Optimum Design for Mixed Effects Non-Linear and generalized Linear Models

Cambridge, August 9-12, 2011

Non-maximum likelihood estimation and statistical inference for linear and nonlinear mixed models

Eugene Demidenko

Dartmouth College

New Hampshire, USA

www.dartmouth.edu/~eugened

Outline

1. The essence of the mixed model.

2. What mixed model has something to do with the Simpson's paradox.

3. Mixed model nomenclature.

4. Why mixed model?

5. Coefficient of determination for random effects.

6. Criterion for existence of the maximum likelihood estimate.

7. Unbiasedeness.

8. Non-iterative estimation of the variance/covariance parameters.
9. Balanced data–the favorable data.
10. Generalized linear mixed model.
11. Binary model with random intercept.
12. Individual and population-based (marginal) model.
13. Laplace approximation for a general nonlinear mixed model.
14. Poisson model.
15. Deterministic and stochastic asymptotics.
16. Challenges of optimal design with mixed model.

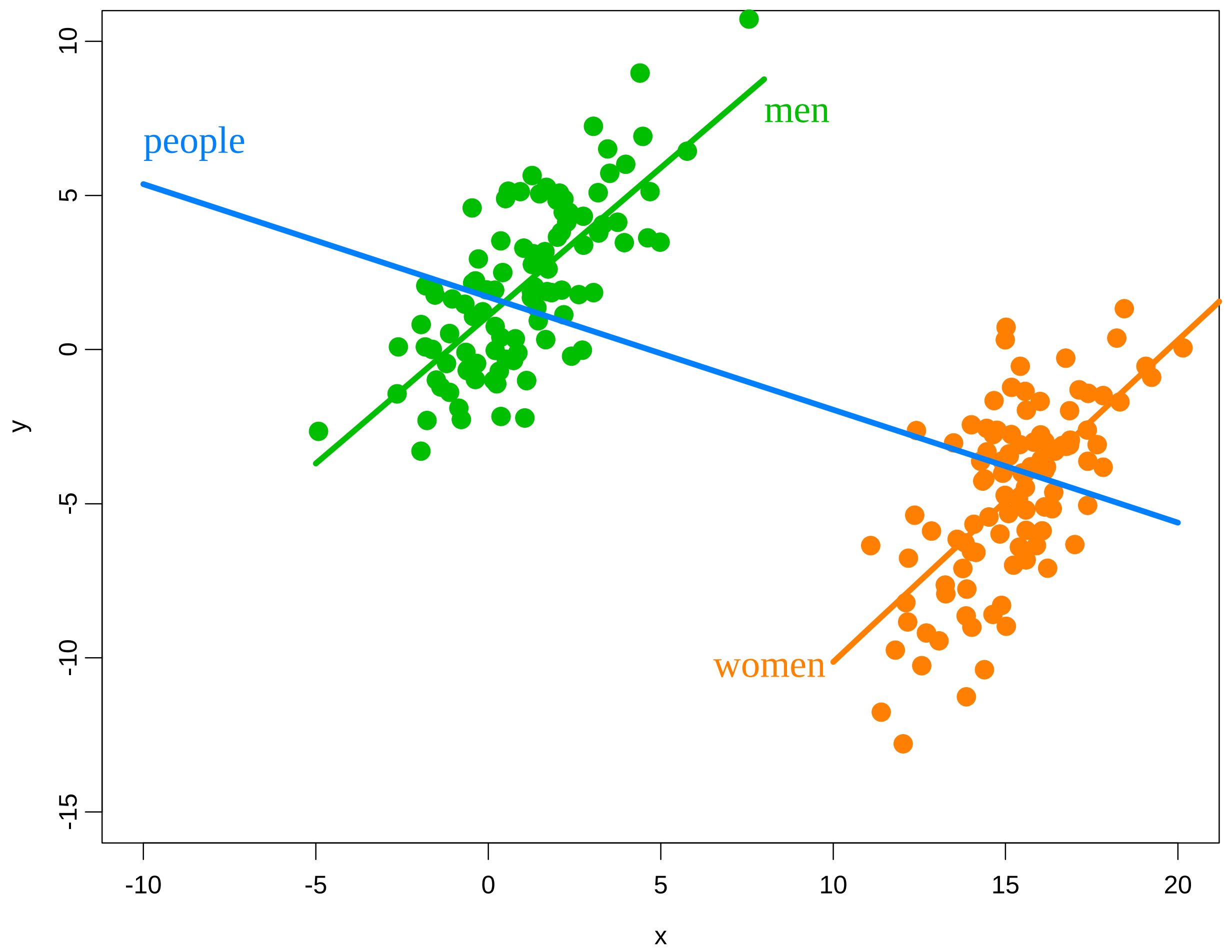
1 What is mixed model and when it's used?

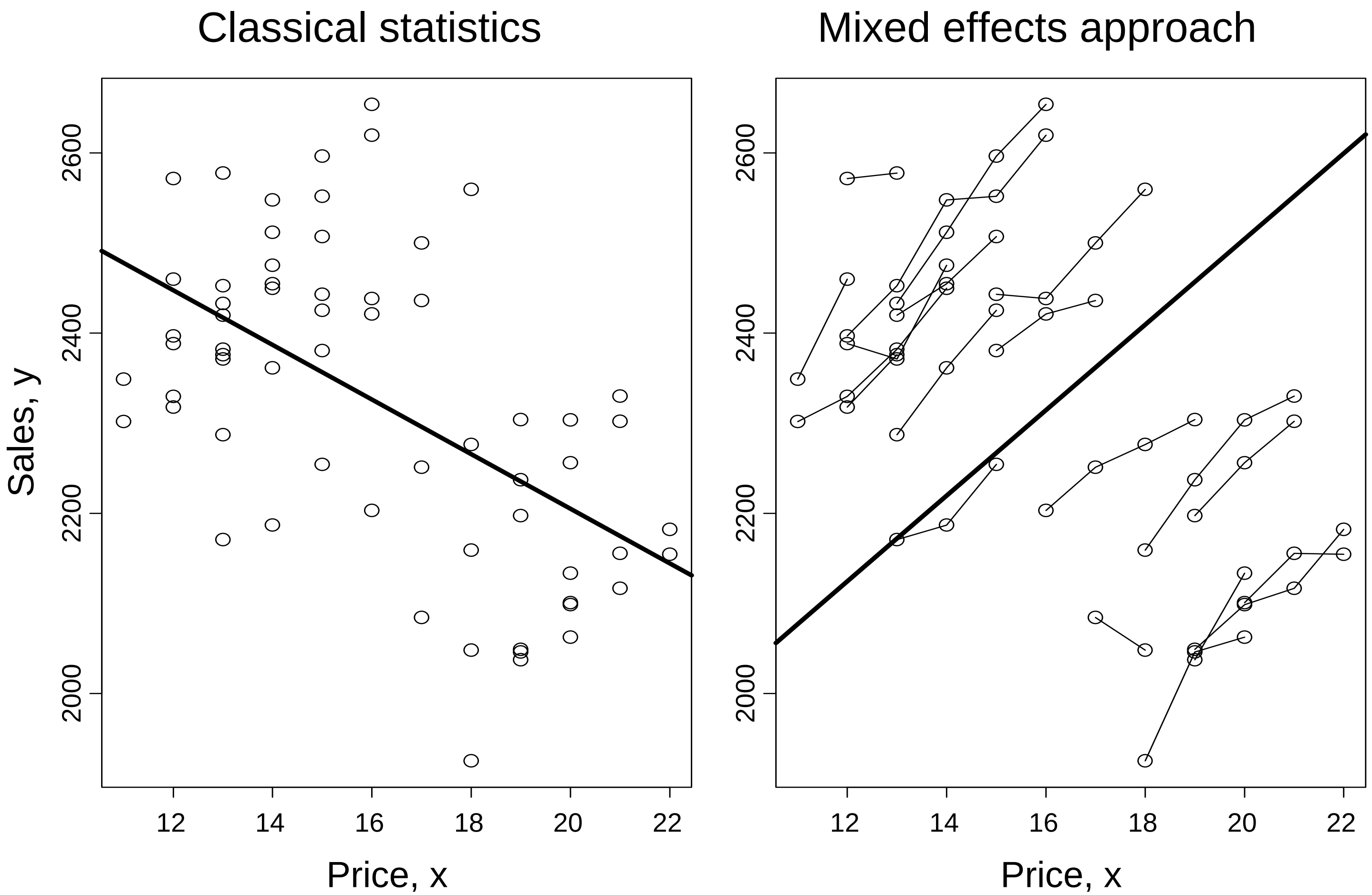
- MM specifies the variance, not the mean, e.g. in regression $y_i = \boldsymbol{\beta}'\mathbf{x}_i + \eta_i$, we may specify $var(\eta_i) = \sigma^2(1 + \gamma z_i^2)$. Consequently, classical OLS unbiasedely estimates $\boldsymbol{\beta}$ (this is true only for linear MM).

- MM is used to model multilevel/clustered data, e.g. pupils from the same school, patients from the same geographical region (MM does not effect fixed effects but increases the variance due to intra-cluster correlation). Consequently, the variances of fixed effect estimates are larger because they are more realistic.

- Often MM is used for longitudinal analysis, e.g. blood pressure over the time for patients with a different baseline (random intercept model).

2 Simpson's paradox: good for men, good for women, bad for people





Classical and mixed effects approaches lead to reverse conclusions.

3 Mixed model nomenclature

	Mixed Model (MM)	
A. Linear MM (LMM)	B. Generalized LMM	C. Nonlinear MM
Meta-analysis model*	Logistic regression (LR)	Nonlinear marginal model
Random intercept model	LR with ri*	Fixed matrix of re*
Growth curve model (GCM)*	Probit regression (PR)*	Varied matrix of re*
Rectangular GCM*	PR with ri*	Type I,II,III Nmm*
Balanced GCM*	Poisson regression (PoiR)*	General Nonlinear MM
General LMM*	PoiR with ri*	
LMM with linear covariance*	Beta-binomial*	

* non-maximum likelihood unbiased estimation (no integration) is available

ri=random intercept; re=random effect

4 Estimation methods nomenclature

- Maximum likelihood for GLMM and nonlinear MM require integration to obtain the marginal distribution (to integrate out the unobservable random effects). The dimension of the integral equals the number of random effects. The MLE is especially difficult for models with high-dimension random effects.

- Linear mixed model: method of moments (MM) or MINQUE estimation for the matrix of random effects and consequent estimation by generalized LS.

- Integral approximation via Laplace approximation $\longrightarrow$ pseudo-likelihood approach/penalized likelihood (quadratic term).

- Model-specific integral estimation, like logistic/probit or Poisson regressions with random intercepts.

5 Why mixed model?

Linear mixed model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad cov(\mathbf{b}_i) = \sigma^2\mathbf{D}$$

The test

$$H_0 : \mathbf{D} = \mathbf{0}$$

is fundamental.

Theorem. Under the null

$$\frac{(S_{OLS} - S_{\min})/(r-m)}{S_{\min}/(N_T - r)} \sim F(r-m, N_T - r)$$

where S_{OLS} is the OLS residual sum of squares and $S_{\min}$ is the residual sum of squares via the dummy variable approach (global minimum RSS),

$$S_{\min} = \min_{\boldsymbol{\beta},\mathbf{b}_1,\ldots,\mathbf{b}_N} \sum_{i=1}^{N} \| \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i \|^2$$

and

$$
\begin{aligned}
r &= rank(\mathbf{W}), \\
\mathbf{W} &= \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \mathbf{Z}_N \end{bmatrix}
\end{aligned}
$$

This test is **exact** and is a generalization of the variance components test (Searle 1992, Khuri 1998).

6 Random effect coefficient of determination

Fixed effect coefficient of determination (the proportion of the variance of y explained by x) :

$$R^2_{\text{fixed effect}} = \mathbf{1} - \frac{\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})'\widehat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})}{\sum_{i=1}^{N}(\mathbf{y}_i - \overline{y}_i\mathbf{1}_i)'\widehat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \overline{y}_i\mathbf{1}_i)}$$

where $\widehat{\mathbf{V}}_i = \mathbf{I} + \mathbf{Z}_i\widehat{\mathbf{D}}\mathbf{Z}_i'$.

Random effect coefficient of determination (the proportion of the variance of random effects on the maximum scale):

$$R^2_{\text{random effect}} = \mathbf{1} - \frac{S_{MM} - S_{\text{min}}}{S_{OLS} - S_{\text{min}}}$$

where

$$S_{MM} = \sum_{i=1}^{N} \| \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{GLS} - \mathbf{Z}_i\widehat{\mathbf{b}}_i \|^2 .$$

7 $R^2_{\text{random effect}}$ interpretation

- $0 \leq R^2_{\text{random effect}} \leq 1$

- If $R^2_{\text{random effect}} = 1$ then $S_{MM} = S_{\text{min}}$ so that random effects become free fixed effects ($\mathbf{D} = \infty$).

- If $R^2_{\text{random effect}} = 0$ then $S_{MM} = S_{LS}$ and random effects are not present (matrix $\mathbf{D} = \mathbf{0}$).

$R^2_{\text{random effect}}$ = the proportion of the variance reduction, due to random effects, on the scale of the maximum reduction.

8 Existence of ML

Theorem. The MLE for LME exists iff

$$S_{\mathsf{min}} > 0.$$

Example of invalid linear MM: fixed and random intercepts cannot be identified simulteneosly,

$$y_{ij} = \alpha_i + \boldsymbol{\gamma}'\mathbf{u}_{ij} + b_i + \varepsilon_{ij}$$

where α_i is the fixed intercept and b_i is the random effect.

Correct model $(\alpha_i = const)$:

$$y_{ij} = \alpha + \boldsymbol{\gamma}'\mathbf{u}_{ij} + b_i + \varepsilon_{ij}$$

9 Existence of a nonnegative matrix $\widehat{\mathrm{D}}_{ML}$

Theorem 1 *Let* $\widehat{\mathrm{e}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{OLS}$ *denote the* $n_i \times 1$ *OLS residual vector and* $\widehat{\sigma}^2_{OLS} = \sum \parallel \widehat{\mathrm{e}}_i \parallel^2 / N_T$ *denote the OLS variance. If the* $k \times k$ *matrix*

$$\sum \mathbf{Z}_i' \widehat{\mathrm{e}}_i \widehat{\mathrm{e}}_i' \mathbf{Z}_i - \widehat{\sigma}^2_{OLS} \sum \mathbf{Z}_i' \mathbf{Z}_i$$

is nonzero nonnegative definite, then $\widehat{\mathrm{D}}_{ML}$ *is a nonnegative definite matrix.*

One random effect $(k = 1)$: if

$$\widehat{\sigma}^2_{OLS} < \frac{\sum (\mathbf{z}_i' \widehat{\mathrm{e}}_i)^2}{\sum \mathbf{z}_i' \mathbf{z}_i}$$

then $\widehat{d}_{ML} > 0$.

10 Unbiased estimation of fixed effects

The most important problem is estimation of matrix D

GLS estimator of fixed effect coefficients:

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N} \mathbf{X}_i'(\mathbf{I} + \mathbf{Z}_i\widehat{\mathbf{D}}\mathbf{Z}_i')^{-1}\mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{X}_i'(\mathbf{I} + \mathbf{Z}_i\widehat{\mathbf{D}}\mathbf{Z}_i')^{-1}\mathbf{y}_i\right].$$

where $cov(\mathbf{b}_i) = \sigma^2\mathbf{D}$.

Theorem. Under normal (or symmetric) distribution, $\widehat{\boldsymbol{\beta}}_{ML}$ is an unbiased estimator (small sample, exact) and for most other noniterative estimators of matrix $\mathbf{D}$.

11 Noniterative estimation of variances

1. MINQUE (minimum norm unbiased estimator)

2. MM (method of moments)

3. VLS (variance least squares)

11.1 MINQUE for σ^2

Quadratic estimator

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{A}\mathbf{y}$$

with minimum norm

$$tr(\mathbf{A}\mathbf{A}') \Rightarrow \text{min}\,.$$

Unbiasedness leads to

$$\mathbf{W}'\mathbf{A}\mathbf{W} = \mathbf{0}, \quad tr(\mathbf{A}) = 1.$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \mathbf{Z}_N \end{bmatrix}$$

The solution

$$\widehat{\sigma}^2_{MINQUE} = \frac{S_{\text{min}}}{\sum n_i - rank(\mathbf{W})}$$

where (remind)

$$S_{\text{min}} = \min_{\boldsymbol{\beta},\mathbf{b}_1,\ldots,\mathbf{b}_N} \sum_{i=1}^{N} \| \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i \|^2$$

is found from a dummy variable approach, the global RSS.

Theorem. MINQUE=RMLE for the balanced random-intercept model

$E\widehat{\sigma}^2_{MINQUE} = \sigma^2$.

12 Method of moments for $\mathbf{D}$

'Estimate' random effects as

$$\widehat{\mathbf{b}}_i = (\mathbf{Z}_i'\mathbf{Z}_i')^{-1}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{OLS})$$

and solve equation

$$E\left(\sum_{i=1}^{N} \widehat{\mathbf{b}}_i\widehat{\mathbf{b}}_i'\right) = \sum_{i=1}^{N} \widehat{\mathbf{b}}_i\widehat{\mathbf{b}}_i'$$

for $\mathbf{D}$.

Special cases are treated, such as balance random intercept model, a close connection to MINQUE and previously derived estimators of variance components model (Searle et al, 1992).

13 Variance least squares for D_*

Estimate using OLS and obtain residuals $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{OLS}$, then minimize

$$\sum_{i=1}^{N} tr\left(\widehat{\mathbf{e}}_i \widehat{\mathbf{e}}_i' - \sigma^2 \mathbf{I} - \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i'\right)^2$$

to get estimates for σ^2 and $\mathbf{D}_*$.

Unbiased version is derived for σ^2 and $\mathbf{D}_*$.

14 Balanced data/model

Balanced data, and the respective model have a remarkable property that the estimate of fixed effects (beta-coefficients) do not depend on the matrix $\mathbf{D}$.

Balanced data have the same number of observations per cluster, $n_i = n$.

Examples:

- Random inetrcept MM: $\mathbf{y}_i^{n\times\mathbf{1}} = a_i + \mathbf{X}^{n\times m}\boldsymbol{\beta}^{m\times\mathbf{1}} + \boldsymbol{\varepsilon}_i$, $a_i = \alpha + b_i$, $i = \mathbf{1}, ..., N, j = \mathbf{1}, .., n$.

- Balanced rectangular linear growth curve: $\mathbf{y}_i = \mathbf{Z}\mathbf{a}_i + \boldsymbol{\varepsilon}_i$, where $\mathbf{a}_i = (\mathbf{I} \otimes \mathbf{q}_i')\boldsymbol{\beta} + \mathbf{b}_i$.

15 Generalized linear mixed model (GLMM)

Exponential family of distributions (McCullagh and Nelder, 1989)

$$f(y;\theta,\phi) = \exp\left[\frac{\theta y - b(\theta)}{\phi} - c(y,\phi)\right],$$

where $b' = \mu$, the inverse link, and $b'' = var(y)$.

For cluster data $\{y_{ij}\}$, assuming that $y_{ij}|\mathbf{u}_i$ are independent and following the exponential family distribution and

$$\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_*).$$

The marginal log-likelihood for the GLMM takes the form

$$l(\boldsymbol{\beta}, \mathbf{D}_*) = -\frac{Nk}{2}\ln(2\pi) - \frac{N}{2}\ln|\mathbf{D}_*| + \sum_{i=1}^{N}\ln\int_{R^k} e^{l_i(\boldsymbol{\beta},\mathbf{u}) - \frac{1}{2}\mathbf{u}'\mathbf{D}_*^{-1}\mathbf{u}} d\mathbf{u},$$

where

$$l_i(\boldsymbol{\beta}, \mathbf{u}) = \sum_{j=1}^{n_i} \left[(\boldsymbol{\beta}'\mathbf{x}_{ij} + \mathbf{u}'\mathbf{z}_{ij}) y_{ij} - b(\boldsymbol{\beta}'\mathbf{x}_{ij} + \mathbf{u}'\mathbf{z}_{ij}) \right]$$

The most popular links are: logit, probit, Poisson.

16 Binary regression with random intercepts

y_{ij} is binary (0 or 1) with

$$\Pr(y_{ij} = 1|a_i) = \mu(a_i + \boldsymbol{\beta}'\mathbf{x}_{ij})$$

where $a_i \sim \mathcal{N}\left(\alpha, \sigma^2\right)$ and μ is the probability function:

1. $\mu' > 0$

2. $\lim_{s\to-\infty} \mu(s) = 0$ and $\lim_{s\to\infty} \mu(s) = 1$.

3. $(\ln \mu)'' < 0$.

Applications: logit in epidemiology, probit in toxicology and econometrics.

17 Individual and population-based models

Probit model with random intercepts

$$\Pr(y_{ij} = 1 | u_i) = \Phi(u_i + \boldsymbol{\beta}'\mathbf{x}_{ij}), \quad u_i \sim \mathcal{N}\left(0, \sigma^2\right)$$

specifies individual (within cluster) data.

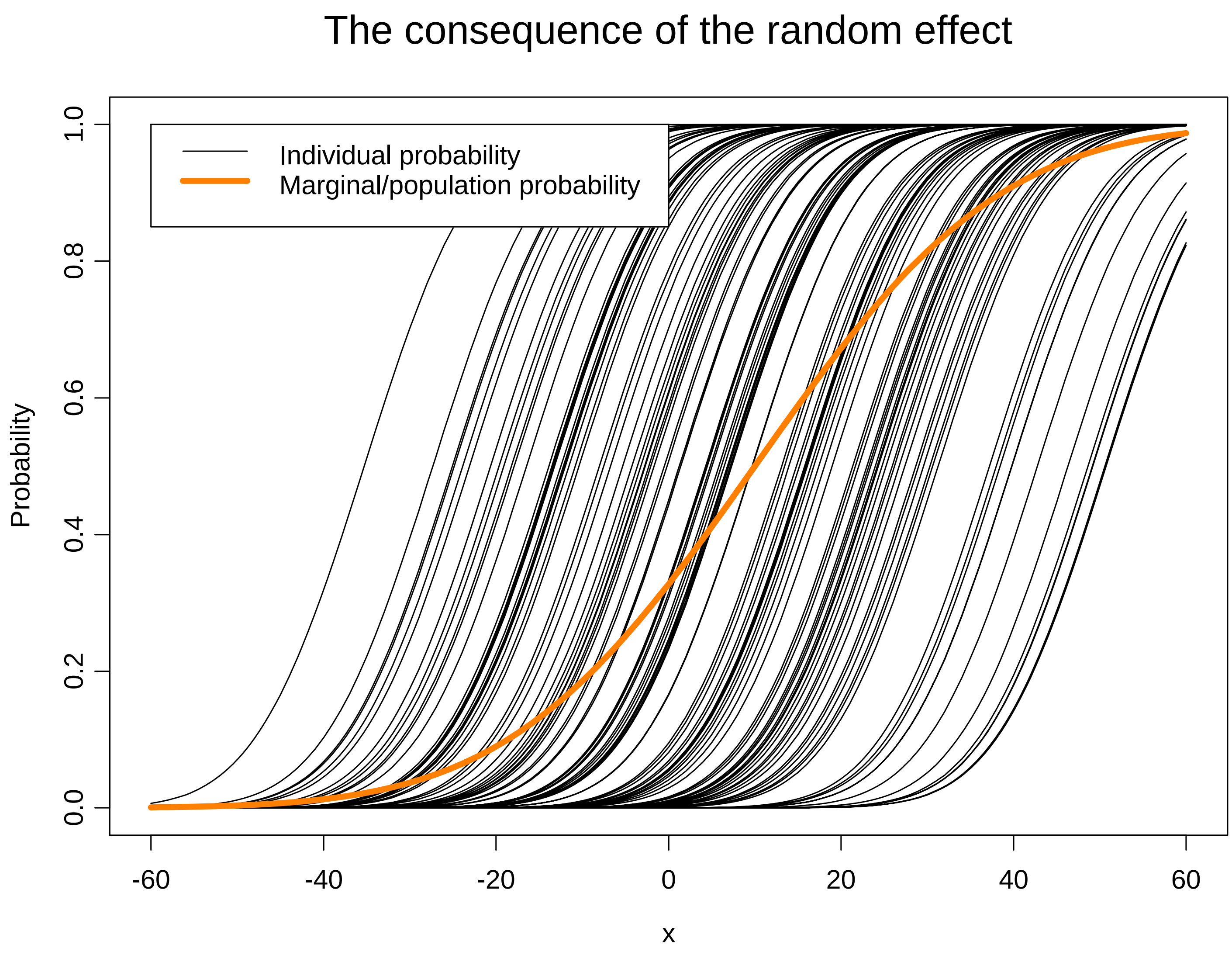
Lemma.

$$E_{u \sim \mathcal{N}(0,\sigma^2)} \Phi(s + u) = \Phi\left(\frac{s}{\sqrt{1 + \sigma^2}}\right).$$

Marginal mean

$$\Pr(y_{ij} = 1) = E(y_{ij}) = \Phi\left(\frac{\boldsymbol{\beta}'}{\sqrt{1 + \sigma^2}} \mathbf{x}_{ij}\right)$$

determines population-based mean/average.



The polulation-based average response is *attenuated*.

18 Laplace approximation of the integral

Let $h(\mathbf{x})$ be a function of $\mathbf{x} \in R^k$, then

$$\int_{R^k} e^{h(\mathbf{x})} d\mathbf{x} \simeq (2\pi)^{k/2} e^{h_{\max}} \left| - \left. \frac{\partial^2 h}{\partial \mathbf{x}^2} \right|_{\mathbf{x}=\mathbf{x}_{\max}} \right|^{-1/2}.$$

Approximate $h(\mathbf{x})$ with the second order Taylor series expansion function at the maximum.

Laplace approximation of the maximum likelihood leads to the penalized quasi-likelihood estimation (Linstrom and Bates, 1992; Breslow and Clayton, 1993).

19 GEE for MM

GEE was developed in eighties (Scott Zeger) for estimation of longitudinal GLM. It was regarded as one of the highlights of statistical science of the last century.

Theorem (consistency and asymptotic normality of EE approach). Let $\mathbf{y}_i^{n_i \times 1} = \mathbf{f}_i(\boldsymbol{\theta}^{m \times 1}) + \boldsymbol{\eta}_i$ where $cov(\boldsymbol{\eta}_i) = \mathbf{V}_i$, may be a function of $\boldsymbol{\theta}$. Then under mild regularity conditions the solution of

$$\sum_{i=1}^{N} \mathbf{H}_i^{m \times n_i}(\mathbf{y}_i - \mathbf{f}_i(\boldsymbol{\theta})) = \mathbf{0}$$

is consistent and asymptotically normal.

The idea of GEE is to use an estimating equation

$$\sum_{i=1}^{N} \mathbf{X}_i' \dot{\boldsymbol{\mu}}_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0},$$

where

$$\begin{aligned} E(\mathbf{y}_i) &= \mu(\mathbf{X}\boldsymbol{\beta}), \quad \dot{\boldsymbol{\mu}}_i = \mu'(\mathbf{X}\boldsymbol{\beta}) \\ \mathbf{V}_i &= \mathbf{D}_i^{1/2} \mathbf{R}_i(\boldsymbol{\gamma}) \mathbf{D}_i^{1/2} \end{aligned}$$

and is $\mathbf{R}_i(\boldsymbol{\gamma})$ a **working** correlation matrix.

Generalized Estimation Equation (GEE) for GLM is an estimation method of nonexistent model.

Fundamental problems of GEE applied to a nongaussian linear model, such as logistic regression, are:

1. It is impossible to generate data for GEE, e.g. when simulations are needed (method for a nonexistent model).

2. The covariance matrix depends on fixed effects so the separate modeling of main and cluster dependence is invalid. Example: logistic regression $E(y) = p$ and $var(y) = p(1-p)$ are connected. It's wrong to write $y = p + \eta$ for binary data y.

3. GEE does not yield a consistent estimation of GLMM or nonlinear model.

4. May be used for marginal nonlinear models or to obtain a starting point for MM iterations

20 Nonlinear mixed model

Hierarchical formulation (normal distribution)

$$\begin{aligned} \mathbf{y}_i^{n_i\times 1}|\mathbf{a}_i^{k\times 1} &\sim \mathcal{N}\left(\mathbf{f}_i(\boldsymbol{\gamma},\mathbf{a}_i),\sigma^2\mathbf{I}\right), \\ \mathbf{a}_i &\sim \mathcal{N}\left(\mathbf{A}_i\boldsymbol{\beta},\sigma^2\mathbf{D}\right) \end{aligned}$$

The marginal log-likelihood function (up to a constant term)

$$\begin{aligned} & l(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\theta}) \\ = & -\frac{1}{2}\left[N\ln|\,\mathbf{D}\,| + \sum_{i=1}^{N}(n_i+k)\ln(\sigma^2) + \sum_{i=1}^{N}\ln\int_{R^k} g_i(\boldsymbol{\gamma},\mathbf{a},\boldsymbol{\beta},\boldsymbol{\theta})d\mathbf{a}\right] \end{aligned}$$

where $\boldsymbol{\theta} = (\sigma^2, vech(\mathbf{D}))$ and

$$\begin{aligned} & g_i(\boldsymbol{\gamma},\mathbf{a},\boldsymbol{\beta},\boldsymbol{\theta}) \\ = & \exp\left\{-\frac{1}{2\sigma^2}\left[\|\,\mathbf{y}_i-\mathbf{f}_i(\boldsymbol{\gamma},\mathbf{a})\,\|^2 + (\mathbf{a}-\mathbf{A}_i\boldsymbol{\beta})'\mathbf{D}^{-1}(\mathbf{a}-\mathbf{A}_i\boldsymbol{\beta})\right]\right\}. \end{aligned}$$

21 Logistic regression with random intercepts

Use the idea of *conditional* regression: to obtain binary outcomes conditioned by the number of successes, $k_i = \sum_{j=1}^{n_i} y_{ij}$. The trick is that then nuisance random intercepts dissappear. The conditional log-likelihood is

$$l_C(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{r} - \sum_{i=1}^{N} \ln \left(\sum_{\mathbf{z} \in \mathbf{Z}_{k_i}^{n_i}} e^{\boldsymbol{\beta}' \sum_{j=1}^{n_i} \mathbf{x}_{ij} z_j} \right),$$

where $\mathbf{r} = \sum_{i=1}^{N} \sum_{j=1}^{n} \mathbf{x}_{ij} y_{ij}$ and z_j is an indicator variable (0 or 1) such that $\sum_{j=1}^{n_i} z_j = k_i$, and $\mathbf{Z}_{k_i}^{n_i}$ is the set of all such zs.

Very close to MLE but requires not large clusters, an obvious connection to the partial likelihood idea in Cox proportional survival models.

Computationally attractive for large number of clusters and small number of observations per cluster (twin studies).

22 Poisson regression with random effects

Let y_{ij} be count (i codes the cluster, j codes the observation in the ith cluster).

For Poisson regression (log link) we have

$$E(y_{ij}|u_i) = e^{\alpha + u_i + \mathbf{b}'\mathbf{x}_{ij}}, \quad u_i \sim \mathcal{N}(0, \sigma^2).$$

The marginal expected value is

$$E(y_{ij}) = e^{\alpha + \sigma^2/2 + \mathbf{b}'\mathbf{x}_{ij}}.$$

The naive Poisson regression is ok: the random effect affects only the intercept term, the slope coefficients remain consistent.

Comparison of the efficiency for six methods of balanced ($n_i = n$) Poisson regression with normally distributed random intercepts.

1. OP=Ordinary Poisson regression (intercept=const).

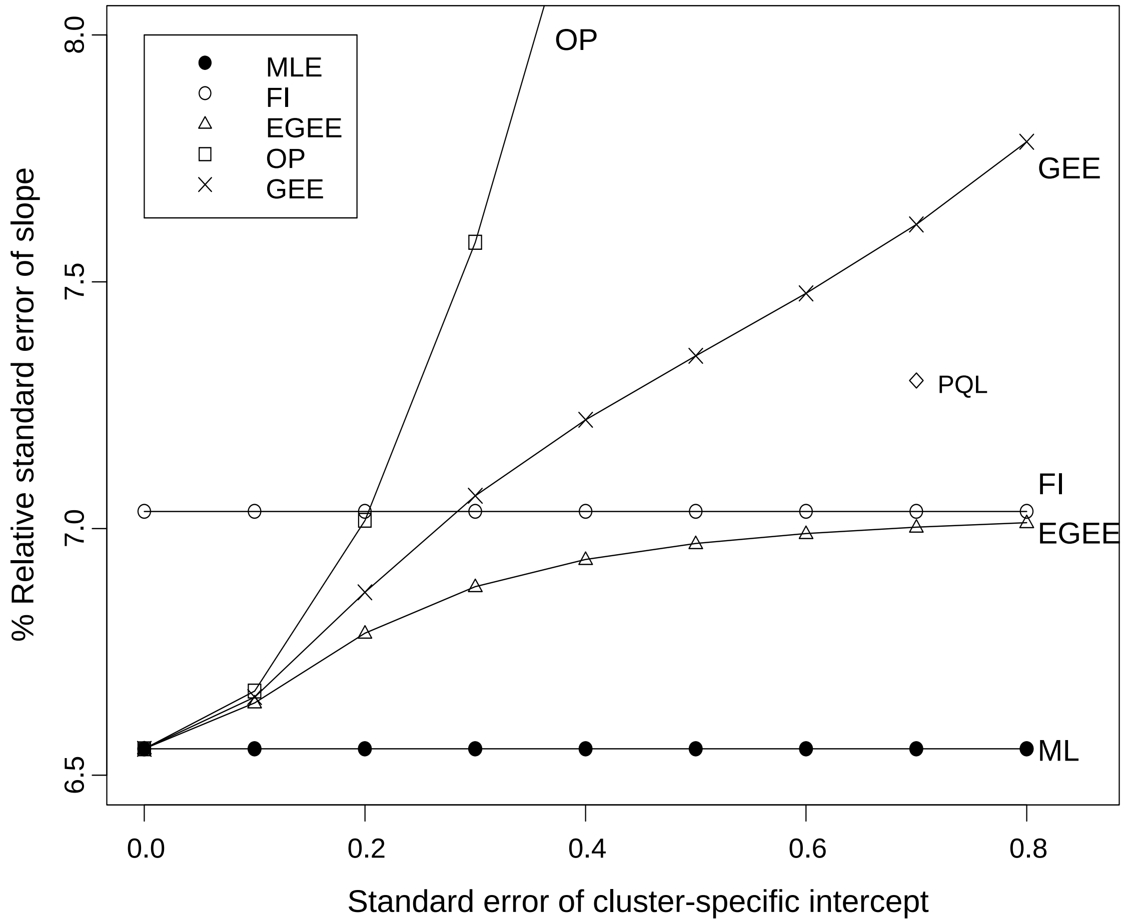
2. FI=Fixed cluster-specific intercept (dummy varaible approach).

3. GEE=compound symmetry correlation structure.

4. EGEE=Exact GEE, $cov(\mathbf{y}_i) = \mathbf{E}_i + (\mathbf{1} - e^{-\sigma^2})\mathbf{e}_i\mathbf{e}_i'$ where $\mathbf{e}_i = e^{\mathbf{b}'\mathbf{x}_i}$ and $\mathbf{E}_i = diag(\mathbf{e}_i)$.

5. Pseudo-likelihood (Laplace approximation) estimation.

6. MLE (integration).



Percent relative SE of slope estimator (β) as a function of SE intercept (σ) in clustered Poisson regression with random intercept. The variance of ML and FI estimators do not depend on the intercept variance, σ^2. EGEE is between FI and ML. The variances of GEE and OP increase with the variance of the intercept. The variance of the PQL estimator at $\sigma = 0.7$ was assessed via Monte Carlo simulations.

23 Deterministic versus stochastic asymptotics

Statisticians spend much time on asymptotic properties of estimators. The outcome of this multi-page research does not vary much from paper to paper: the estimator is consistent and asymptotically normally distributed.

For example, in the case of nonlinear regression $y_i = f(\boldsymbol{\theta}; \mathbf{x}_i) + \varepsilon_i$, the reason why the proofs are so long is because $\mathbf{x}_i$ is treated fixed (nonrandom, the **deterministic approach**).

Stochastic approach: $\mathbf{x}_i$ are random iid $\sim F$ (does not depend on $\boldsymbol{\theta}$). Then the model is treated conditional on $\mathbf{x}_i$ and $(y_i, \mathbf{x}_i)$ are iid with the full log-likelihood

$$\sum_{i=1}^{N} l(y_i|\mathbf{x}_i; \boldsymbol{\theta}) + \sum_{i=1}^{N} d(\mathbf{x}_i).$$

Since $d(\mathbf{x}_i) = F'(\mathbf{x}_i)$ does not depend on $\boldsymbol{\theta}$ the property of $\widehat{\boldsymbol{\theta}}$ does not depend on F. It implies that the

24 Asymptotics for LME, GLMM, NLM

Number of clusters goes to infinity, $N \to \theta$.

Two types of asymptotics:

1. $N \to \infty$ and max $n_i < \infty$: consistency for linear MM, inconsistency for nonlinear MM

2. $N \to \infty$ and min$_i$ $n \to \infty$: consistency for linear and nonlinear MM

25 Optimal designs with mixed model

There are at least three new issues with construction of optimal design for mixed models:

1. Since designs depend on variance parameters (such as matrix $\mathbf{D}$), optimal designs should be constructed in adaptive fashion (even for a linear model).

2. Optimal design may be constructed for population or individual models.

3. Two ways to create optimal designs: (1) to increase number of clusters or (2) to increase the number of observations per cluster, e.g.

$$cov(\widehat{\boldsymbol{\beta}}) = \sigma^2 \left[\sum_{i=1}^{N} \mathbf{X}_i'(\mathbf{I} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i')^{-1} \mathbf{X}_i \right]^{-1}$$

Example: the drug to reduce the blood pressure in longitudinal studies.

Question: to recruit more patients or a longer follow up?

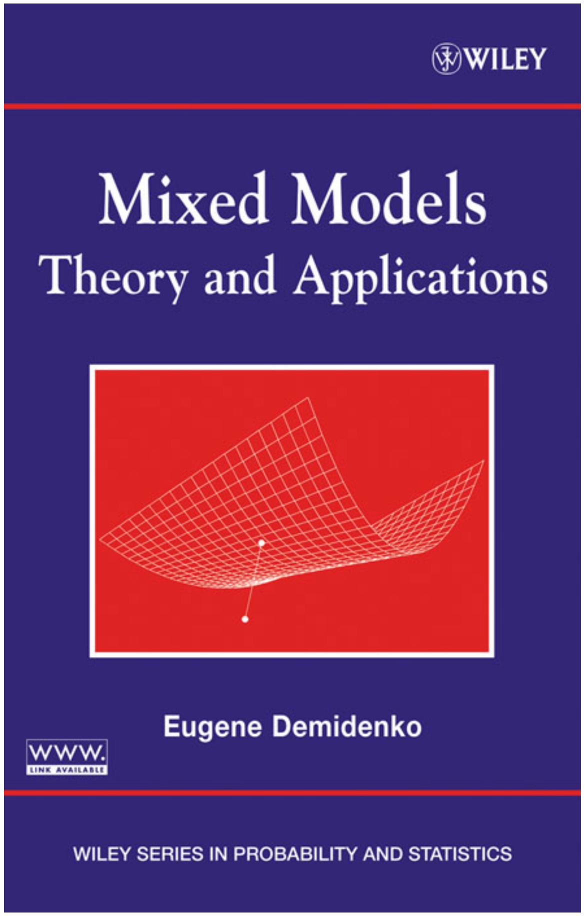


Figure 1: My book (2004, 700+ pages), the second edition is in progress.