta_j|x_{ij} \sim N(0, \psi)
\]

- \( \zeta_j \) independent of \( \epsilon_{ij} \)
- \( \zeta_j \) and \( \epsilon_{ij} \) independent across clusters
- \( \epsilon_{ij} \) independent across units within clusters

Intraclass correlation of latent responses

- Total residual \( \xi_{ij} = \zeta_j + \epsilon_{ij} \) has variance

\[
\text{Var}(\xi_{ij}|x_{ij}) = \begin{cases} 
\psi + 1 & \text{for probit models} \\
\psi + \pi^2/3 & \text{for logit models} 
\end{cases}
\]

- Covariance between total residuals \( \xi_{ij} \) and \( \xi_{ij'} \) of two subjects in same cluster is \( \psi \) and intraclass correlation is

\[
\text{Cor}(\xi_{ij}, \xi_{ij'}|x_{ij}, x_{ij'}) = \begin{cases} 
\psi/(\psi + 1) & \text{for probit models} \\
\psi/(\psi + \pi^2/3) & \text{for logit models} 
\end{cases}
\]

Conditional versus marginal effects

- For probit model, **conditional** or cluster-specific probabilities are

\[
\Pr(y_{ij} > s|\zeta_j, x_{ij}) = \Phi(x_{ij}'\beta + \zeta_j - \kappa_s)
\]

- **Marginal** or population-averaged response probabilities are

\[
\Pr(y_{ij} > s|x_{ij}) = \Pr(y_{ij}^* > \kappa_s) = \Pr(x_{ij}'\beta + \xi_{ij} > \kappa_s) = \Pr(-\xi_{ij} \leq x_{ij}'\beta - \kappa_s) = \Pr \left( \frac{\xi_{ij}}{\sqrt{\psi + 1}} \leq \frac{x_{ij}'\beta - \kappa_s}{\sqrt{\psi + 1}} \right) = \Phi \left( \frac{x_{ij}'\beta - \kappa_s}{\sqrt{\psi + 1}} \right)
\]

- Therefore, marginal effects \( \beta/\sqrt{\psi + 1} \) attenuated compared with conditional effects \( \beta \)

Illustration:

Conditional versus marginal relationship

![Illustration of conditional versus marginal relationship](image-url)

---

cluster-specific (random sample)
- median
- marginal or population-averaged
Multilevel random coefficient models

- Consider clustered longitudinal data with occasions $i$ (level 1) nested in student $j$ (level 2) in schools $k$ (level 3)
- Example of three-level random coefficient model:
  \[ y_{ijk}^* = \left[ c_{(2)ij} + c_{(3)jk} \right] + [\beta_1 + \zeta_{(2)1j} + \zeta_{(3)1k}]x_{1ijk} + \beta_2 x_{2ijk} + \epsilon_{ijk} \]
- $\zeta_{(2)ij}$ and $\zeta_{(3)jk}$ are random intercepts at levels 2 and 3
- $\zeta_{(2)1j}$ and $\zeta_{(3)1k}$ are random coefficients of $x_{1ijk}$ at levels 2 and 3
- Random effects at the same level, e.g. $(\zeta_{(2)ij}, \zeta_{(1ji)})$, bivariate normal with zero means, independent across units $jk$
- General three-level random coefficient model
  \[ y_{ijk}^* = X'_{ijk}\beta + Z_{(2)ijk}\zeta_{(2)jk} + Z_{(3)ijk}\zeta_{(3)jk} + \epsilon_{ijk} \]

Estimation: Approximate methods

- Estimation of multilevel models with categorical responses difficult because likelihood does not generally have closed form
- Penalized Quasilikelihood (PQL)
  - Two versions: First and second order (PQL-1, PQL-2), the latter being better
    - PQL-1 in GLIMMIX in SAS, MLwiN and HLM
    - PQL-2 in MLwiN
  - Even PQL-2 produces biased estimates for small clusters and high intraclass correlations
- Sixth order Laplace in HLM
- H-likelihood in Genstat
- Methods do not provide a likelihood

Estimation: Maximum likelihood

- Numerical integration
  - Gauss-Hermite (ordinary) quadrature used in MIXOR/MIXNO (two-level only), aML, NLMIXED in SAS (two-level only) and gllamm in Stata
  - Adaptive quadrature superior to ‘ordinary quadrature’, particularly for large clusters and high intraclass correlations and available in NLMIXED in SAS (two-level only), gllamm in Stata, glme in S-PLUS
- Monte Carlo integration
  - Simulated maximum likelihood in nlogit
  - Monte Carlo EM - no software?
- MCMC with vague priors approximates maximum likelihood and available in MLwiN and WinBUGS

Example: Cluster randomized trial of sex education in Norway

- Schools randomized to receive special sex education or not
- Assessments before, 6 and 18 months after randomization
- One outcome is question relating to ‘Contraceptive self-efficacy’:
  - “If my partner and I were about to have intercourse without either of us having mentioned contraception, it would be easy for me to produce a condom (if I brought one)”
  - Question answered in terms of five ordinal categories: not at all true of me, slightly true of me, somewhat true of me, mostly true of me, completely true of me
- Multilevel data with responses at occasions (level 1) from 1184 students (level 2) in 46 schools (level 3)
- Only 570 students always responded, 400 responded on some occasions and 114 never responded
Models and estimation using gllamm

- Occasions $t$, students $j$, schools $k$
- Covariates
  - $x_{1t}$ [Time] (0, 1, 3)
  - $x_{2jk}$ [Treat] (yes=1, no=0)
  - $x_{3tjk}$ [Treat] $\times$ [Time]
- Model the probability of exceeding a category $s$, $s = 1, 2, 3, 4$

$$\logit[Pr(y_{tjk} > s | z_{jk}^{(2)}, z_k^{(3)}, x_{tjk})] = \beta_1 x_{1t} + \beta_2 x_{2jk} + \beta_3 x_{3tjk} + z_{jk}^{(2)} + z_k^{(3)} - \kappa_s$$

- Estimation using adaptive quadrature in gllamm:

```
gllamm use treat time treat_time, i(id school) ///
link(ologit) family(binom) adapt
```

Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Single-level model</th>
<th>Two-level model</th>
<th>Three-level model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est (SE)</td>
<td>Est (SE)</td>
<td>Est (SE)</td>
</tr>
<tr>
<td>$\beta_1$ [Time]</td>
<td>-0.12 (0.06)</td>
<td>-0.13 (0.06)</td>
<td>-0.13 (0.06)</td>
</tr>
<tr>
<td>$\beta_2$ [Treat]</td>
<td>-0.05 (0.14)</td>
<td>-0.02 (0.19)</td>
<td>-0.02 (0.19)</td>
</tr>
<tr>
<td>$\beta_3$ [Time] $\times$ [Treat]</td>
<td>0.17 (0.08)</td>
<td>0.17 (0.09)</td>
<td>0.17 (0.09)</td>
</tr>
<tr>
<td>Var($z_{jk}^{(2)}$)</td>
<td>-</td>
<td>2.03 (0.31)</td>
<td>2.03 (0.31)</td>
</tr>
<tr>
<td>Var($z_k^{(3)}$)</td>
<td>-</td>
<td>-</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>-3.54 (0.17)</td>
<td>-4.41 (0.23)</td>
<td>-4.41 (0.23)</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>-2.43 (0.13)</td>
<td>-3.15 (0.19)</td>
<td>-3.15 (0.19)</td>
</tr>
<tr>
<td>$\kappa_3$</td>
<td>-1.18 (0.12)</td>
<td>-1.58 (0.16)</td>
<td>-1.58 (0.16)</td>
</tr>
<tr>
<td>$\kappa_4$</td>
<td>0.16 (0.12)</td>
<td>0.25 (0.15)</td>
<td>0.25 (0.15)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-2531</td>
<td>-2471</td>
<td>-2471</td>
</tr>
</tbody>
</table>

Predicted probabilities

- Conditional probabilities:

$$Pr(y_{tjk} > 2 | \zeta_{jk}^{(2)} = 0, \zeta_k^{(3)} = 0, x_{tjk}) = \frac{\exp(x'_{tjk} \hat{\beta} + 0 + 0 - \hat{\kappa}_2)}{1 + \exp(x'_{tjk} \hat{\beta} + 0 + 0 - \hat{\kappa}_2)}$$

```
gen ul1 = 0
gen u2 = 0
gllapred p_cond, mu us(u) above(2)
```

- Marginal probabilities:

$$Pr(y_{tjk} > 2 | x_{tjk}) = \int \int \frac{\exp(x'_{tjk} \hat{\beta} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} - \hat{\kappa}_2)}{1 + \exp(x'_{tjk} \hat{\beta} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} - \hat{\kappa}_2)} \cdot g(\zeta_{jk}^{(2)} \cdot g(\zeta_k^{(3)} \cdot d\zeta_{jk}^{(2)} \cdot d\zeta_k^{(3)})$$

```
gllapred p_marg, mu marg above(2)
```

Predicted conditional probabilities

- Probability of responding at least ‘mostly true of me’ (category 3)
- Relationship between probability and occasion for two treatment groups, when $\zeta_k^{(3)} = 0$ and $\zeta_{jk}^{(2)}$ is -1, 0 and 1:
**Observed proportions and predicted marginal probabilities**

- - - intervention group, --- control group, ● predicted, ○ observed

**II. Unordered categorical responses or discrete choice**

- Small number of mutually exclusive categories \(a, a = 1, \ldots, A\).
- Categories cannot be ordered a priori
- Examples:
  - Candidate voted for: Obama, McCain, Nader, other
  - Brand of cola preferred: Coca Cola, Pepsi, other
  - Method of birth control used: pill, condom, etc.
  - Diagnosis: Autism, Asperger's syndrome, pervasive developmental disorder, other
- Responses often correspond to discrete choices among alternatives (categories)

**Random utility models**

- Unobserved 'utility' \(U^a_i\) associated with each alternative \(a = 1, \ldots, A\) for unit \(i = 1, \ldots, N\)
- Random utility models composed as
  \[
  U^a_i = V^a_i + \epsilon^a_i
  \]
  - \(V^a_i\) is linear predictor
  - \(\epsilon^a_i\) is residual term (independent over \(i\) and \(a\))
- Alternative \(f\) chosen if
  \[
  U^f_i > U^g_i \quad \text{for all} \quad g \neq f
  \]
  - \(\epsilon^a_i\) independent Gumbel distributed
  - \(\Downarrow\)
  - \(Pr(f_i) = \frac{\exp(V^f_i)}{\sum_{a=1}^{A} \exp(V^a_i)}\) multinomial logit

**Covariate effects on utilities**

Linear predictor for unit \(i\) and alternative \(a\):

\[
V^a_i = m^a + x_i^a g^a + x_i^a b
\]

- Covariates and parameters:
  - \(m^a\) alternative-specific constants
  - \(x_i^a\) varies over subjects (but not alternatives) and has fixed alternative-specific effects \(g^a\)
  - Example: Age of subject
  - \(x_i^a\) varies over alternatives (and possibly subjects) and has fixed effects \(b\), constant across alternatives
  - Example: Cost of treatment alternative
**Identification**

- Probability of choosing alternative 1 among alternatives 1, 2 and 3 can be expressed in terms of utility differences:
  \[ \Pr(U^1 - U^2 > 0 \text{ and } U^1 - U^3 > 0) \]
- Therefore location of \( V_{i1} \) arbitrary:
  \[ \frac{\exp(V_{i1})}{\sum_a \exp(V_{ia})} = \frac{\exp(V_{i1} + c_i)}{\sum_a \exp(V_{ia} + c_i)} \]
- Solution: Last alternative \( S \) serves as reference alternative, set \( m^S = 0 \) and \( g^S = 0 \)

**Abuse of antibiotics in China**

- Acute respiratory tract infection (ARI) can lead to pneumonia and death if not properly treated
- Inappropriate frequent use of antibiotics common in China in 1990’s, leading to drug resistance
- In the 1990’s WHO introduced program of case management for children under 5 with ARI in China
- Data collected on 855 children \( i \) (level 1) treated by 134 doctors \( j \) (level 2) in 36 hospitals \( k \) (level 3) in two counties (one of which was in WHO program)
- **Response variable**: Abuse defined as prescription of antibiotics when there were no clinical indications based on medical files
  1. Abuse of several antibiotics
  2. Abuse of one antibiotic
  3. Correct use of antibiotics (reference category)

**Multilevel random utility models**

- Consider three-level data with patients \( i \) (level 1) treated by doctors \( j \) (level 2) working in hospitals \( k \) (level 3)
- Can include random effects in linear predictor:
  \[ V_{ijk} = [m^a + \gamma_{(2)} + \gamma_{(3)}] + x_{ijk}[g^a + \gamma_{(2)} + \gamma_{(3)}] + x_{ijk}[b + \beta_{(2)} + \beta_{(3)}] \]
- **Random intercepts**: \( \gamma_{(2)} \) and \( \gamma_{(3)} \)
- **Random Coefficients I**: \( \gamma_{(2)} \) and \( \gamma_{(3)} \) are alternative-specific random coefficients for subject-specific covariates \( x_{ijk} \)
- **Random Coefficients II**: \( \beta_{(2)} \) and \( \beta_{(3)} \) are random coefficients for alternative-specific covariates \( x_{ijk} \)

**Covariates**

- 7 covariates \( x_{ijk} \)
  - Patient level \( i \)
    - [Age] Age in years (0-4)
    - [Temp] Body temperature, centered at 36°C
    - [Paymed] Pay for medication (yes=1, no=0)
    - [Selfmed] Self medication (yes=1, no=0)
    - [Wrdiag] Failure to diagnose ARI early (yes=1, no=0)
  - Doctor level \( j \)
    - [DRed] Doctor’s education
      (6 categories from self-taught to medical school)
  - Hospital level \( k \)
    - [WHO] Hospital in WHO program (yes=1, no=0)

Model

- No alternative-specific covariates

\[ V_{ijk}^{\alpha} = [m_{ij}^{\alpha} + \gamma_{0jk}^{(2)} + \gamma_{0k}^{(3)}] + X'_{ijk} \gamma^{\alpha} \]

<table>
<thead>
<tr>
<th>Data:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 0</td>
</tr>
<tr>
<td>1 1 2 1</td>
</tr>
</tbody>
</table>

- Estimation in `gllamm`

```
gen categ1 = alt == 1
gen categ2 = alt == 2
eq c1: categ1
eq c2: categ2
gllamm alt age temp ..., i(doc hosp) nrf(2 2) eqs(c1 c2 c1 c2) link(mlogit) expanded(child choice m) basecat(3) adapt
```

Maximum likelihood estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Abuse several</th>
<th>Abuse one</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>(SE)</td>
</tr>
<tr>
<td>0[Cons]</td>
<td>-5.72</td>
<td>(0.99)</td>
</tr>
<tr>
<td>0[Age]</td>
<td>0.07</td>
<td>(0.09)</td>
</tr>
<tr>
<td>0[Temp]</td>
<td>-0.27</td>
<td>(0.13)</td>
</tr>
<tr>
<td>0[Paymed]</td>
<td>0.92</td>
<td>(0.40)</td>
</tr>
<tr>
<td>0[Selfmed]</td>
<td>-0.86</td>
<td>(0.29)</td>
</tr>
<tr>
<td>0[Wred]</td>
<td>1.85</td>
<td>(0.26)</td>
</tr>
<tr>
<td>0[WHO]</td>
<td>-2.40</td>
<td>(0.62)</td>
</tr>
</tbody>
</table>

Doctor-level variances

| Var(\gamma_{0jk}^{(2)}) | 0.46 | (0.28) | 0.43 | (0.22) |
| Cov(\gamma_{0jk}^{(2)}, \gamma_{0jk}^{(2)}) | -0.44 | (0.13) |

Hospital-level variances

| Var(\gamma_{0k}^{(3)}) | 0.88 | (0.45) | 0.11 | (0.12) |
| Cov(\gamma_{0k}^{(3)}, \gamma_{0k}^{(3)}) | 0.31 | (0.20) |

References

- See also http://www.gllamm.org

Further reading

- Snijders & Bosker (1999): Excellent introduction to multilevel modelling (MLM)
- Rabe-Hesketh & Skrondal (2008): MLM using Stata
- Skrondal & Rabe-Hesketh (2004): Generalized latent variable modeling