Variational Bayes Implementation
Summary of Changes to Original Algorithms

David Ewing

25 January 2026

Dear Dr John,

I hope you're well. Here is an update on the VB project . I've adapted your original algorithms across the four model files.

Starting Point: M2-Hierarchical-Linear-Diagnostic.Rmd

I began with your two core functions (normalmm.Gibbs and VB.mm) and wrapped them with diagnostic tools to fit the way I was making calls previously:

- VB.mm_with_history() tracks $E[\tau_e]$, $E[\tau_u]$, and ELBO per iteration so we can monitor convergence
- Wrapper functions handle 3-chain Gibbs initialisation ($\tau_e = 3/\tau_u = 0.5$, $\tau_e = 0.5/\tau_u = 3$, $\tau_e = 5/\tau_u = 5$)
- Added convergence plots, ELBO trajectories, and 8-panel posterior comparisons
- The underlying algorithm is unchanged from your code

Reference Check: M0-Hierarchical-Diagnostic.Rmd

I then removed these modified fuctions and replaced them with your original functions with no diagnostic modifications, so I can demostrate the same as the code you wrote for me. M2 produces almost identical results, so this serves as a reference check:

- Your exact normalmm.Gibbs() and VB.mm()
- Everything else is M2's diagnostic framework
- Same test conditions ($n = 300$, $\tau_e = 0.5$, $\tau_u = 1.0$, $\beta = (0.5, -2, 3)$)

Simplification: M1-Simple-Linear-Diagnostic.Rmd

This removes the random effects to show how your algorithm reduces to simple linear regression:

- No u parameters or Z matrix ($q = 0$)
- Only τ_e, no τ_u
- normalmm.linear.Gibbs_with_history() samples $\beta \mid \tau_e$ and $\tau_e \mid \beta$
- Uses the same synthetic setup as the other models

Extension to Logistic: M3-Hierarchical-Logistic-Diagnostic.Rmd

M3 extends M2 to binary responses, by

- $y \sim \text{Bernoulli}(\text{logit}^{-1}(X\beta + Zu))$ instead of normal linear
- VB uses Laplace approximation (iterative reweighted least squares) rather than conjugate updates
- Gibbs uses Metropolis-Hastings for $(\beta, u) \mid \tau_u, y$ instead of conjugate sampling
- No τ_e since binary responses have no residual variance
- Same diagnostic framework but by totaly different algorithms

Consistent Test Conditions

All files use the same setup:

- $n = 300$, $p = 3$ fixed effects
- $\beta_{\text{true}} = (0.5, -2, 3)$
- $\tau_{e,\text{true}} = 0.5$ (where applicable); $\tau_{u,\text{true}} = 1.0$
- Q scenarios: 5, 10, 20, 50, 100 groups
- Flat priors ($\alpha = 0$, $\gamma = 0$), seed 82171165
- Three-chain Gibbs initialisation following your pattern

All four files are now generating PDFs with diagnostic plots. To demonstrate under-dispersion in VB for variance components by comparing against your Gibbs sampler across different group sizes and model types.

Can you review and advise, particularly on how to handle the M3 algorithmic difference.

Best regards,

David Ewing