How to fit generalised linear mixed models and keep smiling!

Patty Solomon

School of Mathematical Sciences
University of Adelaide

http://www.maths.adelaide.edu.au/patty.solomon/

1 December 2017

THE UNIVERSITY of ADELAIDE

Australian Government
Australian Research Council

Why am I fitting GLMMs?

To evaluate the performance of Australian and New Zealand intensive care units (ICUs)

ICU bedside area, The Queen Elizabeth Hospital SA

We took the 'best approach' to evaluating ICU performance

Courtesy of your very own Linda Sharples!

J. R. Statist. Soc. A (2007)
170, *Part* 4, *pp.* 865–890

A hierarchical modelling framework for identifying unusual performance in health care providers

David I. Ohlssen, Linda D. Sharples and David J. Spiegelhalter
Medical Research Council Biostatistics Unit, Cambridge, UK

Key idea: identify unusual mortality performance in three stages of analysis.

Three-stages for evaluating ICU performance

Stage 1: Fit *hierarchical logistic regression models* to mortality and identify *potentially* unusual ICUs

Stage 2: estimate a *null model* for 'usual performance'

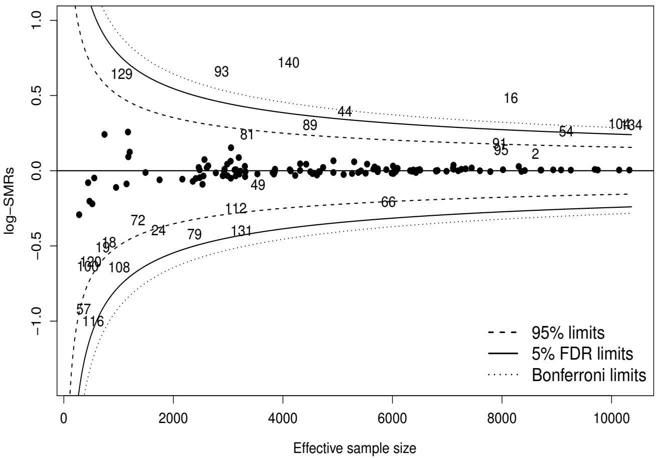
Stage 3: identify and visualise unusual ICUs.

Our approach is frequentist:

- Solomon *et al* *BMC Medical Research Methodology* 2014
- Kasza *et al* *Statistics in Medicine* 2013

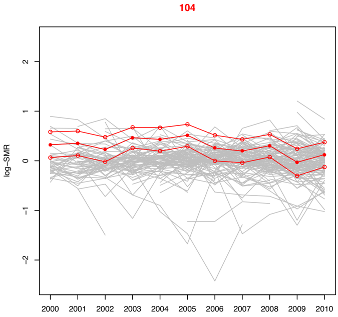
The first complete picture of ICU mortality in Australia

Log-SMRs versus effective sample size for Australian & NZ ICUs
2000 – 2010

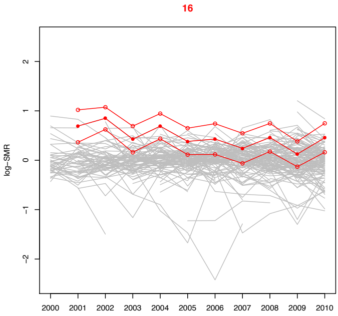


I would avoid Queensland ICUs . . .

QLD, Metropolitan

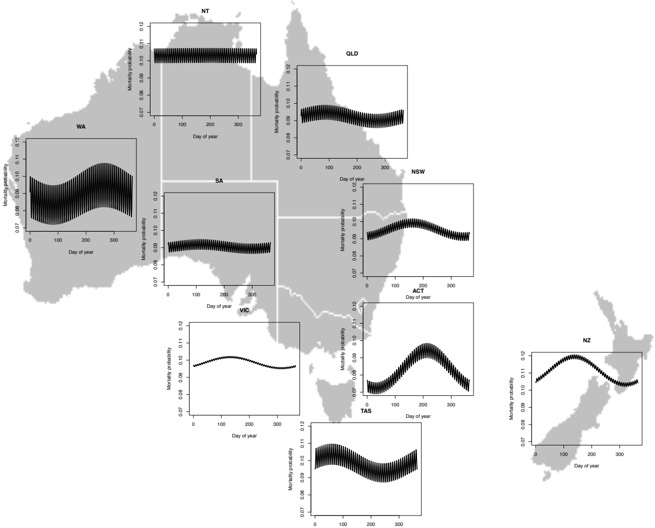


QLD, Private



We found seasonal differences in mortality for the first time

Annual and weekly cycles for respiratory patients



Things I learnt from this work . . .

- Comprehensive risk adjustment at both patient and hospital levels important $\Rightarrow$ **need complex models**.
- Accurate parameter estimates important $\Rightarrow$ **method of model fitting matters**.
- Waiting for the models to converge was like watching grass grow
  - took about as long
  - and was about as exciting.
- Data owners ANZICS CORE threatened 'grievous data withdrawal' if the media reported the results.

The Australian and New Zealand Intensive Care Society (ANZICS) Adult Patient Database (APD)

- One of the largest bi-national databases in the world
- Has collected voluntary patient admissions since 1995
- Currently $>$ 1.5m admissions; 167 of 214 eligible ICUs participated to 2013
- Data collected on *age, sex, diagnostic category, surgical status, ventilation status, hospital level, geographical locality, transfers, etc,* patient severity score APACHE III
- APACHE = Acute Physiology And Chronic Health Evaluation score (3rd revision): worst value in 24 hours post admission
- We use in-hospital mortality.

The database resembles a Swiss-cheese.

Snapshot of the ANZICS APD 2000 – 2010

Table 2 Characteristics of ANZICS APD study patients by year, 2000-2010

Hosp. admit year	*n* (%)	Hosp. mort. (%)	ICU mort. (%)	APIII mean (sd)	Age mean (sd)	Vent. (%)	Transfer (%)
2000	20,888 (4.0)	17.3	11.1	53.7 (30.9)	58.9 (19.3)	48.1	8.9
2001	26,353 (5.0)	15.8	10.1	52.6 (30.3)	59.6 (19.2)	44.0	9.6
2002	32,380 (6.2)	15.3	9.9	51.7 (29.7)	60.0 (18.9)	42.6	9.4
2003	37,082 (7.1)	14.4	9.2	51.5 (29.0)	60.4 (18.8)	41.0	9.1
2004	43,132 (8.2)	13.6	8.5	51.5 (28.4)	60.7 (18.6)	40.3	8.8
2005	49,093 (9.4)	12.9	8.2	50.9 (28.4)	60.6 (18.6)	40.1	8.8
2006	54,323 (10.4)	12.1	7.8	51.0 (28.2)	61.1 (18.8)	38.5	8.4
2007	57,187 (10.9)	12.0	7.8	51.0 (28.4)	61.0 (18.7)	37.6	8.5
2008	61,667 (11.8)	11.7	7.5	51.8 (28.7)	60.8 (18.8)	39.3	8.4
2009	67,015 (12.8)	11.3	7.3	51.8 (28.4)	60.8 (18.8)	39.3	8.4
2010	74,342 (14.2)	10.5	6.8	50.8 (27.6)	61.1 (18.8)	37.5	8.3

Mortality declined over the decade.

Data are hierarchical: Dataframe

ICU	patid	mortality	APACHEIII	variables
1	1	0	49	x_{11}
1	2	1	88	x_{12}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	n_1	0	59	x_{1n_1}
2	1	1	91	x_{21}
2	2	0	45	x_{22}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	n_2	0	94	x_{2n_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	1	1	49	x_{m1}
m	2	0	147	x_{m2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	n_m	1	57	x_{mn_m}

A random intercept and slope model for ICU mortality

$$Y_{ij} = \begin{cases} 1 & \text{if patient } j \text{ in ICU } i \text{ died in-hospital} \\ 0 & \text{alive at discharge} \end{cases}$$

- ICU mean β_0, $U_{i0} \sim N(0, \tau_0)$ ICU random effects.
- x_{ij} is the patient's APACHE III score
- APIII slope β_1, $U_{i1} \sim N(0, \tau_1)$ APIII random effects.
- Then

$$Y_{ij} | (x_{ij}, \boldsymbol{U}_i, \boldsymbol{\tau}) \sim \text{Bernoulli}(\pi_{ij})$$

- where

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + U_{i0} + x_{ij}(\beta_1 + U_{i1})$$

Similar to but not the same as, Zhang *et al Stats in Med* (2011)

More general models for ICU mortality

For explanatory variables $\boldsymbol{x}_{ij}$,

$$Y_{ij} | (\boldsymbol{x}_{ij}, \boldsymbol{U}_i, \boldsymbol{\tau}) \sim \text{Bernoulli}(\pi_{ij}) \quad \texttt{Model A}$$

For random structure with design vector $\boldsymbol{z}_{ij}$

$$\text{logit}(\pi_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{U}_i \qquad \texttt{Models B,C}$$

$$+ \qquad U_{it} \sim N(0, \tau_2)$$

For our simple model,

$$\boldsymbol{x}_{ij}^T = (1, x_{ij}) = \boldsymbol{z}_{ij}^T$$

We began with R: results using `glmer` (**Laplace**)

Models A,B,C fitted to the ANZICS APD 2000 – 2010
522, 911 patients from 144 ICUs

M	Fixed effects	Random effects	Levels	Completion	Time **hours**
A	67	int + APIII slope	2	Without error	8.5
B	67	A + ICU-years	3	Failed to converge†	17.8
C	133	B	3	Aborted by user	$>$ 177.0

†produced poor estimates

Three key issues when fitting GLMMs

- **Accurate parameter estimation**
- **Computing time**
- **Model selection**

We want procedures that will provide accurate estimates of the parameters of interest in a timely manner.

Parameter estimation: the profiled likelihood

- Use maximum likelihood
- Need to marginalise over the $\boldsymbol{U}_i$s
- Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau})$ be the unknown parameter vector
- the *profiled likelihood*

$$L_p(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \phi(\boldsymbol{u}_i; \boldsymbol{\tau}) d\boldsymbol{u}_i$$

where

$$\phi(\boldsymbol{u}_i; \boldsymbol{\tau}) \sim MVN(\boldsymbol{0}, \boldsymbol{\tau}).$$

Parameter estimation: two approaches

Let $\log L_p = \ell_p$. We want

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_p$$

1. Gold standard: approximate ℓ_p then maximise
   - Adaptive Gaussian Hermite quadrature (aGHQ)
2. Linearisation of the model
   - Laplace approximation
   - This is the same as aGHQ when $Q = 1$.

Parameter estimation, $\boldsymbol{\theta}$

We can write

$$
\begin{aligned}
\ell_p(\boldsymbol{\theta}; \boldsymbol{y}) &= a(\boldsymbol{\theta}) + \sum_{i=1}^{m} \log \int g(\boldsymbol{\theta}, \boldsymbol{u}_i)\, d\boldsymbol{u}_i \\
&\approx a(\boldsymbol{\theta}) + \sum_{i=1}^{m} \log G_i^{(q)}(\boldsymbol{\theta}) \\
&= a(\boldsymbol{\theta}) + b^{(q)}(\boldsymbol{\theta})
\end{aligned}
$$

Then

- ***Estimate*** $b^{(q)}(\boldsymbol{\theta})$
- ***Optimise*** to obtain *argmax*$_{\boldsymbol{\theta}}$; then
- ***Iterate*** until a minimum change threshold is met.

Software (available at our institution) evaluated

Software/package	Routine/function
Stata	`melogit, meqrlogit (xtmelogit)`
SAS	`NLMIXED, GLIMMIX`
R/lme4	`glmer`
ADMB*	`ADMB-RE`
R/glmmADMB	`glmmADMB`
S-Plus	`nlme`
Matlab	`fitglme`
SPSS	`GENLINMIXED`

*Automatic Differentiation Model Builder project

http://www.admb-project.org

Fake data: random intercept and slope model

$$Y_{ij} | (x_{ij}, \boldsymbol{U}_i, \boldsymbol{\tau}) \sim \text{Bernoulli}(\pi_{ij})$$

$$\texttt{logit}(\pi_{ij}) = \beta_0 + U_{i0} + x_{ij}(\beta_1 + U_{i1})$$

for $i = 1, 2, \ldots, 500$ and $j = 1, 2, \ldots$.

Assuming:

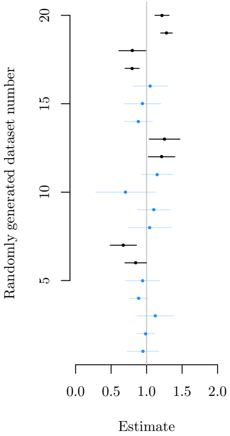
- $\beta_0 = \beta_1 = 1$, $x_{ij} \sim N(0, 1)$
- $\boldsymbol{U}_i = \begin{bmatrix} U_{i0} \\ U_{i1} \end{bmatrix} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\tau})$
    - $\boldsymbol{\tau} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$

1. Parameter accuracy

Results are presented as 'spine plots'

- 1,000 datasets were randomly generated
- 95% confidence interval for each dataset given a horizontal line
- Spine is true value which should be covered by 95% of CIs
- Horizontal lines that do not cover the true value are **black**.

Estimate the Type I error, α.

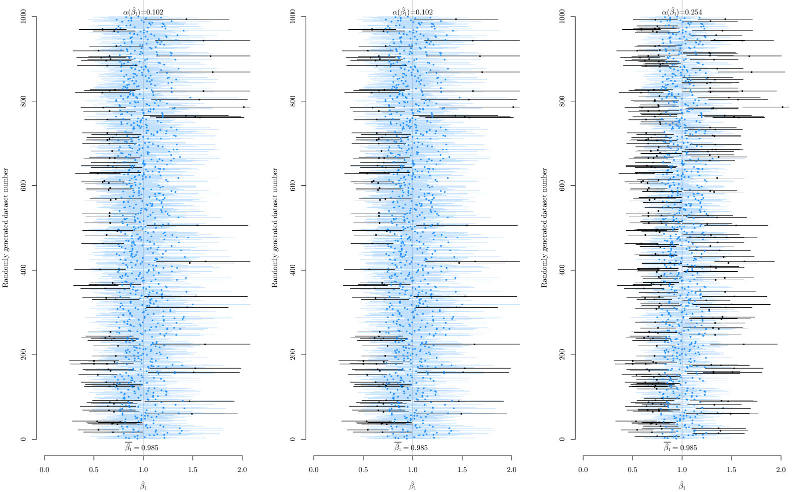


Estimation of β_1 using Laplace, $n_i = 3$

$\beta_1 = 1$	glmer	glmmADMB	ADMB	GLIMMIX	meqrlogit	fitglme	GENLINMIXED
$\alpha\left(\hat{\beta}_1\right)$	0.182	0.102	0.102	0.102	0.102	0.254	1.000
$\hat{\beta}_1$	0.982	0.985	0.985	0.985	0.985	0.985	0.454

Estimation of β_1 using Laplace $n_i = 3$

$\beta_1 = 1$	glmer	glmmADMB	ADMB	GLIMMIX	meqrlogit	fitglme	GENLINMIXED
$\alpha\left(\hat{\beta}_1\right)$	0.182	0.102	0.102	0.102	0.102	0.254	1.000
$\bar{\hat{\beta}}_1$	0.982	0.985	0.985	0.985	0.985	0.985	0.454

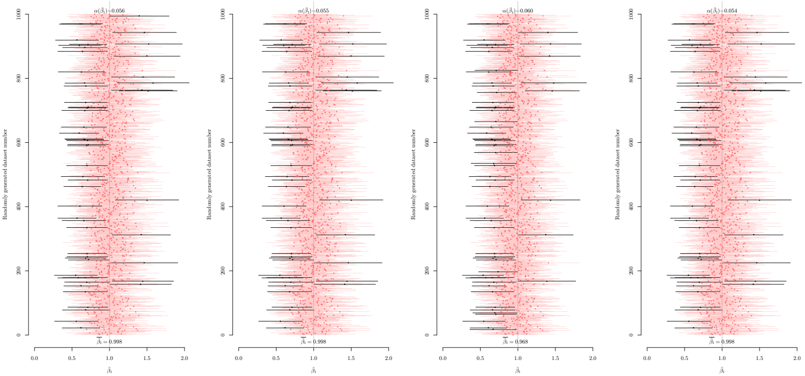


Estimation of β_1: PQL

$\beta_1 = 1$	glmer	glmmADMB	ADMB	GLIMMIX	meqrlogit	fitglme	GENLINMIXED
$\alpha\left(\hat{\beta}_1\right)$	0.182	0.102	0.102	0.102	0.102	0.254	1.000
$\bar{\hat{\beta}}_1$	0.982	0.985	0.985	0.985	0.985	0.985	0.454

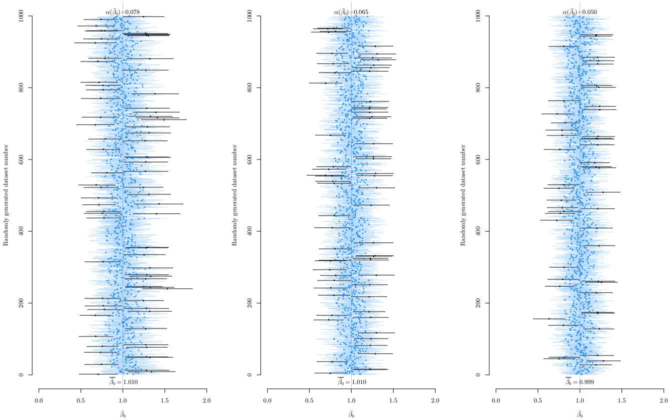
Estimation of β_1 using aGHQ=7, $n_i = 3$

$\beta_1 = 1$	ADMB	GLIMMIX	NLMIXED	meqrlogit
$\alpha\left(\hat{\beta}_1\right)$	0.056	0.055	0.060	0.054
$\overline{\hat{\beta}}_1$	0.998	0.998	0.968	0.998



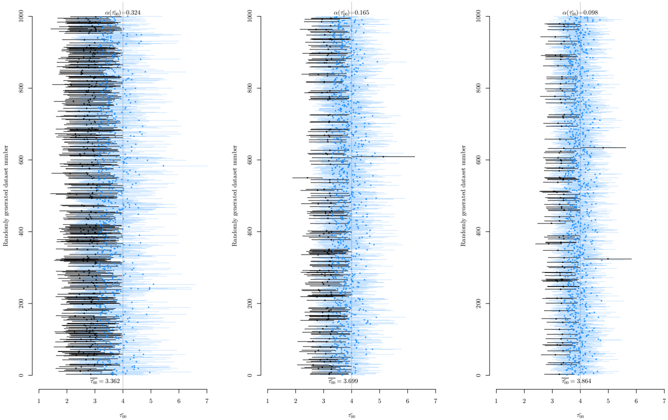
Fixed effects in GLIMMIX using Laplace: increasing n_i

$\beta_0 = 1$	$n_i = 5$	$n_i = 10$	$n_i = 25$
$\alpha(\hat{\beta}_0)$	0.078	0.065	0.050
$\bar{\hat{\beta}}_0$	1.010	1.010	0.999



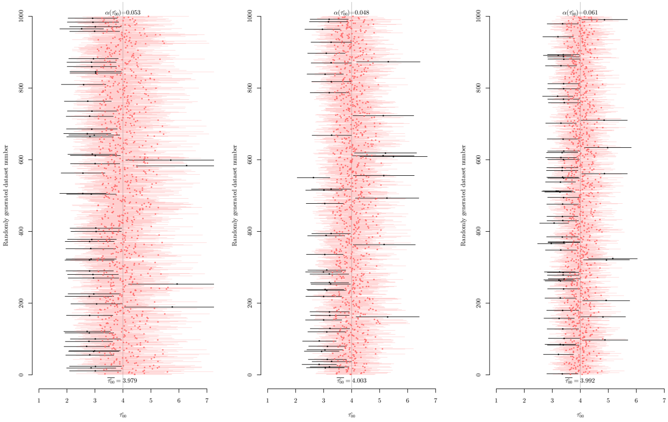
Variance components in GLIMMIX & Laplace: increasing n_i

$\tau_0 = 4$	$n_i = 5$	$n_i = 10$	$n_i = 25$
$\alpha(\hat{\tau}_0)$	0.324	0.165	0.098
$\bar{\hat{\tau}}_0$	3.362	3.699	3.864



Variance components in GLIMMIX and $Q = 7$: increasing n_i

$\tau_0 = 4$	$n_i = 5$	$n_i = 10$	$n_i = 25$
$\alpha(\hat{\tau}_0)$	0.053	0.048	0.061
$\bar{\hat{\tau}}_0$	3.979	4.003	3.992



Variance components in Stata: increasing Q, $n_i = 3$

$\sqrt{\tau_1} = 2$	$Q = 1$	$Q = 2$	$Q = 3$	$Q = 4$	$Q = 5$	$Q = 6$	$Q = 7$
$\alpha\left(\sqrt{\hat{\tau}_1}\right)$	0.373	0.954	0.204	0.035	0.037	0.040	0.033
$\sqrt{\bar{\hat{\tau}}_1}$	1.671	1.456	1.757	1.949	1.927	2.033	1.990

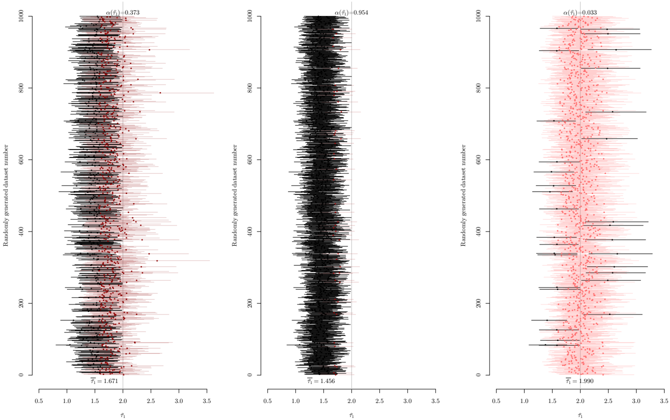


Illustration: behaviour of $Q = 1$

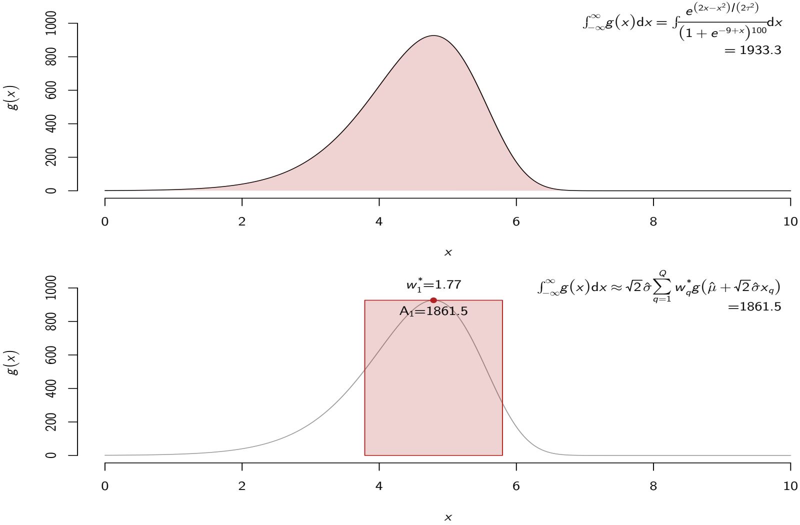
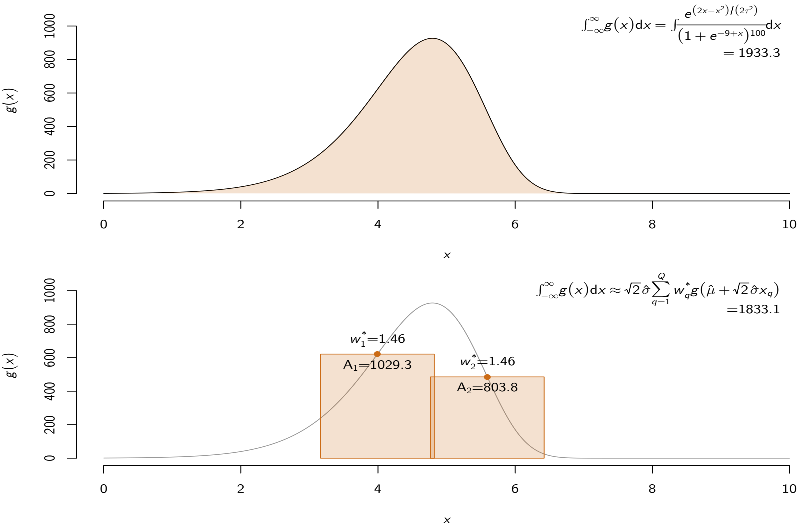
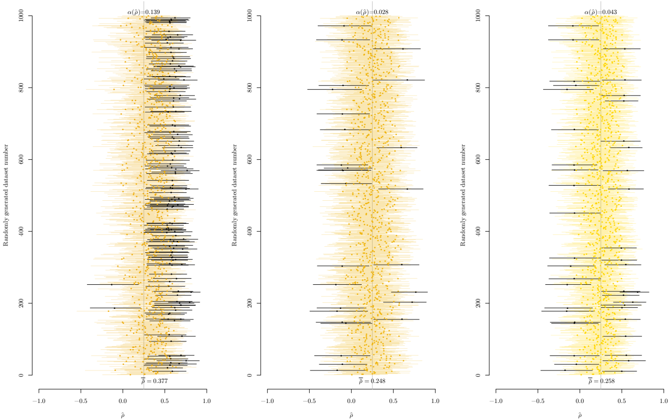


Illustration: behaviour of $Q = 2$

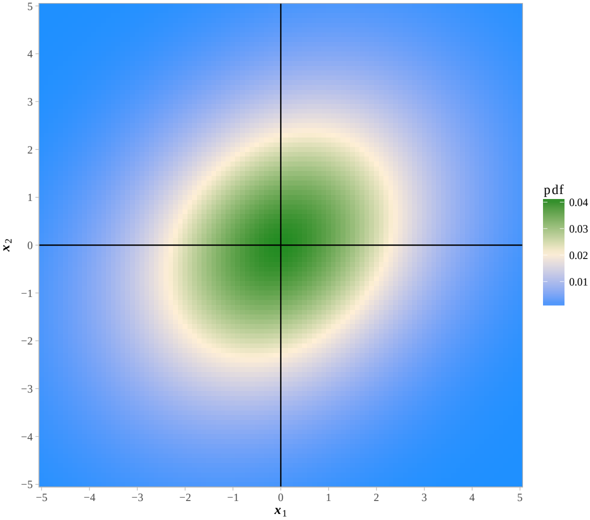


ρ estimates in Stata, $n_i = 3$

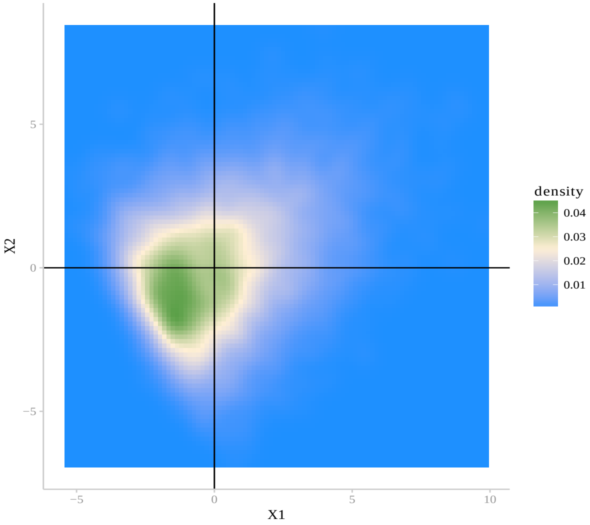
$\rho = 0.25$	$Q=1$	$Q=2$	$Q=3$	$Q=4$	$Q=5$	$Q=6$	$Q=7$
$\alpha(\hat{\rho})$	0.139	0.028	0.028	0.049	0.036	0.046	0.043
$\bar{\hat{\rho}}$	0.377	0.248	0.243	0.248	0.259	0.251	0.258



Simulation: bivariate normal with $\rho = 0.25$

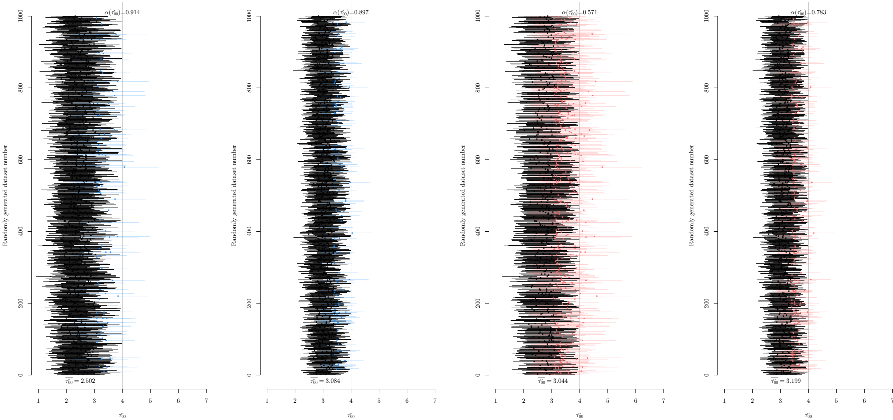


Simulation: skew-normal random effects



GLIMMIX Laplace and $Q = 7$ for τ_0, skew-normal

$\tau_0 = 4$	$n_i = 5$	$n_i = 20$	$n_i = 5$	$n_i = 20$
$\alpha(\hat{\tau}_0)$	0.914	0.897	0.571	0.783
$\bar{\hat{\tau}}_0$	2.502	3.084	3.044	3.199



2. Computing time in minutes

Extended random intercept and slope model, fake data:

$$\texttt{logit}\,(\pi_{ij}) = \overbrace{\beta_0 + \beta_1 x_{1ij}}^{\text{fixed}} + \overbrace{\sum_{k=2}^{21} \beta_k x_{kij}}^{\text{fixed/noise}} + \overbrace{U_{i0} + U_{i1} x_{1ij}}^{\text{random}}$$

for

$\beta_0 = \beta_1 = 1, \beta_2 = \ldots = \beta_{21} = 0$, $x_{kij} \sim N(0, 1)$, and

$i = 1, 2, \ldots, 200$, $j = 1, 2, \ldots, n_i$.

Computing time in minutes

Mac Pro (2010): 2 $\times$ 2.93GHz 6-Core Intel Xeon, 32GB DDR3, SSD

Method	Software	$n_i = 10$	$n_i = 100$	$n_i = 1000$
Laplace	`melogit`	1$^{\times}$	1	3
	`meqrlogit`	2	5	32
	`GLIMMIX`†	0	0	3
	`NLMIXED`††	2	21	187
	`ADMB-RE`	2	14	$\geq$4320$^{\times}$
	`glmer`	2	3	16
	`fitglme`	0	1	17

$^{\times}$ Required change to default convergence tolerance, otherwise repeated optimisation until maximum iterations

† Run on a virtual machine and not called from command line

†† Required starting values that were chosen at random

$^{\times}$ No result produced (reasons currently unknown)

Computing time in minutes

Method	Software	$n_i = 10$	$n_i = 100$	$n_i = 1000$
aGHQ ($Q = 7$)	`melogit`	0	0	342$^{\times}$
	`meqrlogit`	6	12	90
	`GLIMMIX`†	0	1	17
	`NLMIXED`††	18	203	4299
	`ADMB-RE`	2	17	$\geq$4320$^{\times}$

$^{\times}$ Gradient/convergence error

$^{\times}$ No result produced (reasons currently unknown)

Putting it all together: Models for ANZICS APD

Model fitting computing times (hours)

Software	Routine	Estimation	Model A	Model B	Model C
R	`glmer`	Laplace	8.5	17.8	>177
SAS	`GLIMMIX`	Laplace	1.9	4.1	††
		aGHQ (Q=7)	3.8	†††	†††
Stata	`melogit`	Laplace*	**2.7**	$>$ **24**	$>$ **72**
		aGHQ (Q=3)	0.09	0.19	0.32
		aGHQ (Q=5)	0.13	0.65	1.32
		aGHQ (Q=7)	0.18	$>$ **144****	$>$ **216**

†† Unable to fit because "obtaining MVQU estimates as starting values for the covariance parameters failed"

††† Unable to fit because "insufficient resources to perform adaptive quadrature with 7 quadrature points"

* **Not recommended**

** Memory loss 32GB - aborted on iteration 3

Some Don'ts and Dos

- **Never** use `SPSS` for glmms
- **Don't use**
  - aGHQ with $Q = 2$
  - `glmer` for models with more than a random intercept
  - `intmethod(laplace)` in `Stata`
  - Laplace for estimating variance components.

On a more positive note . . . **do** use

- aGHQ with $Q = 7$: gives reasonably accurate estimates
- GLIMMIX in `SAS`: fastest for aGHQ using simple models
- melogit in `Stata` for more complex models: $Q < 7$?
- Laplace for model selection with AIC.

Acknowledgements

Thank you to Dr Ty Stanford, now sadly (for me) working in the private sector.

THE UNIVERSITY of ADELAIDE

Much of the computation was made feasible using the command line parallel computing utility: `GNU Parallel`. Please see http://www.gnu.org/s/parallel or the *;login: The USENIX Magazine* article (O. Tange; 2011) for more details.

GNU parallel