Slide 1

# Estimation and prediction in generalized linear latent and mixed models: A practical perspective

Anders Skrondal
Division of Epidemiology
Norwegian Institute of Public Health

joint work with
Sophia Rabe-Hesketh (King's College London)
and Andrew Pickles (The University of Manchester)

Todagesmøde i Dansk Selskab for Teoretisk Statistik
Aarhus, 19.-20. November 2002

Slide 2

## OUTLINE

**I.** A general model framework for latent variable modelling: Generalised Linear Latent And Mixed Models (GLLAMM)

**II.** Estimation methods: When do they work and fail?

- Methods assuming normal latent variables
  - Marginal and Penalized Quasi-Likelihood
  - Maximum Likelihood
    - Gaussian quadrature
    - Adaptive quadrature
- Non-Parametric Maximum Likelihood

**III.** Prediction methods for latent variables: When do they work and fail?

- Empirical Bayes assuming normal latent variables
- Empirical Bayes based on non-parametric maximum likelihood

**IV.** Conclusion

Slide 3

# I. GLLAMM

- Response model: Generalised linear model conditional on latent variables
  - Linear predictor: latent variables as factors or random coefficients
  - Links and distributions
- Structural model
  - Regressions of latent variables on observed variables
  - Regressions of latent variables on other latent variables
- Distribution of the latent variables (disturbances)
  - Multivariate normal
  - Discrete

Slide 4

## GLLAMM: Links and families

- The conditional expectation of the response is 'linked' to the linear predictor

$$g(E[y|\mathbf{x}, \boldsymbol{\eta}, \mathbf{z}]) = \nu$$

- The conditional distribution of the response is from the exponential family:

| Links |
|---|
| identity |
| reciprocal |
| logarithm |
| logit |
| probit |
| scaled probit |
| compl. log-log |

| Families |
|---|
| Gaussian |
| gamma |
| Poisson |
| binomial |

| Ordinal responses |
|---|
| ordinal logit |
| ordinal probit |
| ordinal compl. log-log |
| scaled ord. probit |

| Nominal & Rankings |
|---|
| multinomial logit |

- Heteroscedasticity: Dispersion or scale parameters can be modelled as $\log \sigma = \mathbf{z}^{(1)\prime}\boldsymbol{\alpha}$

Slide 5

## GLLAMM: Linear Predictor

$$\nu = \mathbf{x}'\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_m^{(l)\prime} \boldsymbol{\lambda}_m^{(l)} \quad \text{for identification, } \lambda_{m1}^{(l)} = 1$$

- Fixed part: $\mathbf{x}'\boldsymbol{\beta}$ as usual
- Random part:
  - $\eta_m^{(l)}$ is $m$th latent variable at level $l$, $m = 1, \cdots, M_l$, $l = 2, \cdots, L$
  - $\eta_m^{(l)}$ can be a **factor** or a **random coefficient**
  - $\mathbf{z}_m^{(l)}$ are variables and $\boldsymbol{\lambda}_m^{(l)}$ are parameters

Slide 6

## Random coefficient models in GLLAMM

- One covariate multiplies each latent variable,

$$\eta_m^{(l)} z_m^{(l)} \quad (\lambda_m^{(l)} = 1)$$

- e.g. Latent growth curve model for individuals $j$ (level 2) observed at times $t_{ij}$, $i = 1, \cdots, n_j$ (level 1)

$$\nu_{ij} = \beta_1 + \beta_2 t_{ij} + \eta_{1j}^{(2)} + \eta_{2j}^{(2)} t_{ij}$$

$\beta_1$, $\beta_2$ : mean intercept and slope

$\eta_{1j}^{(2)}$, $\eta_{2j}^{(2)}$: random deviations of the subject-specific intercepts and slopes from their means

## Factor models in GLLAMM

Slide 7

- A linear combination of dummy variables for the items multiplies each latent variable,

$$\eta_m^{(l)} \mathbf{z}_m^{(l)\prime} \boldsymbol{\lambda}_m^{(l)}$$

- e.g. One-factor model for items $i$, $i = 1, \cdots, I$ (level 1) and subjects $j$ (level 2)

$$\begin{aligned} \nu_{ij} &= \beta_1 \delta_{1i} + \cdots + \beta_I \delta_{Ii} + \eta_j^{(2)} (\delta_{1i} + \lambda_2^{(2)} \delta_{2i} \cdots + \lambda_I^{(2)} \delta_{Ii}) \\ &= \beta_i + \eta_j^{(2)} \lambda_i^{(2)}, \end{aligned}$$

where

$$\delta_{pi} = \begin{cases} 1 & \text{if } p = i \\ 0 & \text{otherwise} \end{cases}$$

$\beta_i$: intercept for item $i$

$\eta_j^{(2)}$: common factor

$\lambda_i^{(2)}$: factor loading for item $i$, $\lambda_1^{(2)} = 1$

| unit $j$ | item $i$ | $\delta_{1i}$ | $\delta_{2i}$ | $\cdots$ | $\delta_{Ii}$ | $y_{ij}$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | $\cdots$ | 0 | $y_{11}$ |
| 1 | 2 | 0 | 1 | $\cdots$ | 0 | $y_{21}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| 1 | $I$ | 0 | 0 | $\cdots$ | 1 | $y_{I1}$ |

## GLLAMM: Structural model

Slide 8

- Regressions of latent variables on other latent and explanatory variables

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}$$

- $\boldsymbol{\eta} = (\eta_1^{(2)}, \eta_2^{(2)}, \cdots, \eta_{M_2}^{(2)}, \cdots, \eta_1^{(l)}, \cdots, \eta_{M_l}^{(l)}, \cdots, \eta_{M_L}^{(L)})'$ ($M$ elements)
  - factors
  - random coefficients
- $\mathbf{B}$ is an upper diagonal $M \times M$ matrix of regression coefficients
- $\boldsymbol{\Gamma}$ is an $M \times p$ matrix of regression coefficients
- $\mathbf{w}$ is a $p$ dimensional vector of explanatory variables
- $\boldsymbol{\zeta}$ is an $M$ dimensional vector of errors/disturbances (same level as corresponding elements in $\boldsymbol{\eta}$)

Slide 9

## GLLAMM: Latent variable distributions

- Multivariate normal disturbances $\boldsymbol{\zeta}^{(l)}$ with covariance matrices $\boldsymbol{\Psi}^{(l)}$, independent across levels
- Discrete
  - Mixture model
    - latent class model
    - latent profile model
  - Nonparametric estimator of a continuous, discrete or mixed distribution

Slide 10

# II. ESTIMATION METHODS FOR GLMM's

- Marginal Quasi Likelihood (MQL) and Penalized Quasi Likelihood (PQL)
- Maximum Likelihood (ML)
  - Gaussian Quadrature (GQ)
  - Adaptive Quadrature (AQ)
- Non-Parametric Maximum Likelihood (NPML)
- Markov Chain Monte Carlo (MCMC)
  [Will not cover MCMC in this talk!]

Slide 11

## For simplicity: Random intercept model

- Linear predictor:
  - Two-level random intercept model (unit $i$, cluster $j$)

$$\nu_{ij} \;=\; \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j^{(2)}, \quad \zeta_j^{(2)} \sim \mathrm{N}(0, \psi^{(2)})$$

  - Three-level model random intercept model (unit $i$, cluster $j$, supercluster $k$)

$$\nu_{ijk} \;=\; \mathbf{x}'_{ijk}\boldsymbol{\beta} + \zeta_{jk}^{(2)} + \zeta_k^{(3)}, \quad \zeta_{jk}^{(2)} \sim \mathrm{N}(0, \psi^{(2)}), \zeta_k^{(3)} \sim \mathrm{N}(0, \psi^{(3)})$$

- Latent response formulation for dichotomous observed responses:
  - Threshold model:

$$y^*_{ijk} \;=\; \nu_{ijk} + \epsilon_{ijk},$$

$$y_{ijk} = \begin{cases} 1 & \text{if } y^*_{ijk} > 0 \\ 0 & \text{if } y^*_{ijk} \le 0 \end{cases}$$

  - Type of model:
    - LOGIT random intercept model if $\epsilon_{ijk} \sim$ Logistic, var$(\epsilon_{ijk}) = \pi^2/3$
    - PROBIT random intercept model if $\epsilon_{ijk} \sim \mathrm{N}(0,1)$

Slide 12

## Intra-class correlation

- Intra-class correlations for latent responses
  - Two-level logit random intercept model

$$\rho = \mathrm{cor}(y^*_{ij}, y^*_{i'j}|\mathbf{x}_{ij}) = \frac{\psi^{(2)}}{\psi^{(2)} + \pi^2/3}$$

  - Three-level logit random intercept model

$$\rho_a \;=\; \mathrm{cor}(y^*_{ijk}, y^*_{i'jk}|\mathbf{x}_{ijk}) = \frac{\psi^{(2)} + \psi^{(3)}}{\psi^{(2)} + \psi^{(3)} + \pi^2/3}$$

$$\rho_b \;=\; \mathrm{cor}(y^*_{ijk}, y^*_{i'jk}|\mathbf{x}_{ijk}, \zeta_k^{(3)}) = \frac{\psi^{(2)}}{\psi^{(2)} + \pi^2/3}$$

$$\rho_c \;=\; \mathrm{cor}(y^*_{ijk}, y^*_{i'j'k}|\mathbf{x}_{ijk}) = \frac{\psi^{(3)}}{\psi^{(2)} + \psi^{(3)} + \pi^2/3}$$

- Designing simulations:
  Relation between $\psi^{(2)}$ and $\rho$ in two-level model

| | $\psi^{(2)}$ | 0.01 | 0.10 | 0.50 | 1 | 2 | 4 | 10 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| logit: | $\rho$ | 0.00 | 0.03 | 0.13 | 0.23 | 0.38 | 0.55 | 0.75 | 0.86 |
| probit: | $\rho$ | 0.01 | 0.09 | 0.33 | 0.50 | 0.67 | 0.80 | 0.91 | 0.95 |

Slide 13

## MQL and PQL: Approximate linear mixed models

- Conditional expectation of the response in generalized linear mixed models

$$\mu_{ij} \;=\; h(\nu_{ij}),$$

where $h(\cdot)$ is the inverse link function.

- Linearize by expanding $h(\nu_{ij})$ as a 1st order Taylor series around the 'current' value of the linear predictor $\tilde{\nu}_{ij} = \mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}}+\tilde{\zeta}_j^{(2)}$

$$\begin{aligned} y_{ij} &\approx h(\tilde{\nu}_{ij}) + \mathbf{x}'_{ij}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})h'(\tilde{\nu}_{ij}) \\ &+(\zeta_j^{(2)}-\tilde{\zeta}_j^{(2)})h'(\tilde{\nu}_{ij}) \\ &+\epsilon_{ij}, \end{aligned}$$

where $\epsilon_{ij}$ is a heteroscedastic error term with variance $\phi V(\tilde{\mu}_{ij})$ corresponding to the chosen distribution

(e.g. Schall, 1991; Goldstein, 1991; Wolfinger, 1993; Breslow and Clayton, 1993; McGilchrist, 1994)

- Reformulate as approximate linear mixed model

$$\underbrace{y_{ij}-\omega_{ij}}_{y^{\bullet}_{ij}} \;\approx\; \underbrace{\mathbf{x}'_{ij}h'(\tilde{\nu}_{ij})\boldsymbol{\beta}}_{\mathbf{x}^{\bullet\prime}_{ij}\boldsymbol{\beta}} + \underbrace{\zeta_j^{(2)}h'(\tilde{\nu}_{ij})}_{\zeta_j^{\bullet(2)}} + \epsilon_{ij},$$

with offset $\omega_{ij} = h(\tilde{\nu}_{ij}) - (\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \tilde{\zeta}_j^{(2)})h'(\tilde{\nu}_{ij})$.

- Estimate as if linear mixed model

Slide 14

## MQL and PQL cont'd

- Extend Taylor expansion to 2nd order for the random part

(e.g. Goldstein, 1995; Goldstein and Rasbash, 1996)

$$\begin{aligned} y_i &\approx h(\tilde{\nu}_{ij}) + \mathbf{x}'_{ij}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})h'(\tilde{\nu}_{ij}) \\ &+(\zeta_j^{(2)}-\tilde{\zeta}_j^{(2)})h'(\tilde{\nu}_{ij}) \\ &+\underbrace{\frac{1}{2}(\zeta_j^{(2)}-\tilde{\zeta}_j^{(2)})^2h''(\tilde{\nu}_{ij})} \\ &+\epsilon_{ij}. \end{aligned}$$

- Types of estimators:

| | 1st order | 2nd order |
|---|---|---|
| $\tilde{\zeta}_j^{(2)}=0$ | **MQL-1** | **MQL-2** |
| $\tilde{\zeta}_j^{(2)}\approx$EB | **PQL-1** | **PQL-2** |

- PQL better approximation than MQL and second order expansions of the random part improve on first order
- Software:
  - HLM (also higher order Laplace for two-level models)
  - MLwiN (also bias-correcting bootstrap)

Slide 15

## Performance of MQL & PQL

- Pros:
  - Computationally efficient
  - Perform well when posterior is close to normal:
    - Conditional distribution of responses given random effects close to normal
      - Poisson distribution with mean 7 or greater (e.g. McCullagh and Searle, 2001)
      - Proportions with large binomial denominators
    - Large cluster size
- Cons:
  - Downward bias of fixed effects, random effect variances and their standard errors for
    - Dichotomous data with small cluster sizes (e.g. Rodriguez and Goldman, 1995, 2001; Breslow and Lin, 1995; Lin and Breslow, 1995; Goldstein and Rasbash, 1996; Browne and Draper, 2002)
    - 'Large' variance components
  - No likelihood!
  - Difficult to assess performance for a given application

Slide 16

## Misbehaviour of MQL & PQL: Simulation

- 500 datasets simulated to have the same structure as data from the 1987 Guatemalan National Survey of Maternal and Child Health (Browne and Draper, 2002)
- Outcome: whether women received prenatal care or not
- Simulated 2449 births $i$ (level 1) by 1558 women $j$ (level 2) from 161 communities $k$ (level 3) from a three-level logit random intercept model with one covariate at each level:

$$\eta_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2jk} + \beta_3 x_{3k} + \zeta_{jk}^{(2)} + \zeta_k^{(3)}$$

| | | Estimate | | Coverage 95%CI | |
|---|---|---|---|---|---|
| | True | MQL-1 | PQL-2 | MQL-1 | PQL-2 |
| $\beta_0$ | **0.65** | 0.47 | 0.61 | 77 | 92 |
| $\beta_1$ | **1.0** | 0.74 | 0.95 | 69 | 96 |
| $\beta_2$ | **1.0** | 0.75 | 0.96 | 18 | 91 |
| $\beta_3$ | **1.0** | 0.73 | 0.94 | 70 | 90 |
| $\psi^{(2)}$ | **1.0** | 0.03 | 0.57 | 0 | 27 |
| $\psi^{(3)}$ | **1.0** | 0.55 | 0.89 | 2 | 78 |

$\rho_a = 0.38$, $\rho_b = 0.23$ and $\rho_c = 0.19$.

Slide 17

## Gaussian Quadrature (GQ)

- Exploit conditional independence of responses given latent variables

$$f(y_{1j} \cap \cdots \cap y_{n_j j}|\zeta_j^{(2)};\boldsymbol{\beta},\boldsymbol{\lambda}) = \prod_i f(y_{ij}|\zeta_j^{(2)};\boldsymbol{\beta},\boldsymbol{\lambda})$$

- 'Marginal' likelihood contribution

$$\ell_j(\boldsymbol{\beta},\boldsymbol{\lambda},\psi) \;=\; \int \phi(\zeta_j^{(2)};0,\psi)\prod_i f(y_{ij}|\zeta_j^{(2)};\boldsymbol{\beta},\boldsymbol{\lambda})\mathrm{d}\zeta_j^{(2)}$$

- Change of variable $[u_j = \zeta_j^{(2)}/\sqrt{\psi}]$

$$\ell_j(\boldsymbol{\beta},\boldsymbol{\lambda},\psi) \;=\; \int \phi(u_j)\prod_i f(y_{ij}|\sqrt{\psi}u_j)\mathrm{d}u_j$$

- Gaussian quadrature approximation
  (e.g. Bock & Lieberman, 1970; Butler and Moffitt, 1982)

$$\ell_j(\boldsymbol{\beta},\boldsymbol{\lambda},\psi) \;\approx\; \sum_{r=1}^{R} W_r \prod_i f(y_{ij}|\sqrt{\psi}A_r),$$

  - $r = 1,..,R$ quadrature points
  - $W_r$ quadrature weights
  - $A_r$ locations
- Approximation exact if $\prod_i f(y_{ij}|\sqrt{\psi}u_j)$ were a $(2R-1)$th polynomial in $u_j$
- Software: aML, gllamm, MIXOR(etc.)

Slide 18

## Performance of GQ

- Pros:
  - Works well for dichotomous responses with small to moderate cluster sizes, precisely where MQL/PQL fails (e.g. Rodriguez and Goldman, 2001; Stryhn et al., 2000; Rabe-Hesketh et al., 2001).
  - Performance easily assessed by comparing solutions with different numbers of quadrature points.
- Cons:
  - Computationally intensive, particularly for many random effects. Large number of quadrature points often needed to closely approximate the likelihood (e.g. Crouch and Spiegelman, 1990)
  - Problematic if the posterior distribution has a sharp peak
    - ∗ Dichotomous responses with:
      - · large cluster sizes (e.g. Lee, 2000)
      - · high intraclass correlation (e.g. Lesaffre & Spiessens, 2001).
    - ∗ Counts (conditionally Poisson distributed) (e.g. Albert and Follmann, 2000).
    - ∗ Continuous responses (conditionally normally distributed)

## Adaptive Quadrature (AQ)

- Reformulation of likelihood contribution
  (e.g. Naylor and Smith, 1982; Liu and Pierce, 1994; Pinheiro and Bates, 1995; Rabe-Hesketh, Skrondal and Pickles, 2002)

$$\ell_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, \psi) = \int \underbrace{\phi(u_j;\mu_j,\sigma_j^2)} \left[ \frac{\phi(u_j) \prod_i f(y_{ij}|\sqrt{\psi}u_j)}{\underbrace{\phi(u_j;\mu_j,\sigma_j^2)}} \right] \mathrm{d}u_j$$

  - $\phi(u_j;\mu_j,\sigma_j^2)$: normal density approximating posterior (approximately proportional to the numerator in [ ])

**Slide 19**

- Change of variable $[v_j = (u_j - \mu_j)/\sigma_j]$

$$\ell_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, \psi) = \int \frac{\phi(v_j)}{\sigma_j} \left[ \frac{\phi(\sigma_j v_j + \mu_j) \prod_i f(y_{ij}|\sqrt{\psi}(\sigma_j v_j + \mu_j))}{\frac{1}{\sqrt{2\pi}\sigma_j} \exp(-v_j^2/2)} \right] \sigma_j \mathrm{d}v_j$$

- Approximation via adaptive quadrature

$$\ell_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, \psi) \approx \sum_{r=1}^{R} W_r \left[ \frac{\phi(\sigma_j A_r + \mu_j) \prod_i f(y_{ij}|\sqrt{\psi}(\sigma_j A_r + \mu_j))}{\frac{1}{\sqrt{2\pi}\sigma_j} \exp(-A_r^2/2)} \right]$$
$$= \sum_{r=1}^{R} \omega_{jr} \prod_i f(y_{ij}|\sqrt{\psi}\alpha_{jr})$$

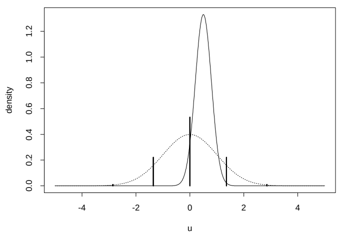
  - $\alpha_{jr} = \sigma_j A_r + \mu_j$
  - $\omega_{jr} = \sqrt{2\pi}\sigma_j \exp(A_r^2/2)\phi(\sigma_j A_r + \mu_j) W_r$

- Software: gllamm, SAS NLMIXED (only two-level models)

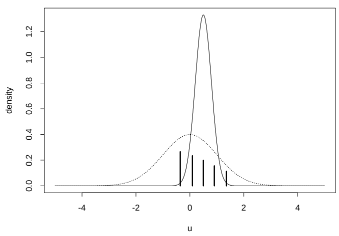
## Adpative quadrature places quadrature locations under peak of integrand

**Slide 20**

Quadrature




Adaptive quadrature




Prior (dotted curve) and posterior (solid curve) densities

(Integrand is proportional to posterior)

Slide 21

## Performance of AQ

- Pros:
  - Works well
    - if the posterior densities are nearly normal (e.g. large cluster sizes and/or counts.)
    - if the posterior densities are highly non-normal and not too peaked (e.g. dichotomous responses, small cluster sizes and moderate intraclass correlation)
    - even when the posterior densities are highly non-normal but with sharp peaks, e.g. dichotomous responses, small cluster sizes and large intraclass correlation
      (Rabe-Hesketh, Skrondal and Pickles, submitted)
  - Implementation in gllamm more computationally efficient than ordinary quadrature
- Cons:
  - Computationally intensive, especially for many latent variables. But: spherical rules improves efficiency
    (Rabe-Hesketh, Skrondal and Pickles, submitted)
  - Less computationally efficient than MQL/PQL

Slide 22

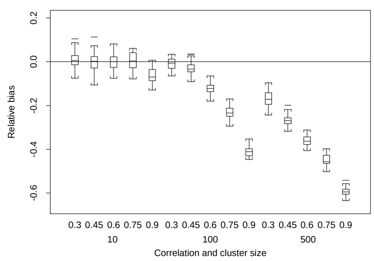
## Behaviour of GQ and AQ: Simulation

100 simulations from two-level probit random intercept model

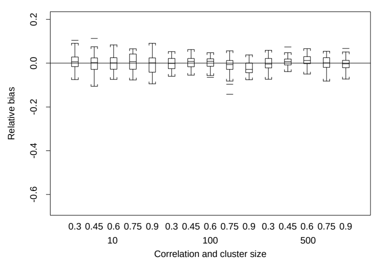
$$\eta_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2j} + \zeta_j^{(2)}$$

Relative bias $(\sqrt{\hat{\psi}^{(2)}} - \sqrt{\psi^{(2)}})/\sqrt{\psi^{(2)}}$ for different intra-class correlations $\rho$ and cluster sizes $n_j$.

**GQ**




**AQ**



Slide 23

## Behaviour of AQ: Poisson example

- A randomized longitudinal epilepsy trial comparing a new drug with placebo. Outcomes are seizure counts $y_{ij}$ for patient $j$ during the two weeks before visit $i$. Data from Leppik et.al, 1987, analysed by e.g. Thall and Vail, 1990; Breslow and Clayton, 1993
- Covariates (Breslow and Clayton, 1993):
  - lbas: log of a quarter of number of seizures in eight week preceding entry into the trial
  - treat: dummy variable for treatment
  - lbas_trt: interaction between lbas and treat
  - v4: dummy variable for fourth visit to account for drop in seizure counts during fourth interval
- Log-linear (Poisson regression) model with a random intercept for subjects

$$\log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j^{(2)}.$$

(Model II in Breslow and Clayton, 1993)

Slide 24

## Behaviour of AQ: Poisson example cont'd

| | Model II | | |
|---|---|---|---|
| | PQL-1 | GQ(20) | AQ(10) |
| *Fixed effects* | | | |
| lbas | 0.87 (0.14) | 0.94 (——) | 0.88 (0.13) |
| treat | -0.91 (0.41) | -1.42 (0.22) | -0.93 (0.40) |
| labs_trt | 0.33 (0.21) | 0.59 (0.11) | 0.34 (0.20) |
| lage | 0.47 (0.35) | 0.70 (0.30) | 0.48 (0.35) |
| v4 | -0.16 (0.05) | -0.16 (0.05) | -0.16 (0.05) |
| *Random effect* | | | |
| SD of intercept | 0.53 (0.06) | 0.51 (0.05) | 0.50 (0.06) |

Slide 25

## Non-parametric maximum likelihood estimation
## NPMLE

- The likelihood contribution has the form

$$l_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\pi}_R, \mathbf{z}_R) \;=\; \sum_{r=1}^{R} \pi_r \prod_i f(y_{ij}|\zeta_j^{(2)}\!=\!z_r)$$

  - $z_r$ are location parameters, $\pi_r$ are mass parameters and $\mathbf{z}_R$, $\boldsymbol{\pi}_R$ are the corresponding $R$-dimensional parameter vectors.
  - Gaussian quadrature has the same form with $z_r$ replaced by $\sqrt{\psi}A_r$ and $\pi_r$ replaced by $W_r$. But: crucial difference that $A_r$ and $W_r$ fixed apriori and not estimated.

  (e.g. Aitkin, 1999; Rabe-Hesketh, Pickles and Skrondal, revised)

- NPMLE:
  - Increase number of mass-points until likelihood does not increase
  - Mass-points introduced one by one using the concept of a directional derivative
- Software: gllamm

Slide 26

## Example:
## Logistic regr. with covariate measurement error

- Problem
  - Effect of fibre intake (continuous, measured twice on a subset of subjects) on coronary heart disease (CHD present/absent) (Morris, Marr and Clayton, 1977)
- Data and notation
  - Responses are dietary fibre intake ($i\!=\!1,2$) and coronary heart disease ($i=3$)
  - $\eta_j$ is $j$th subject's true dietary intake
- Measurement model for fibre intake: $y_{1j}$, $y_{2j}$ conditionally independently normally distributed with

$$\begin{aligned} E[y_{ij}|\eta_j] &= \beta_i + \eta_j, \quad i = 1,2 \; (\lambda_1\!=\!\lambda_2\!=\!1) \\ \mathrm{var}[y_{ij}|\eta_j] &= \sigma^2 \quad \text{[Measurement error variance]} \end{aligned}$$

- Disease model: $y_{i3}$ conditionally Bernoulli with

$$\mathrm{logit}(E[y_{3j}|\eta_j]) = \beta_3 + \eta_j\lambda_3 \quad [\lambda_3 \text{ is } \log(\mathrm{OR})]$$

Slide 27

## Structural equation models

1. **Solely direct effect of $x$ on $y_3$**

$$\begin{aligned} \text{Measurement model}: E[y_{ij}|\eta_j] &= \beta_i + \eta_j, \quad j = 1,2 \\ \text{Disease model}: \text{logit}(E[y_{3j}|\eta_j]) &= \beta_3 + \beta_4 x_j + \eta_j \lambda_3 \end{aligned}$$

2. **Solely indirect effect of $x$ on $y_3$**

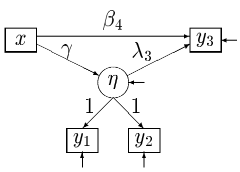
$$\eta_j = \gamma x_j + \zeta_j$$

$$\begin{aligned} \text{Measurement model}: E[y_{ij}|\eta_j] &= \beta_i + \gamma x_j + \zeta_j \\ \text{Disease model}: \text{logit}(E[y_{3j}|\eta_j]) &= \beta_3 + \gamma \lambda_3 x_j + \zeta_j \lambda_3 \end{aligned}$$

$\Rightarrow$ Requires nonlinear constraints or structural model

3. **Both direct and indirect effects of $x$ on $y_3$**

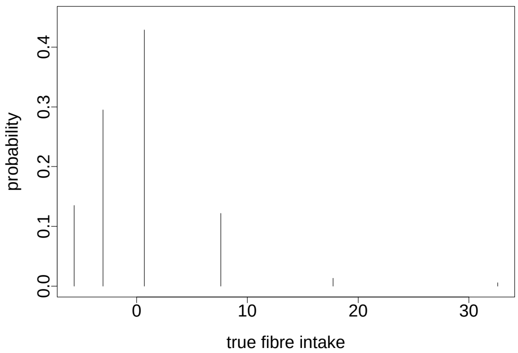
$$\begin{aligned} \text{Measurement model}: E[y_{ij}|\eta_j] &= \beta_i + \gamma x_j + \zeta_j \\ \text{Disease model}: \text{logit}(E[y_{3j}|\eta_j]) &= \beta_3 + (\beta_4 + \gamma\lambda_3) x_j + \zeta_j \lambda_3 \end{aligned}$$




Slide 28

## Normal vs. non-parametric exposure distribution

| | GQ(60) | | NPMLE(6) | |
|---|---|---|---|---|
| Parameters | Estimates | SE | Estimates | SE |
| $\lambda_3$ | -0.13 | 0.05 | -0.15 | 0.06 |
| $\sigma^2$ | 6.95 | 1.14 | 6.13 | 0.86 |
| $\psi^{(2)}$ | 23.66 | 0.52 | 24.93 | —– |
| | Log-likelihood=-1372.4 | | Log-likelihood=-1319.8 | |



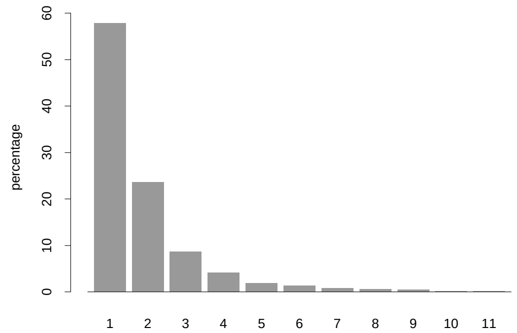
Slide 29

## Simulation: Normal true exposure distribution

| | | Quadrature (60 points) | | | |
|---|---|---|---|---|---|
| Parameter | True value | Mean | SD | Bias | RMSE |
| $\lambda_3$ | **-0.13** | -0.13 | 0.04 | 0.00 | 0.04 |
| $\sigma^2$ | **6.86** | 7.04 | 1.08 | 0.18 | 1.09 |
| $\psi^{(2)}$ | **23.81** | 23.35 | 2.64 | -0.46 | 2.67 |
| | | NPMLE | | | |
| Parameter | True value | Mean | SD | Bias | RMSE |
| $\lambda_3$ | **-0.13** | -0.13 | 0.04 | 0.00 | 0.04 |
| $\sigma^2$ | **6.86** | 6.59 | 1.11 | -0.27 | 1.14 |
| $\psi^{(2)}$ | **23.81** | 23.82 | 2.63 | 0.01 | 2.62 |

Slide 30

## Skewed true exposure distribution



Slide 31

## Simulation: Skewed true exposure distribution

| | | Quadrature (60 points) | | | |
|---|---|---|---|---|---|
| Parameter | True value | Mean | SD | Bias | RMSE |
| $\lambda_3$ | **-0.13** | -0.10 | 0.04 | 0.03 | 0.05 |
| $\sigma^2$ | **6.86** | 7.06 | 1.13 | 0.20 | 1.14 |
| $\psi^{(2)}$ | **23.81** | 23.00 | 4.83 | -0.81 | 4.88 |
| | | NPMLE | | | |
| Parameter | True value | Mean | SD | Bias | RMSE |
| $\lambda_3$ | **-0.13** | -0.14 | 0.08 | -0.01 | 0.08 |
| $\sigma^2$ | **6.86** | 6.42 | 0.98 | -0.44 | 1.07 |
| $\psi^{(2)}$ | **23.81** | 23.63 | 4.71 | -0.18 | 4.69 |

Slide 32

# III. PREDICTION OF LATENT VARIABLES

- Empirical Bayes (EB) prediction: posterior mean of latent variable given response with estimates plugged in
- Normally distributed latent variable:
  - Posterior mean for unit $j$

$$\tilde{\zeta}_j^{\mathsf{EB}} = \mathrm{E}[\zeta_j|\mathbf{y}_j;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\psi}] = \frac{\int \zeta_j \phi(\eta_j;0,\widehat{\psi}) \prod_i f(y_{ij}|\zeta_j;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda})\mathrm{d}\zeta_j}{\ell_j(\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\psi})}$$

  - Posterior variance for unit $j$

$$\mathrm{var}[\zeta_j|\mathbf{y}_j;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\psi}] = \frac{\int (\zeta_j)^2 \phi(\zeta_j;0,\widehat{\psi}) \prod_i f(y_{ij}|\zeta_j;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda})\mathrm{d}\zeta_j}{\ell_j(\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\psi})} - \mathrm{E}[\zeta_j|\mathbf{y}_j;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\psi}]^2$$

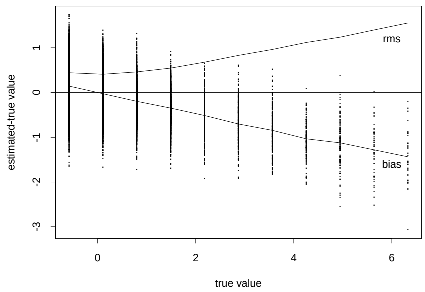
  - The integrals are evaluated using adaptive quadrature
- NPMLE-based prediction:

$$\tilde{\zeta}_j^{\mathsf{EB}} = \mathrm{E}[\zeta_j|\mathbf{y}_j;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\boldsymbol\pi}_K,\widehat{\mathbf{z}}_K] = \frac{\sum_{k=1}^K \widehat{z}_k \widehat{\pi}_k \prod_{i=1} f(y_{ij}|\zeta_j = \widehat{z}_k;\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\boldsymbol\pi}_K,\widehat{\mathbf{z}}_K)}{l_j(\widehat{\boldsymbol\beta},\widehat{\boldsymbol\lambda},\widehat{\boldsymbol\pi}_K,\widehat{\mathbf{z}}_K)}$$

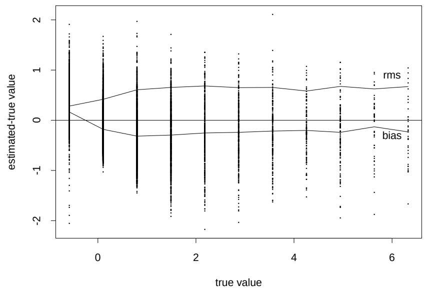
Slide 33

## Performance: Prediction for skewed exposure

**Normality assumed**




**NPMLE**




Slide 34

## Variances of latent variable prediction

1. Posterior variance:

$$\mathrm{var}(\zeta_j|\mathbf{y}_j, \mathbf{x}_j, \widehat{\boldsymbol{\theta}})$$

2. Marginal sampling variance:

$$\mathrm{var}_{\mathbf{y}}(\tilde{\zeta}_j^{\mathsf{EB}}|\mathbf{x}_j, \widehat{\boldsymbol{\theta}})$$

   'Diagnostic' variance in Goldstein (1995)

3. Conditional sampling variance:

$$\mathrm{var}_{\mathbf{y}}(\tilde{\zeta}_j^{\mathsf{EB}}|\zeta_j, \mathbf{x}_j, \widehat{\boldsymbol{\theta}})$$

4. Prediction error variance (marginal):

$$\mathrm{var}_{\mathbf{y}}(\tilde{\zeta}_j^{\mathsf{EB}} - \zeta_j|\mathbf{x}_j, \widehat{\boldsymbol{\theta}})$$

   'Comparative' variance in Goldstein (1995)

Slide 35

## Obtaining the standard errors of prediction

- Posterior variance obtained by numerical integration (as shown)
- Approximate expression for prediction error variance

$$\text{var}[\tilde{\zeta}_j^{\text{EB}} - \zeta_j | \mathbf{x}_j; \widehat{\boldsymbol{\theta}}] \approx \text{var}[\zeta_j | \mathbf{y}_j, \mathbf{x}_j; \widehat{\boldsymbol{\theta}}].$$

- Approximate expression for the marginal sampling variance

$$\begin{aligned}\text{var}[\zeta_j | \mathbf{x}_j; \widehat{\boldsymbol{\theta}}] &= \text{E}_{\mathbf{y}}\big[\text{var}[\zeta_j | \mathbf{y}_j, \mathbf{x}_j; \widehat{\boldsymbol{\theta}}]\big] \\ &\quad + \text{var}_{\mathbf{y}}\underbrace{\big[\text{E}[\zeta_j | \mathbf{y}_j, \mathbf{x}_j; \widehat{\boldsymbol{\theta}}]\big]}_{\tilde{\zeta}_j^{\text{EB}}}.\end{aligned}$$

  This can be rewritten as

$$\begin{aligned}\text{var}_{\mathbf{y}}[\tilde{\zeta}_j^{\text{EB}} | \mathbf{x}_j; \widehat{\boldsymbol{\theta}}] &= \text{var}[\zeta_j | \mathbf{x}_j; \widehat{\boldsymbol{\theta}}] - \text{E}_{\mathbf{y}}\left[\text{var}[\zeta_j | \mathbf{y}_j, \mathbf{x}_j; \widehat{\boldsymbol{\theta}}]\right] \\ &\approx \widehat{\psi} - \text{var}[\zeta_j | \mathbf{y}_j, \mathbf{x}_j; \widehat{\boldsymbol{\theta}}].\end{aligned}$$

Slide 36

# IV. CONCLUSION

- Adaptive quadrature appears to work well in all situations
  $\Longrightarrow$ MQL/PQL and ordinary Gaussian quadrature obsolete?
- Estimation based on normality appears to be quite robust against misspecification of the latent variable distribution
- Prediction based on normality does <u>not</u> appear to be robust against misspecification of the latent variable distribution
  $\Longrightarrow$ Non-parametric maximum likelihood should be used for prediction of latent variables
- Suggested methodology for estimation and prediction is
  - applicable for entire GLLAMM framework (not just random intercept models!)
  - implemented in gllamm software running in STATA

Slide 37

## Selected references from GLLAMM project

- GLLAMM framework:
  - *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/ CRC, to appear. (A.Skrondal & S.Rabe-Hesketh).
  - "Generalised multilevel structural equation modelling". *Psychometrika*, 2003, in press (S.Rabe-Hesketh, A.Skrondal & A.Pickles).
  - "Parametrization of multivariate random effects models for categorical data". *Biometrics*, 57: 1256-1264, 2001. (S.Rabe-Hesketh & A.Skrondal).
  - "Multilevel logistic regression for polytomous data and rankings". *Psychometrika*, 2003, in press (A.Skrondal & S.Rabe-Hesketh).
- gllamm software:
  - *GLLAMM Manual*. London: Institute of Psychiatry, 2001. [128p]. (S.Rabe-Hesketh, A.Pickles & A.Skrondal). `www.iop.kcl.ac.uk/IoP/Departments/BioComp/programs/gllamm.html`
- Adaptive quadrature:
  - "Reliable estimation of generalized linear mixed models using adaptive quadrature". *The Stata Journal*, 2: 1-21, 2002. (S.Rabe-Hesketh, A.Skrondal & A.Pickles).
  - "Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects". Submitted to *Journal of Econometrics*. (S.Rabe-Hesketh, A.Skrondal & A.Pickles).
- Non-parametric maximum likelihood:
  - "Correcting for covariate measurement error in generalised linear models using nonparametric maximum likelihood estimation". Revised for *Statistical Modelling*. (S.Rabe-Hesketh, A.Pickles & A.Skrondal).