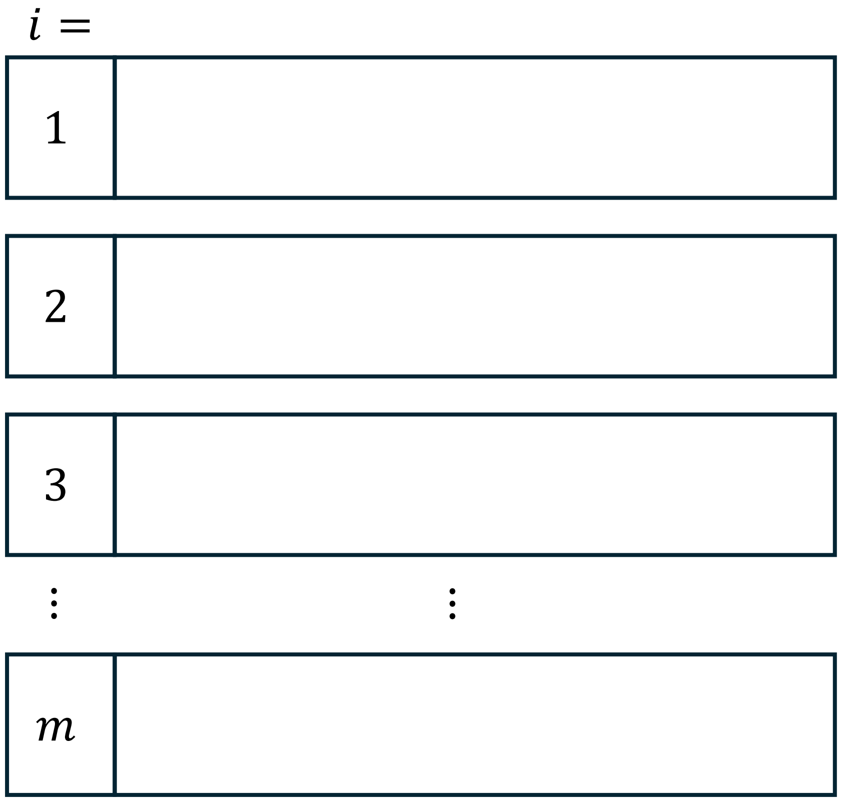
Asymptotics for Gaussian Variational Approximation in Generalised Linear Mixed Models

Nelson Chua[1,2], Francis Hui[1], Luca Maestrini[1], Alan Welsh[1]

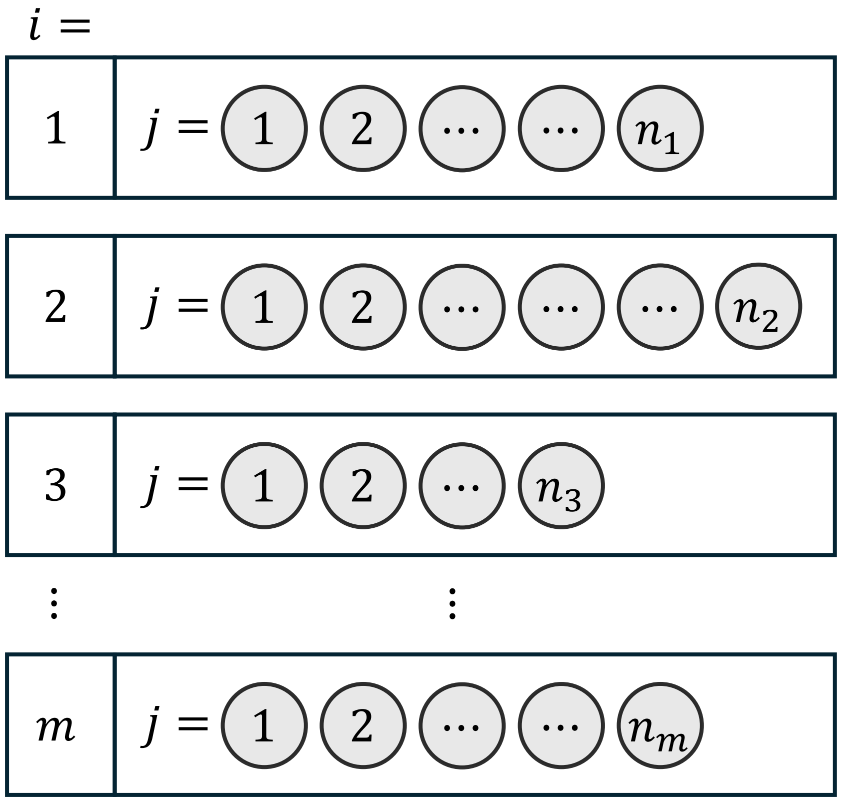
[1] *Australian National University, Canberra*

[2] *Australian Bureau of Statistics, Canberra*

Independent-cluster GLMM



Independent-cluster GLMM



Independent-cluster GLMM

Response variable

$$(y_{ij}|\boldsymbol{u}_i) \sim \exp\{y_{ij}\eta_{ij} - b(\eta_{ij}) + c(y_{ij})\}$$

$$\eta_{ij} = \boldsymbol{z}_{ij}^T(\boldsymbol{\beta}_0 + \boldsymbol{u}_i) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_1$$

Independent-cluster GLMM

Response variable

$$(y_{ij}|\boldsymbol{u}_i) \sim \exp\{y_{ij}\eta_{ij} - b(\eta_{ij}) + c(y_{ij})\}$$

$$\eta_{ij} = \boldsymbol{z}_{ij}^T(\boldsymbol{\beta}_0 + \boldsymbol{u}_i) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_1$$

Random effects

$$\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

d-dimensional

Independent-cluster GLMM

Response variable

$$(y_{ij}|\boldsymbol{u}_i) \sim \exp\{y_{ij}\eta_{ij} - b(\eta_{ij}) + c(y_{ij})\}$$

$$\eta_{ij} = \boldsymbol{z}_{ij}^T(\boldsymbol{\beta}_0 + \boldsymbol{u}_i) + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_1$$

Random effects

$$\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

d-dimensional

(Full) log-likelihood

$$\ell(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\Sigma}) = \sum_{i=1}^{m} \underbrace{\log \int_{\mathbb{R}^d} \left[\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{u}_i) f(\boldsymbol{u}_i) \right] d\boldsymbol{u}_i}_{\log f(\boldsymbol{y}_i)}$$

Variational approximation

For any density $q(\cdot)$, the following holds:

$$\underbrace{\log f(\boldsymbol{y}_i)}_{\substack{\text{Log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} = \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Variational log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i | \boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

Variational approximation

For any density $q(\cdot)$, the following holds:

$$\underbrace{\log f(\boldsymbol{y}_i)}_{\substack{\text{Log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} = \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Variational log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

Choose $q(\cdot)$ so that the variational log-likelihood is...	faster to compute than the log-likelihood	**and**	a good approximation of the log-likelihood

Variational approximation

For any density $q(\cdot)$, the following holds:

$$\underbrace{\log f(\boldsymbol{y}_i)}_{\substack{\text{Log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} = \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Variational log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

Avoid multidimensional integration

Choose $q(\cdot)$ so that the variational log-likelihood is…	faster to compute than the log-likelihood	and	a good approximation of the log-likelihood

Variational approximation

For any density $q(\cdot)$, the following holds:

$$\underbrace{\log f(\boldsymbol{y}_i)}_{\substack{\text{Log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} = \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Variational log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} + \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i | \boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Kullback-Leibler divergence} \\ \text{(Non-negative quantity)}}}$$

Avoid multidimensional integration

KL divergence near zero

Choose $q(\cdot)$ so that the variational log-likelihood is... faster to compute than the log-likelihood **and** a good approximation of the log-likelihood

Gaussian variational approximation (GVA)

$\boxed{\text{Choose } q(\boldsymbol{u}_i) \coloneqq q(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \text{ density}}$

"Variational parameters"
"Tuning parameters"

$$\underbrace{\log f(\boldsymbol{y}_i)}_{\substack{\text{Log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} = \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Variational log-likelihood} \\ \text{component for } \boldsymbol{y}_i}} + \underbrace{\int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i}_{\substack{\text{Kullback-Leibler divergence} \\ \text{(Non-negative quantity)}}}$$

Does choosing a Gaussian variational density...	avoid multidimensional integration	**and**	result in a KL divergence near zero?

Gaussian variational approximation (GVA)

Choose $q(\boldsymbol{u}_i) \coloneqq q(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$ density

"Variational parameters"
"Tuning parameters"

$$\log f(\boldsymbol{y}_i) = \int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

$$= -\sum_{j=1}^{n_i} \int_{\mathbb{R}} b\left(\eta_{ij} + \{\boldsymbol{z}_{ij}^T \boldsymbol{\Lambda}_i \boldsymbol{z}_{ij}\}^{1/2} s\right) \phi(s)\, ds$$

$$+\boldsymbol{y}_i^T \boldsymbol{\eta}_i + \frac{1}{2}\log\left|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}_i\right| - \frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}_i) + \mathrm{const.}$$

Ormerod and Wand (2012)

Does choosing a Gaussian variational density...	avoid multidimensional integration?	and	result in a KL divergence near zero?	**Yes!**

Gaussian variational approximation (GVA)

$\boxed{\text{Choose } q(\boldsymbol{u}_i) \coloneqq q(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \text{ density}}$

"Variational parameters"
"Tuning parameters"

$$\log f(\boldsymbol{y}_i) = \int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

$$N(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Lambda}}_i) \overset{?}{\approx} f(\boldsymbol{u}_i|\boldsymbol{y}_i)$$

Does choosing a Gaussian variational density…	avoid multidimensional integration	and	result in a KL divergence near zero?

Gaussian variational approximation (GVA)

$\boxed{\text{Choose } q(\boldsymbol{u}_i) \coloneqq q(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \text{ density}}$

"Variational parameters"
"Tuning parameters"

$$\log f(\boldsymbol{y}_i) = \int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

$$N(\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\Lambda}}_i) \overset{☺}{\approx} \underbrace{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}$$

Asymptotically normal
(*Bernstein-von Mises Theorem*)

Does choosing a Gaussian variational density...	avoid multidimensional integration	and	result in a KL divergence near zero?	**It seems promising!**

Gaussian variational approximation (GVA)

Choose $q(\boldsymbol{u}_i) := q(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$ density

"Variational parameters"
"Tuning parameters"

$$\log f(\boldsymbol{y}_i) = \int_{\mathbb{R}^d} \log\left\{\frac{f(\boldsymbol{y}_i, \boldsymbol{u}_i)}{q(\boldsymbol{u}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i + \int_{\mathbb{R}^d} \log\left\{\frac{q(\boldsymbol{u}_i)}{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}\right\} q(\boldsymbol{u}_i)\, d\boldsymbol{u}_i$$

$$N(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Lambda}}_i) \overset{☺}{\approx} \underbrace{f(\boldsymbol{u}_i|\boldsymbol{y}_i)}$$

Asymptotically normal
(*Bernstein-von Mises Theorem*)

$\widehat{\boldsymbol{\mu}}_i \approx \mathbb{E}[\boldsymbol{u}_i|\boldsymbol{y}_i] \quad \widehat{\boldsymbol{\Lambda}}_i \approx \mathrm{Var}[\boldsymbol{u}_i|\boldsymbol{y}_i]$

Useful for random effects estimation and inference?

Does choosing a Gaussian variational density...	avoid multidimensional integration	**and**	result in a KL divergence near zero?	**It seems promising!**

Using variational parameters for inference

Gaussian Variational Approximate Inference for **Generalized Linear Mixed Models**

Ormerod and Wand 2012

So, in summary, the maximizing variational parameters, $\widehat{\mu}$ and $\widehat{\Lambda}$, can be used for predicting the random effects and measuring their variability.

Variational Approximations for **Generalized Linear Latent Variable Models**

Hui, Warton et al. 2017

For example, $\hat{\boldsymbol{a}}$ serves as the variational version of both the empirical Bayes and maximum a-posteriori estimate of the smoothing coefficients, while $\hat{\boldsymbol{A}}$ is an estimate of the posterior covariance matrix. The multivariate normality of $h(\boldsymbol{\beta}|\boldsymbol{a}, \boldsymbol{A})$ also

Semiparametric Regression Using Variational Approximations

Hui, You et al. 2019

algorithm, as was seen in Section 3. In summary, the Gaussian VA approach quite naturally lends itself to the problem of predicting latent variables and constructing ordination plots, with $\hat{\boldsymbol{a}}_i$ can be used as the point predictions and $\hat{\boldsymbol{A}}_i$ can be used to construct prediction regions around these points.

Research question

Response variable

$$(y_{ij}|u_i) \sim \exp\{y_{ij}\eta_{ij} - b(\eta_{ij}) + c(y_{ij})\}$$

$$\eta_{ij} = \beta_0 + u_i + \beta_1 x_{ij}$$

Random effects

$$u_i \sim N(0, \sigma^2)$$

one-dimensional

Variational parameter estimates

$$\hat{\mu}_i \approx \mathbb{E}[u_i|\boldsymbol{y}_i] \quad \hat{\lambda}_i \approx \mathrm{Var}[u_i|\boldsymbol{y}_i]$$

Is $\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2}\hat{\lambda}_i^{1/2}\right]$ a good confidence interval for u_i?

Simulation setup

1000 simulated datasets

Bernoulli random intercept model

$$(y_{ij}|u_i) \sim Bern(e^{\eta_{ij}}/\{1 + e^{\eta_{ij}}\})$$

$$\eta_{ij} = \beta_0 + u_i + \beta_1 x_{ij} \quad x_{ij} \sim N(0,1)$$

$\beta_0 = -1$ $\beta_1 = 1$

Random effects

$$u_i \sim N(0, \sigma^2)$$

one-dimensional

Simulation setup

1000 simulated datasets

Bernoulli random intercept model

$$(y_{ij}|u_i) \sim Bern(e^{\eta_{ij}}/\{1+e^{\eta_{ij}}\})$$

$$\eta_{ij} = \beta_0 + u_i + \beta_1 x_{ij} \quad x_{ij} \sim N(0,1)$$

$\beta_0 = -1$ $\beta_1 = 1$

Random effects

$u_i \sim N(0, \sigma^2)$

one-dimensional

1000 sets of GVA estimates

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, \hat{\mu}_1, \hat{\lambda}_1, \hat{\mu}_2, \hat{\lambda}_2, \dots, \hat{\mu}_m, \hat{\lambda}_m)$$

Simulation setup

1000 simulated datasets

Bernoulli random intercept model

$$(y_{ij}|u_i) \sim Bern(e^{\eta_{ij}}/\{1+e^{\eta_{ij}}\})$$

$$\eta_{ij} = \beta_0 + u_i + \beta_1 x_{ij} \quad x_{ij} \sim N(0,1)$$

$\beta_0 = -1$ $\beta_1 = 1$

Random effects

$u_i \sim N(0, \sigma^2)$

one-dimensional

1000 sets of GVA estimates

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, \hat{\mu}_1, \hat{\lambda}_1, \hat{\mu}_2, \hat{\lambda}_2, \ldots, \hat{\mu}_m, \hat{\lambda}_m)$$

Empirical coverage probability

$\alpha = 0.1$

$$\frac{1}{1000} \cdot \# \left\{ u_1 \in \left[\hat{\mu}_1 \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_1^{1/2} \right] \right\}$$

Simulation setup

1000 simulated datasets

Bernoulli random intercept model

$$(y_{ij}|u_i) \sim Bern(e^{\eta_{ij}}/\{1+e^{\eta_{ij}}\})$$

$$\eta_{ij} = \beta_0 + u_i + \beta_1 x_{ij} \quad x_{ij} \sim N(0,1)$$

$\beta_0 = -1$ $\beta_1 = 1$

Random effects

$u_i \sim N(0, \sigma^2)$

one-dimensional

1000 sets of GVA estimates

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, \hat{\mu}_1, \hat{\lambda}_1, \hat{\mu}_2, \hat{\lambda}_2, \ldots, \hat{\mu}_m, \hat{\lambda}_m)$$

Empirical coverage probability

$\alpha = 0.1$

$$\frac{1}{1000} \cdot \#\left\{u_1 \in \left[\hat{\mu}_1 \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_1^{1/2}\right]\right\}$$

Repeat for different m (number of clusters) and $\bar{n}$ (average cluster size)

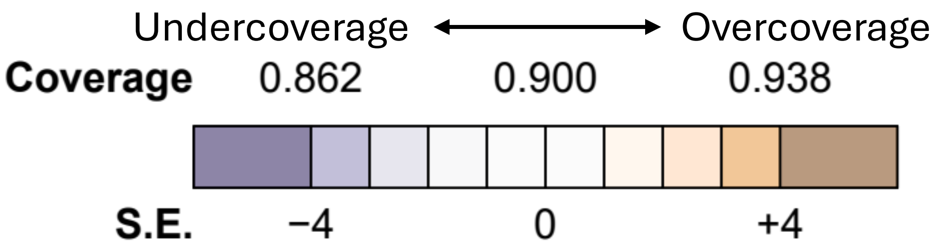
Coverage for u_1

Confidence interval: $\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2} \hat{\lambda}_1^{1/2}\right]$

Cluster size ($n_1 = \bar{n}$)

Number of clusters (m)

m \ n̄	10	20	40	80	160	320
10	0.708	0.746	0.736	0.669	0.570	0.487
20	0.806	0.824	0.813	0.762	0.695	0.602
40	0.868	0.868	0.872	0.832	0.799	0.723
80	0.887	0.882	0.876	0.861	0.839	0.805
160	0.889	0.880	0.890	0.869	0.868	0.847
320	0.895	0.882	0.893	0.878	0.883	0.863



Variational mean asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij} | u] \right)^{-1}$$

$$\hat{\mu}_i - u_i = \frac{1}{\sqrt{m}} \underbrace{\left[\frac{1}{\sqrt{m}} \sum_{k=1}^{m} u_k \right]}_{A} + \frac{1}{\sqrt{n_i}} \underbrace{\left[\frac{v(u_i)}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(y_{ij} - \mathbb{E}[y_{ij} | u_i] \right) \right]}_{B} + O_p\left(\max\{m^{-1}, n_i^{-1}\} \right)$$

Similar results seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)
Ning, Hui and Welsh 2025 (Penalised quasi-likelihood)

Variational mean asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij}|u] \right)^{-1}$$

$$\hat{\mu}_i - u_i = \frac{1}{\sqrt{m}} \underbrace{\left[\frac{1}{\sqrt{m}} \sum_{k=1}^{m} u_k \right]}_{A} + \frac{1}{\sqrt{n_i}} \underbrace{\left[\frac{v(u_i)}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(y_{ij} - \mathbb{E}[y_{ij}|u_i]\right) \right]}_{B} + O_p\left(\max\{m^{-1}, n_i^{-1}\}\right)$$

Similar results seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)
Ning, Hui and Welsh 2025 (Penalised quasi-likelihood)

Conditional on...	Asymptotic distributions		
	A	B	$\hat{\mu}_i - u_i$
All of $u_1, \ldots, u_m$			
Only u_i			
None of $u_1, \ldots, u_m$			

Variational mean asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij} | u] \right)^{-1}$$

$$\hat{\mu}_i - u_i = \frac{1}{\sqrt{m}} \underbrace{\left[\frac{1}{\sqrt{m}} \sum_{k=1}^{m} u_k \right]}_{A} + \frac{1}{\sqrt{n_i}} \underbrace{\left[\frac{v(u_i)}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(y_{ij} - \mathbb{E}[y_{ij} | u_i] \right) \right]}_{B} + O_p\left(\max\{m^{-1}, n_i^{-1}\} \right)$$

Similar results seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)
Ning, Hui and Welsh 2025 (Penalised quasi-likelihood)

Conditional on...	Asymptotic distributions		
	A	B	$\hat{\mu}_i - u_i$
All of $u_1, \dots, u_m$	Deterministic (and not necessarily zero)	$N(0, v(u_i))$	$N\left(\bar{u}, \frac{v(u_i)}{n_i} \right)$
Only u_i			
None of $u_1, \dots, u_m$			

Variational mean asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i}\sum_{j=1}^{n_i} \mathrm{var}[y_{ij}|u]\right)^{-1}$$

$$\hat{\mu}_i - u_i = \frac{1}{\sqrt{m}}\underbrace{\left[\frac{1}{\sqrt{m}}\sum_{k=1}^{m} u_k\right]}_{A} + \frac{1}{\sqrt{n_i}}\underbrace{\left[\frac{v(u_i)}{\sqrt{n_i}}\sum_{j=1}^{n_i}\left(y_{ij} - \mathbb{E}[y_{ij}|u_i]\right)\right]}_{B} + O_p\left(\max\{m^{-1}, n_i^{-1}\}\right)$$

Similar results seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)
Ning, Hui and Welsh 2025 (Penalised quasi-likelihood)

Conditional on...	Asymptotic distributions		
	A	B	$\hat{\mu}_i - u_i$
All of $u_1, \dots, u_m$	Deterministic (and not necessarily zero)	$N(0, v(u_i))$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$
Only u_i	$N(0, \sigma^2)$	$N(0, v(u_i))$	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$
None of $u_1, \dots, u_m$			

Variational mean asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions…

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij} | u] \right)^{-1}$$

$$\hat{\mu}_i - u_i = \frac{1}{\sqrt{m}} \underbrace{\left[\frac{1}{\sqrt{m}} \sum_{k=1}^{m} u_k \right]}_{A} + \frac{1}{\sqrt{n_i}} \underbrace{\left[\frac{v(u_i)}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(y_{ij} - \mathbb{E}[y_{ij} | u_i] \right) \right]}_{B} + O_p\left(\max\{m^{-1}, n_i^{-1}\}\right)$$

Similar results seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)
Ning, Hui and Welsh 2025 (Penalised quasi-likelihood)

Conditional on…	Asymptotic distributions		
	A	B	$\hat{\mu}_i - u_i$
All of $u_1, \dots, u_m$	Deterministic (and not necessarily zero)	$N(0, v(u_i))$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$
Only u_i	$N(0, \sigma^2)$	$N(0, v(u_i))$	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$
None of $u_1, \dots, u_m$	$N(0, \sigma^2)$	$\int N(0, v(z)) N(z; 0, \sigma^2)\, dz$	$\int N\left(0, \frac{v(z)}{n_i} + \frac{\sigma^2}{m}\right) N(z; 0, \sigma^2)\, dz$

Variational variance asymptotics

$$v(u) = \left(\frac{1}{n_i}\sum_{j=1}^{n_i} \mathrm{var}[y_{ij}|u]\right)^{-1}$$

Conditional on...	Asymptotic distribution of $\hat{\mu}_i - u_i$	Plug-in confidence interval	Targeted quantity
All of $u_1, \dots, u_m$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$		
Only u_i	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$		
None of $u_1, \dots, u_m$	$\int N\left(0, \frac{v(z)}{n_i} + \frac{\sigma^2}{m}\right) N(z; 0, \sigma^2)\, dz$		

Variational variance asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij}|u] \right)^{-1}$$

$$\hat{\lambda}_i = \frac{v(u_i)}{n_i} + O_p(n_i^{-3/2})$$

Similar result seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)

Conditional on...	Asymptotic distribution of $\hat{\mu}_i - u_i$	Plug-in confidence interval	Targeted quantity
All of $u_1, \dots, u_m$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$		
Only u_i	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$		
None of $u_1, \dots, u_m$	$\int N\left(0, \frac{v(z)}{n_i} + \frac{\sigma^2}{m}\right) N(z; 0, \sigma^2)\, dz$		

Variational variance asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij}|u] \right)^{-1}$$

$$\hat{\lambda}_i = \boxed{\frac{v(u_i)}{n_i}} + O_p(n_i^{-3/2})$$

Similar result seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)

Conditional on...	Asymptotic distribution of $\hat{\mu}_i - u_i$	Plug-in confidence interval	Targeted quantity
All of $u_1, \dots, u_m$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$	$\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_i^{1/2}\right]$	$u_i - \bar{u}$
Only u_i	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$		
None of $u_1, \dots, u_m$	$\int N\left(0, \frac{v(z)}{n_i} + \frac{\sigma^2}{m}\right) N(z; 0, \sigma^2)\, dz$		

Variational variance asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions...

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij} | u] \right)^{-1}$$

$$\hat{\lambda}_i = \boxed{\frac{v(u_i)}{n_i}} + O_p(n_i^{-3/2})$$

Similar result seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)

Conditional on...	Asymptotic distribution of $\hat{\mu}_i - u_i$	Plug-in confidence interval	Targeted quantity
All of $u_1, \ldots, u_m$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$	$\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_i^{1/2}\right]$	$u_i - \bar{u}$
Only u_i	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$	$\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} (\hat{\lambda}_i + \hat{\sigma}^2/m)^{1/2}\right]$	u_i
None of $u_1, \ldots, u_m$	$\int N\left(0, \frac{v(z)}{n_i} + \frac{\sigma^2}{m}\right) N(z; 0, \sigma^2)\, dz$		

Variational variance asymptotics

If $m, \bar{n} \to \infty$, then under certain regularity conditions…

$$v(u) = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \mathrm{var}[y_{ij} | u] \right)^{-1}$$

$$\hat{\lambda}_i = \boxed{\frac{v(u_i)}{n_i}} + O_p(n_i^{-3/2})$$

Similar result seen in:
Hall, Pham et al. 2011 (GVA with Poisson response)

Conditional on…	Asymptotic distribution of $\hat{\mu}_i - u_i$	Plug-in confidence interval	Targeted quantity
All of $u_1, \dots, u_m$	$N\left(\bar{u}, \frac{v(u_i)}{n_i}\right)$	$\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_i^{1/2}\right]$	$u_i - \bar{u}$
Only u_i	$N\left(0, \frac{v(u_i)}{n_i} + \frac{\sigma^2}{m}\right)$	$\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} (\hat{\lambda}_i + \hat{\sigma}^2/m)^{1/2}\right]$	u_i
None of $u_1, \dots, u_m$	$\int N\left(0, \frac{v(z)}{n_i} + \frac{\sigma^2}{m}\right) N(z; 0, \sigma^2)\, dz$	$[\hat{\mu}_i \pm \text{quantile of mixture dsn}]$	u_i

Revisiting coverage for u_1

Cluster size ($n_1 = \bar{n}$)

$\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2}\hat{\lambda}_1^{1/2}\right]$

$\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2}(\hat{\lambda}_1 + \hat{\sigma}^2/m)^{1/2}\right]$

$[\hat{\mu}_1 \pm \text{quantile of mixture dsn}]$

Number of clusters (m)

m \ $\bar{n}$	10	20	40	80	160	320
10	0.708	0.746	0.736	0.669	0.570	0.487
20	0.806	0.824	0.813	0.762	0.695	0.602
40	0.868	0.868	0.872	0.832	0.799	0.723
80	0.887	0.882	0.876	0.861	0.839	0.805
160	0.889	0.880	0.890	0.869	0.868	0.847
320	0.895	0.882	0.893	0.878	0.883	0.863

Coverage 0.862 0.900 0.938

S.E. −4 0 +4

Revisiting coverage for u_1

Cluster size ($n_1 = \bar{n}$)

Number of clusters (m)

$\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2}\hat{\lambda}_1^{1/2}\right]$

m \ $\bar{n}$	10	20	40	80	160	320
10	0.708	0.746	0.736	0.669	0.570	0.487
20	0.806	0.824	0.813	0.762	0.695	0.602
40	0.868	0.868	0.872	0.832	0.799	0.723
80	0.887	0.882	0.876	0.861	0.839	0.805
160	0.889	0.880	0.890	0.869	0.868	0.847
320	0.895	0.882	0.893	0.878	0.883	0.863

$\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2}(\hat{\lambda}_1 + \hat{\sigma}^2/m)^{1/2}\right]$

m \ $\bar{n}$	10	20	40	80	160	320
10	0.744	0.808	0.843	0.843	0.851	0.851
20	0.826	0.863	0.883	0.875	0.886	0.882
40	0.879	0.876	0.895	0.890	0.890	0.903
80	0.892	0.893	0.895	0.890	0.892	0.911
160	0.891	0.883	0.899	0.883	0.900	0.900
320	0.896	0.884	0.900	0.884	0.899	0.896

$[\hat{\mu}_1 \pm \text{quantile of mixture dsn}]$

Coverage 0.862 0.900 0.938

S.E. −4 0 +4

Revisiting coverage for u_1

Cluster size ($n_1 = \bar{n}$)

Number of clusters (m)

$\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2} \hat{\lambda}_1^{1/2}\right]$

m \ $\bar{n}$	10	20	40	80	160	320
10	0.708	0.746	0.736	0.669	0.570	0.487
20	0.806	0.824	0.813	0.762	0.695	0.602
40	0.868	0.868	0.872	0.832	0.799	0.723
80	0.887	0.882	0.876	0.861	0.839	0.805
160	0.889	0.880	0.890	0.869	0.868	0.847
320	0.895	0.882	0.893	0.878	0.883	0.863

$\left[\hat{\mu}_1 \pm \Phi^{-1}_{1-0.1/2} (\hat{\lambda}_1 + \hat{\sigma}^2/m)^{1/2}\right]$

m \ $\bar{n}$	10	20	40	80	160	320
10	0.744	0.808	0.843	0.843	0.851	0.851
20	0.826	0.863	0.883	0.875	0.886	0.882
40	0.879	0.876	0.895	0.890	0.890	0.903
80	0.892	0.893	0.895	0.890	0.892	0.911
160	0.891	0.883	0.899	0.883	0.900	0.900
320	0.896	0.884	0.900	0.884	0.899	0.896

$[\hat{\mu}_1 \pm \text{quantile of mixture dsn}]$

m \ $\bar{n}$	10	20	40	80	160	320
10	0.947	0.914	0.891	0.859	0.855	0.850
20	0.947	0.926	0.914	0.886	0.895	0.881
40	0.951	0.935	0.927	0.908	0.898	0.901
80	0.962	0.941	0.934	0.903	0.903	0.911
160	0.963	0.938	0.928	0.908	0.905	0.895
320	0.963	0.939	0.931	0.907	0.908	0.896

Coverage 0.862 0.900 0.938

S.E. −4 0 +4

Answer to research question*

Is $\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2}\hat{\lambda}_i^{1/2}\right]$ a good confidence interval for u_i?

It has good coverage if m dominates n_i

BUT

we do similarly or better by using the adjusted variance $\hat{\lambda}_i + \hat{\sigma}^2/m$

*In the context of a random intercept independent-cluster GLMM

Other findings*

Inference for...

Model parameters $\beta_0, \beta_1, \sigma^2$	GVA estimation is asymptotically fully efficient
Random effects u_i	`lme4` confidence interval is similar to $\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_i^{1/2}\right]$ `glmmTMB` confidence interval is similar to $\left[\hat{\mu}_i \pm \Phi^{-1}_{1-\alpha/2} (\hat{\lambda}_i + \hat{\sigma}^2/m)^{1/2}\right]$
Linear predictor $\eta_{ij} = \beta_0 + u_i + \beta_1 x_{ij}$	$\left[\hat{\eta}_{ij} \pm \Phi^{-1}_{1-\alpha/2} \hat{\lambda}_i\right]$ has good coverage for large enough m, n_i

*In the context of a random intercept independent-cluster GLMM

Thank you!

I am sponsored by the SSA PhD top-up scholarship.

Statistical Society of Australia