

Supplement to: Fast approximate inference for multivariate longitudinal data

DAVID M. HUGHES*, MARTA GARCÍA-FIÑANA

Department of Health Data Science, University of Liverpool, United Kingdom

dmhughes@liverpool.ac.uk

MATT P. WAND

School of Mathematical and Physical Sciences, University of Technology Sydney, Australia

SUMMARY

This document gives details of the derivation of the mean field variational Bayes (MFVB) approximation to a multivariate generalised linear model. The optimal form of each q -density is derived, and an expression for the log lower bound on the marginal log-likelihood is given. Additional simulation results are also provided.

Key words: Multivariate mixed models; Mean field variational Bayes; Generalised linear mixed model; Bayesian computing; Markov chain Monte Carlo; Repeated measurements.

*To whom correspondence should be addressed.

1. DERIVATION OF OPTIMAL q^* DENSITIES

To find optimal densities for each part of the factorisation of $q(\theta)$ we use the following result.

$$q_i^*(\theta_i) \propto \exp\{E_{-\theta_i}[\log p(\mathbf{y}, \theta)]\} \quad (1.1)$$

For the MGLMM presented in the main manuscript, we have

$$\begin{aligned} E_{-\theta_i}[\log p(\mathbf{y}, \theta)] &= E_{-\theta_i}[\log p(\mathbf{y}, |\beta, \mathbf{u}, \boldsymbol{\Sigma}_\varepsilon)] + E_{-\theta_i}[\log p(\beta, \mathbf{u} | \boldsymbol{\Sigma}_R)] + E_{-\theta_i}[\log p(\boldsymbol{\Sigma}_R | a_1, \dots, a_q)] \\ &\quad + \sum_{k=1}^q E_{-\theta_i}[\log p(a_k)] + \sum_{r=1}^{R_c} \{E_{-\theta_i}[\log p(\sigma_{\varepsilon_r}^2 | a_{\varepsilon_r})] + E_{-\theta_i}[\log p(a_{\varepsilon_r})]\} \end{aligned}$$

1.1 Derivation of $q^*(\sigma_{\varepsilon_r}^2)$

We require for each $r = 1, \dots, R_c$

$$\begin{aligned} E_{-\sigma_{\varepsilon_r}^2}[\log p(\mathbf{y}, \theta)] &= E_{-\sigma_{\varepsilon_r}^2}[\log p(\mathbf{y}, |\beta, \mathbf{u}, \boldsymbol{\Sigma}_\varepsilon)] + E_{-\sigma_{\varepsilon_r}^2}[\log p(\sigma_{\varepsilon_r}^2 | a_{\varepsilon_r})] + \text{const} \\ &= -\frac{1}{2}E_{-\sigma_{\varepsilon_r}^2}[\log |\boldsymbol{\Sigma}_\varepsilon|] - \frac{1}{2}E_{-\sigma_{\varepsilon_r}^2}[(\mathbf{y} - \mathbf{C}\bar{\nu})^T \boldsymbol{\Sigma}_\varepsilon^{-1}(\mathbf{y} - \mathbf{C}\bar{\nu})] - \frac{3}{2} \log \sigma_{\varepsilon_r}^2 \\ &\quad - \frac{1}{\sigma_{\varepsilon_r}^2} E_{-\sigma_{\varepsilon_r}^2}[1/a_{\varepsilon_r}] + \text{const} \\ &= - \left(\frac{\frac{1}{2}(\|\mathbf{y}_r - \mathbf{C}_r \mu_{q(\beta_r, \mathbf{u}_r)}\|^2 + \text{tr}[\mathbf{C}_r^T \mathbf{C}_r \boldsymbol{\Sigma}_{q(\beta_r, \mathbf{u}_r)})] + \mu_{q(1/a_{\varepsilon_r})}}{\sigma_{\varepsilon_r}^2} \right) \\ &\quad \left\{ -\frac{1}{2} \left(\sum_{i=1}^m n_{ir} + 1 \right) - 1 \right\} \log \sigma_{\varepsilon_r}^2 + \text{const} \end{aligned}$$

Here, the subscript r denotes the rows of the overall model relevant to the r th continuous longitudinal marker. Hence, $q^*(\sigma_{\varepsilon_r}^2)$ is an inverse-Gamma($1/2(\sum_{i=1}^m n_{ir} + 1)$, $B_{q(\sigma_{\varepsilon_r}^2)}$) distribution with

$$B_{q(\sigma_{\varepsilon_r}^2)} = \mu_{q(1/\sigma_{\varepsilon_r}^2)} + \frac{1}{2} \{ \|\mathbf{y}_r - \mathbf{C}_r \mu_{q(\beta_r, \mathbf{u}_r)}\|^2 + \text{tr}(\mathbf{C}_r^T \mathbf{C}_r \boldsymbol{\Sigma}_{q(\beta_r, \mathbf{u}_r)}) \}$$

1.2 Derivation of $q^*(a_{\varepsilon_r})$

We require for each $r = 1, \dots, R_c$,

$$\begin{aligned} E_{-a_{\varepsilon_r}}[\log p(\mathbf{y}, \theta)] &= E_{-a_{\varepsilon_r}}[\log p(\sigma_{\varepsilon_r}^2 | a_{\varepsilon_r})] + E_{-a_{\varepsilon_r}}[\log p(a_{\varepsilon_r})] + \text{const} \\ &= -2 \log a_{\varepsilon_r} - \frac{\left(E_{-a_{\varepsilon_r}}\left[\frac{1}{\sigma_{\varepsilon_r}^2}\right] + A_{\varepsilon_r}^{-2} \right)}{a_{\varepsilon_r}} + \text{const} \end{aligned}$$

Hence $q^*(a_{\varepsilon_r})$ is an inverse-Gamma($1, B_{q(a_{\varepsilon_r})}$) distribution with

$$B_{q(a_{\varepsilon_r})} = \mu_{q(1/\sigma_{\varepsilon_r}^2)} + A_{\varepsilon_r}^{-2}.$$

1.3 Derivation of $q^*(a_k)$

We require for each $k = 1, \dots, q$,

$$\begin{aligned} E_{-a_k}[\log p(\mathbf{y}, \theta)] &= E_{-a_k}[\log p(\boldsymbol{\Sigma}_R | a_1, \dots, a_q)] + E_{-a_k}[\log p(a_k)] + \text{const} \\ &= -\frac{(\nu + q + 2)}{2} \log(a_k) - \frac{E_{-a_k}[\boldsymbol{\Sigma}_R]_{kk} + A_k^{-2}}{a_k} + \text{const} \\ &= -\frac{(\nu + q + 2)}{2} \log(a_k) - \frac{\nu \mathbf{M}_{q(\boldsymbol{\Sigma}_R^{-1})_{kk}} + A_k^{-2}}{a_k} + \text{const} \end{aligned}$$

So $q^*(a_k)$ is an inverse-Gamma($1/2(\nu + q), B_{q(a_k)}$) with

$$B_{q(a_k)} = \nu \mathbf{M}_{q(\boldsymbol{\Sigma}_R^{-1})_{kk}} + A_k^{-2}.$$

1.4 Derivation of $q^*(\Sigma_R)$

We require

$$\begin{aligned}
E_{-\Sigma_R}[\log p(\mathbf{y}, \theta)] &= E_{-\Sigma_R}[\log p(\mathbf{u}|\Sigma_R)] + E_{-\Sigma_R}[\log p(\Sigma_R|a_1, \dots, a_q)] + \text{const} \\
&= -\frac{(m + \nu + 2q)}{2} \log |\Sigma_R| - \frac{1}{2} E_{-\Sigma_R}[\mathbf{u}^T (I_m \otimes \Sigma_R^{-1}) \mathbf{u}] \\
&\quad - E_{-\Sigma_R}[\frac{1}{2} \text{tr}(2\nu \text{diag}\{1/a_1, \dots, 1/a_q\} \Sigma_R^{-1})] + \text{const} \\
&= -\frac{1}{2} \text{tr} \left(\left\{ \sum_{i=1}^m (\mu_{q(u_i)} \mu_{q(u_i)}^T + \Sigma_{q(u_i)}) + 2\nu \text{diag}(\mu_{q(1/a_1)}, \dots, \mu_{q(1/a_q)}) \right\} \Sigma_R^{-1} \right) \\
&\quad - \frac{(m + \nu + 2q)}{2} \log |\Sigma_R| + \text{const}
\end{aligned}$$

Hence $q^*(\Sigma_R)$ is an inverse-Wishart($\nu + q + m - 1, B_{q(\Sigma_R)}$) with

$$B_{q(\Sigma_R)} = \sum_{i=1}^m (\mu_{q(u_i)} \mu_{q(u_i)}^T + \Sigma_{q(u_i)}) + 2\nu \text{diag}(\mu_{q(1/a_1)}, \dots, \mu_{q(1/a_q)})$$

1.5 Derivation of $q^*(\beta, \mathbf{u})$

When all of the longitudinal markers are continuous (and assumed Gaussian), it is relatively trivial to show the $q^*(\beta, \mathbf{u})$ is multivariate normal. However, once any of the longitudinal markers are binary or counts (assumed to be Poisson), we can no longer obtain a nice distribution for $q^*(\beta, \mathbf{u})$. Instead we take the approach of Rohde and Wand (2016) and specify that

$$q^*(\beta, \mathbf{u}, \mu_{q(\beta, \mathbf{u})}, \Sigma_{q(\beta, \mathbf{u})}) \sim N(\mu_{q(\beta, \mathbf{u})}, \Sigma_{q(\beta, \mathbf{u})})$$

in order to gain tractability. By result 2 of Rohde and Wand (2016), the necessary updates for $\mu_{q(\beta, \mathbf{u})}$ and $\Sigma_{q(\beta, \mathbf{u})}$ are

$$\begin{aligned}
\nu_{q(\beta, \mathbf{u})} &\leftarrow D_{\mu_{q(\beta, \mathbf{u})}} \text{Non Entropy}(q, \mu_{q(\beta, \mathbf{u})}, \Sigma_{q(\beta, \mathbf{u})})^T \\
\Sigma_{q(\beta, \mathbf{u})} &\leftarrow - \{ H_{\mu_{q(\beta, \mathbf{u})}} \text{Non Entropy}(q, \mu_{q(\beta, \mathbf{u})}, \Sigma_{q(\beta, \mathbf{u})}) \}^{-1} \\
\mu_{q(\beta, \mathbf{u})} &\leftarrow \mu_{q(\beta, \mathbf{u})} + \Sigma_{q(\beta, \mathbf{u})} \nu_{q(\beta, \mathbf{u})}
\end{aligned}$$

where $D_x f$ is the derivate vector of f with respect to x and $H_x f$ is the Hessian matrix of f , the second derivative with respect to x , and

$$\begin{aligned}
\text{Non Entropy}(q, \mu_{q(\beta, \mathbf{u})}, \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}) &= E_q [\log p(\mathbf{y}, |\beta, \mathbf{u}, \boldsymbol{\Sigma}_\varepsilon)] + E_q [\log p(\beta, \mathbf{u} | \boldsymbol{\Sigma}_R)] \\
&= \mathbf{y}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] \mathbf{C} \mu_{q(\beta, \mathbf{u})} - \mathbf{1}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] E_q [b(\mathbf{C}\tilde{\nu})] + \mathbf{1}^T E_q [c(\mathbf{y}, \phi)] \\
&\quad - \frac{(p + mq)}{2} \log 2\pi - \frac{1}{2} E_q \left[\log \left| \begin{array}{cc} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{array} \right| \right] \\
&\quad - \frac{1}{2} \left\{ \mu_{q(\beta, \mathbf{u})}^T \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \mu_{q(\beta, \mathbf{u})} \right. \\
&\quad \left. + \text{tr} \left(\begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})} \right) \right\} \tag{1.2}
\end{aligned}$$

in our case. See Rohde and Wand (2016) for a more general definition of non-entropy. In the fol-

lowing, we will make use of a change of variable. Note that by our specification, $\mathbf{C}\tilde{\nu} \sim N(\mathbf{C}\mu_{q(\beta, \mathbf{u})}, \mathbf{C}\boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}\mathbf{C}^T) =$

$N(\mu, \sigma^2)$. Then we define a new variable, $x = (\mathbf{C}\tilde{\nu} - \mu)/\sigma$, and we have that $x \sim N(0, 1)$. Then,

from 1.2, and using rules 3.3.6 and 3.3.2 of Wand (2002) we have that,

$$\begin{aligned}
D_{\mu_{q(\beta, \mathbf{u})}} \text{Non Entropy}(q, \mu_{q(\beta, \mathbf{u})}, \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}) &= (\mathbf{y} - D_{\mu_{q(\beta, \mathbf{u})}} \{E_q [b(\mathbf{C}\tilde{\nu})]\})^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] \mathbf{C} \\
&\quad - \mu_{q(\beta, \mathbf{u})}^T \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \\
&= (\mathbf{y} - E_x [b'(\sigma x + \mu)])^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] \mathbf{C} - \mu_{q(\beta, \mathbf{u})}^T \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
H_{\mu_{q(\beta, \mathbf{u})}} \text{Non Entropy}(q, \mu_{q(\beta, \mathbf{u})}, \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}) &= D_{\mu_{q(\beta, \mathbf{u})}} \{D_{\mu_{q(\beta, \mathbf{u})}} \text{Non Entropy}(q, \mu_{q(\beta, \mathbf{u})}, \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})})^T\} \\
&= D_{\mu_{q(\beta, \mathbf{u})}} \{ \mathbf{C}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] (\mathbf{y} - E_x [b'(\sigma x + \mu)]) \\
&\quad - \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \mu_{q(\beta, \mathbf{u})} \} \\
&= -\mathbf{C}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] D_{\mu_{q(\beta, \mathbf{u})}} \{E_x [b'(\sigma x + \mu)]\} - \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \\
&= -\mathbf{C}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] E_x [\text{diag}(b''(\sigma x + \mu))] \mathbf{C} - \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix}
\end{aligned}$$

Hence the Rohde and Wand updates for $q^*(\beta, \mathbf{u})$ in the MGLMM defined in our paper are given by

$$\begin{aligned}\nu_{q(\beta, \mathbf{u})} &\leftarrow \mathbf{C}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}](y - E_x[b'(\sigma x + \mu)]) - \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \mu_{q(\beta, \mathbf{u})} \\ \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})} &\leftarrow \left\{ \mathbf{C}^T E[\boldsymbol{\Sigma}_\varepsilon^{-1}] \text{diag} E_x[b''(\sigma x + \mu)] \mathbf{C} + \begin{bmatrix} \sigma_\beta^{-2} I_p & 0 \\ 0 & I_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \right\}^{-1} \\ \mu_{q(\beta, \mathbf{u})} &\leftarrow \mu_{q(\beta, \mathbf{u})} + \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})} \nu_{q(\beta, \mathbf{u})}\end{aligned}$$

It remains to define the cumulant functions and derivatives $b(\cdot)$, $b'(\cdot)$ and $b''(\cdot)$ for each of Gaussian, Poisson, and binary longitudinal markers. Note that the derivative $b'(\sigma x + \mu)$ is taken with respect to $(\sigma x + \mu)$. The chain rule is applied in the overall notation.

If \mathbf{Y}_r is a Gaussian marker, then $b(\sigma x + \mu) = 1/2(\sigma x + \mu)^2$, and so $E_x[b'(\sigma x + \mu)] = \sigma E_x[x] + \mu = \mu = \mathbf{C}\mu_{q(\beta, \mathbf{u})}$, and $E_x[b''(\sigma x + \mu)] = \mathbf{1}$.

If \mathbf{Y}_r is a Poisson marker, then $b(\sigma x + \mu) = b'(\sigma x + \mu) = b''(\sigma x + \mu) = e^{\sigma x + \mu}$ and so $E_x[b'(\sigma x + \mu)] = E_x[b''(\sigma x + \mu)] = \exp(\mathbf{C}\mu_{q(\beta, \mathbf{u})} + 1/2 \text{diagonal}(\mathbf{C}\boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}\mathbf{C}^T))$.

If \mathbf{Y}_r is a binary marker, then $b(\sigma x + \mu) = \log(1 + e^{\sigma x + \mu})$. Defining $\text{expit}(x) = \text{logit}^{-1}(x) = 1/(1 + \exp(-x))$, we have that $b'(\sigma x + \mu) = \text{expit}(\sigma x + \mu)$ and $b''(\sigma x + \mu) = b'(\sigma x + \mu)/(1 + \exp(\sigma x + \mu))$. It then follows that

$$E_x[b'(\sigma x + \mu)] = \int_{-\infty}^{\infty} \text{expit}(\sigma x + \mu) \phi(x) dx = \mathbf{B}_0(\mu, \sigma^2)$$

and with the help of integration by parts,

$$\begin{aligned}E_x[b''(\sigma x + \mu)] &= \int_{-\infty}^{\infty} \frac{b'(\sigma x + \mu)}{(1 + \exp(\sigma x + \mu))} \\ &= \frac{1}{\sigma} \int_{-\infty}^{\infty} x \text{expit}(\sigma x + \mu) \phi(x) dx \\ &= \mathbf{B}_1(\mu, \sigma^2)\end{aligned}$$

Following [Nolan and Wand \(2017\)](#) we avoid quadrature routines by approximating $\mathbf{B}_0(\mu, \sigma^2)$ and $\mathbf{B}_1(\mu, \sigma^2)$ using a weighted mixture of normal distribution using the Monahan-Stefanski weights

with $k = 8$ mixture components. Nolan and Wand (2017) state that

$$\mathbf{B}_0(\mu, \sigma^2) \approx \sum_{i=1}^k p_{k,i} \Phi \left(\frac{\mu s_{k,i}}{\sqrt{1 + \sigma^2 s_{k,i}^2}} \right)$$

$$\mathbf{B}_1(\mu, \sigma^2) \approx \sigma \sum_{i=1}^k \frac{p_{k,i} s_{k,i}}{\sqrt{1 + \sigma^2 s_{k,i}^2}} \phi \left(\frac{\mu s_{k,i}}{\sqrt{1 + \sigma^2 s_{k,i}^2}} \right)$$

In our case then the necessary MFVB updates are given by

$$\begin{aligned} \boldsymbol{\Omega} &\leftarrow \sqrt{\mathbf{1}_n \mathbf{1}_8^T + \mathbf{C} \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})} \mathbf{C}^T (\mathbf{s}^2)^T} \\ \tilde{\mathbf{B}}_0 &\leftarrow \Phi \left(\frac{\mathbf{C} \boldsymbol{\mu}_{q(\beta, \mathbf{u})} \mathbf{s}^T}{\boldsymbol{\Omega}} \mathbf{p} \right) \\ \tilde{\mathbf{B}}_1 &\leftarrow \left\{ \phi \left(\frac{\mathbf{C} \boldsymbol{\mu}_{q(\beta, \mathbf{u})} \mathbf{s}^T}{\boldsymbol{\Omega}} \right) / \boldsymbol{\Omega} \right\} \mathbf{p} \odot \mathbf{s} \end{aligned}$$

The updates for each marker can be substituted back into the Rohde and Wand updates using the relevant rows for each marker.

2. STREAMLINING

The updates for $\boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}$ involve calculating the inverse of a large matrix. However, this matrix has a block sparse structure, and we can utilise this structure, and matrix algebra results, in order to streamline our algorithm, and make substantial speed gains. This streamlining procedure is outlined in more detail in Lee and Wand (2016), and we just give the relevant details for our

model here. The inverse of $\Sigma_{q(\beta, \mathbf{u})}$ is arranged as follows.

$$\Sigma_{q(\beta, \mathbf{u})}^{-1} = \begin{bmatrix} F & G_1 & \cdots & \cdots & G_m \\ G_1^T & H_1^{-1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_m^T & 0 & \cdots & 0 & H_m^{-1} \end{bmatrix}$$

with $F = X^T E[\Sigma_{\varepsilon}^{-1}] \text{diag}\{E_x[b''(\sigma x + \mu)]\} X + \sigma_{\beta}^{-2} I_p$

$$G_i = X_i^T E[\Sigma_{\varepsilon_i}^{-1}] \text{diag}\{E_x[b''(\sigma x + \mu)]\}_i Z_i$$

$$H_i = \left\{ Z_i^T E[\Sigma_{\varepsilon_i}^{-1}] \text{diag}\{E_x[b''(\sigma x + \mu)]\}_i Z_i + M_{q(\Sigma_R^{-1})} \right\}$$

$$\mu = \mathbf{C} \mu_{q(\beta, \mathbf{u})}$$

$$\sigma^2 = \mathbf{C} \Sigma_{q(\beta, \mathbf{u})} \mathbf{C}^T$$

Using corollary 8.5.12 of [Harville \(2008\)](#), the inverse can be found as

$$\Sigma_{q(\beta, \mathbf{u})} = \begin{bmatrix} \Sigma_{q(\beta)} & \Lambda_{q(\beta, \mathbf{u})} \\ \Lambda_{q(\beta, \mathbf{u})}^T & \Sigma_{q(\mathbf{u})} \end{bmatrix}$$

where $\Sigma_{q(\beta)} = \left(F - \sum_{i=1}^m G_i H_i G_i^T \right)^{-1}$

$$\Lambda_{q(\beta, \mathbf{u})} = - [\Sigma_{q(\beta)} G_1 H_1, \dots, \Sigma_{q(\beta)} G_m H_m]$$

$$\Sigma_{q(\mathbf{u}_i)} = H_i + H_i G_i^T \Sigma_{q(\beta)} G_i H_i$$

The updates involving $\Sigma_{q(\beta, \mathbf{u})}$ can now be streamlined which largely involves many calculations being performed only on blocks of the data containing the rows for a particular individual and so substantially improves computation time. Specifically, simple algebraic manipulations result in the updates for $\mu_{q(\beta, \mathbf{u})}$ and $B_{q(\sigma_{\varepsilon_r}^2)}$ given in Algorithm 1 of our main manuscript. We can also streamline the calculation of $1/2 \log |\Sigma_{q(\beta, \mathbf{u})}|$ which occurs in the marginal log lower bound $\log \underline{p}(\mathbf{y}, q)$, using Theorem 13.3.8 of [Harville \(2008\)](#).

3. DERIVATION OF A LOWER BOUND ON THE MARGINAL LOG-LIKELIHOOD

We control the convergence of our MFVB algorithm through monitoring the bound on the marginal log-likelihood, $\log \underline{p}(\mathbf{y}, q)$. This is calculated as follows.

$$\begin{aligned}
\log \underline{p}(\mathbf{y}, q) &= E_q[\log p(\mathbf{y}, \theta) - \log q(\theta)] \\
&= E_q[\log p(\mathbf{y}_G | \beta, \mathbf{u}, \boldsymbol{\Sigma}_\varepsilon)] + E_q[\log p(\mathbf{y}_P | \beta, \mathbf{u})] + E_q[\log p(\mathbf{y}_B | \beta, \mathbf{u})] + E_q[\log p(\beta, \mathbf{u} | \boldsymbol{\Sigma}_R)] \\
&\quad + E_q[\log p(\boldsymbol{\Sigma}_R | a_1, \dots, a_q)] + \sum_{r=1}^{R_c} \{E_q[\log p(\sigma_{\varepsilon_r}^2 | a_{\varepsilon_r})] + E_q[\log p(a_{\varepsilon_r})]\} + \sum_{k=1}^q E_q[\log p(a_k)] \\
&\quad - E_q[\log q(\beta, \mathbf{u})] - E_q[\log q(\boldsymbol{\Sigma}_R)] - \sum_{r=1}^{R_c} \{E_q[\log q(\sigma_{\varepsilon_r}^2)] + E_q[\log q(a_{\varepsilon_r})]\} - \sum_{k=1}^q E_q[\log q(a_k)]
\end{aligned}$$

Here, we use \mathbf{y}_G , \mathbf{y}_P and \mathbf{y}_B to denote all the stacked Gaussian, Poisson and binary longitudinal observations respectively. The number of observations of each type of marker are denoted by n_G , n_P and n_B respectively. We also make use of matrices \mathbf{E}_1 , \mathbf{E}_2 and \mathbf{E}_3 as $n \times n$ diagonal matrices with value 1 for rows corresponding to Gaussian, Poisson and binary markers respectively and 0 elsewhere. We now provide an expression for each part of $\log \underline{p}(\mathbf{y}, q)$ and then state the full

expression.

$$\begin{aligned}
E_q[\log p(\mathbf{y}_G|\beta, \mathbf{u}, \boldsymbol{\Sigma}_\varepsilon)] &= -\frac{nG}{2} \log(2\pi) - \frac{1}{2} E_q[\log |\boldsymbol{\Sigma}_\varepsilon|] - \frac{1}{2} E_q[(\mathbf{y} - \mathbf{C}\tilde{\nu})^T \mathbf{E}_1^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{E}_1 (\mathbf{y} - \mathbf{C}\tilde{\nu})] \\
E_q[\log p(\mathbf{y}_P|\beta, \mathbf{u})] &= \mathbf{y}^T \mathbf{E}_2 \mathbf{C} \mu_{q(\beta, \mathbf{u})} - \mathbf{1}^T \mathbf{E}_2 \log \mathbf{y}! - \mathbf{1}^T \mathbf{E}_2 \exp(\mathbf{C} \mu_{q(\beta, \mathbf{u})} + 1/2 \text{diagonal}(\mathbf{C} \boldsymbol{\Sigma}_{q(\beta, \mathbf{u})} \mathbf{C}^T)) \\
E_q[\log p(\mathbf{y}_B|\beta, \mathbf{u})] &= \mathbf{y}^T \mathbf{E}_3 \mathbf{C} \mu_{q(\beta, \mathbf{u})} - E_q[\mathbf{1}^T \mathbf{E}_3 \log(1 + \exp(\mathbf{C}\tilde{\nu}))] \\
E_q[\log p(\beta, \mathbf{u}|\boldsymbol{\Sigma}_R)] &= \frac{-(p+mq)}{2} \log(2\pi) - \frac{1}{2} E_q \left[\log \begin{vmatrix} \sigma_\beta^{-2} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{vmatrix} \right] \\
&\quad - \frac{1}{2} E_q \left[\tilde{\nu}^T \begin{bmatrix} \sigma_\beta^{-2} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_m \otimes E[\boldsymbol{\Sigma}_R^{-1}] \end{bmatrix} \tilde{\nu} \right] \\
E_q[\log p(\boldsymbol{\Sigma}_R|a_1, \dots, a_q)] &= -\log C_{q, \nu+q-1} + \frac{\nu+q-1}{2} E_q[\log |2\nu \text{diag}\{1/a_1, \dots, 1/a_q\}|] \\
&\quad - \frac{\nu+2q}{2} E_q[\log |\boldsymbol{\Sigma}_R|] - \nu E_q[\text{tr}(\text{diag}\{1/a_1, \dots, 1/a_q\} \boldsymbol{\Sigma}_R^{-1})] \\
E_q[\log p(\sigma_{\varepsilon_r}^2|a_{\varepsilon_r})] &= -\frac{1}{2} E_q[\log(a_{\varepsilon_r})] - \log \Gamma(1/2) - \frac{3}{2} E_q[\log \sigma_{\varepsilon_r}^2] - E_q[1/a_{\varepsilon_r} \sigma_{\varepsilon_r}^2] \\
E_q[\log p(a_{\varepsilon_r})] &= \frac{1}{2} \log(A_{\varepsilon_r}^{-2}) - \log \Gamma(1/2) - \frac{3}{2} E_q[\log a_{\varepsilon_r}] - A_{\varepsilon_r}^{-2} E_q[1/a_{\varepsilon_r}] \\
E_q[\log p(a_k)] &= \frac{1}{2} \log(A_k^{-2}) - \log \Gamma(1/2) - \frac{3}{2} E_q[\log a_k] - A_k^{-2} E_q[1/a_k] \\
E_q[\log q(\beta, \mathbf{u})] &= \frac{-(p+mq)}{2} (1 + \log(2\pi)) - \frac{1}{2} \log |\boldsymbol{\Sigma}_{q(\beta, \mathbf{u})}| \\
E_q[\log q(\boldsymbol{\Sigma}_R)] &= -\log C_{q, \nu+q+m-1} + \frac{\nu+q+m-1}{2} E_q[\log |B_{q(\boldsymbol{\Sigma}_R)}|] - \frac{\nu+2q+m}{2} E_q[\log |\boldsymbol{\Sigma}_R|] \\
&\quad - \frac{1}{2} E_q[\text{tr}(B_{q(\boldsymbol{\Sigma}_R)} \boldsymbol{\Sigma}_R^{-1})] \\
E_q[\log q(\sigma_{\varepsilon_r}^2)] &= \frac{1}{2} \left(\sum_{i=1}^m n_{ir} + 1 \right) E_q[\log B_{q(\sigma_{\varepsilon_r}^2)}] - \log \Gamma(1/2 \left(\sum_{i=1}^m n_{ir} + 1 \right)) \\
&\quad - \left\{ \frac{1}{2} \left(\sum_{i=1}^m n_{ir} + 1 \right) + 1 \right\} E_q[\log \sigma_{\varepsilon_r}^2] - E_q \left[\frac{B_{q(\sigma_{\varepsilon_r}^2)}}{\sigma_{\varepsilon_r}^2} \right] \\
E_q[\log q(a_{\varepsilon_r})] &= E_q[\log B_{q(a_{\varepsilon_r})}] - \log \Gamma(1) - 2E_q[\log a_{\varepsilon_r}] - E_q \left[\frac{B_{q(a_{\varepsilon_r})}}{a_{\varepsilon_r}} \right] \\
E_q[\log q(a_k)] &= \frac{1}{2} (\nu+q) E_q[\log B_{q(a_k)}] - \log \Gamma(1/2(\nu+q)) - (1/2(\nu+q) + 1) E_q[\log a_k] \\
&\quad - E_q \left[\frac{B_{q(a_k)}}{a_k} \right]
\end{aligned}$$

In the above, $C_{d,A}$ is the normalising factor for the inverse-Wishart distribution given by

$$C_{d,A} = 2^{Ad/2} \pi^{d(d-1)/4} \prod_{i=1}^d \Gamma \left(\frac{A+1-i}{2} \right).$$

After evaluation of the expectations, and cancellation, we arrive at the expression for the lower bound on the marginal log-likelihood,

$$\begin{aligned}
\log \underline{p}(\mathbf{y}, q) &= \frac{p + mq}{2} - n_G \log 2\pi - \log C_{q, \nu+q-1} + \log C_{q, \nu+q+m-1} \\
&\quad - (R_c + q/2) \log \pi + q \log \Gamma(1/2(\nu + q)) - \frac{p}{2} \log(\sigma_\beta^2) + \frac{q(\nu + q + 1)}{2} \log 2\nu \\
&\quad + 1/2 \log |\Sigma_{q(\beta, \mathbf{u})}| + \mathbf{y}^T \mathbf{E}_2 \mathbf{C} \mu_{q(\beta, \mathbf{u})} - \mathbf{1}^T \mathbf{E}_2 \log \mathbf{y}! \\
&\quad - \mathbf{1}^T \mathbf{E}_2 \exp(\mathbf{C} \mu_{q(\beta, \mathbf{u})} + 1/2 \text{diagonal}(\mathbf{C} \Sigma_{q(\beta, \mathbf{u})} \mathbf{C}^T)) \\
&\quad + \mathbf{y}^T \mathbf{E}_3 \mathbf{C} \mu_{q(\beta, \mathbf{u})} - \mathbf{1}^T \mathbf{E}_3 \tilde{\mathbf{B}}_0 - \frac{(\nu + q + m - 1)}{2} \log |B_{q(\Sigma_R)}| - \frac{1}{2\sigma_\beta^2} (\mu_{q(\beta)}^T \mu_{q(\beta)} + \text{tr}(\Sigma_{q(\beta)})) \\
&\quad + \sum_{r=1}^{R_c} \left\{ \log \Gamma(1/2(\sum_{i=1}^m n_{ir} + 1)) - \log A_{\varepsilon r - 1/2(\sum_{i=1}^m n_{ir} + 1)} \log B_{q(\sigma_{\varepsilon r}^2) - B_{q(a_{\varepsilon r})} + \mu_{q(1/a_{\varepsilon r})} \mu_{q(1/\sigma_{\varepsilon r}^2)}} \right\} \\
&\quad + \sum_{k=1}^q \left\{ \nu \mathbf{M}_{q(\Sigma_R^{-1})_{kk}} \mu_{q(1/a_k)} - \log A_k - \frac{1}{2}(\nu + q) \log B_{q(a_k)} \right\}
\end{aligned}$$

4. ADDITIONAL SIMULATION DETAILS

The final part of this supplement presents some accuracy results from the simulation study. The main manuscript presents the accuracy for the case where there are $m = 100$ patients. In the supplement, Figures 4 and 4 report the accuracy for the simulations with $m = 1000$ and $m = 10,000$ individuals respectively.

REFERENCES

- HARVILLE, DAVID A. (2008). *Matrix algebra from a statistician's perspective*. Springer, New York.
- LEE, CATHY YUEN YI AND WAND, MATT P. (2016). Streamlined mean field variational bayes for longitudinal and multilevel data analysis. *Biometrical Journal* **58**(4), 868–895.
- NOLAN, TUI H AND WAND, MATT P. (2017). Accurate logistic variational message passing: algebraic and numerical details. *Stat* **6**(1), 102–112.

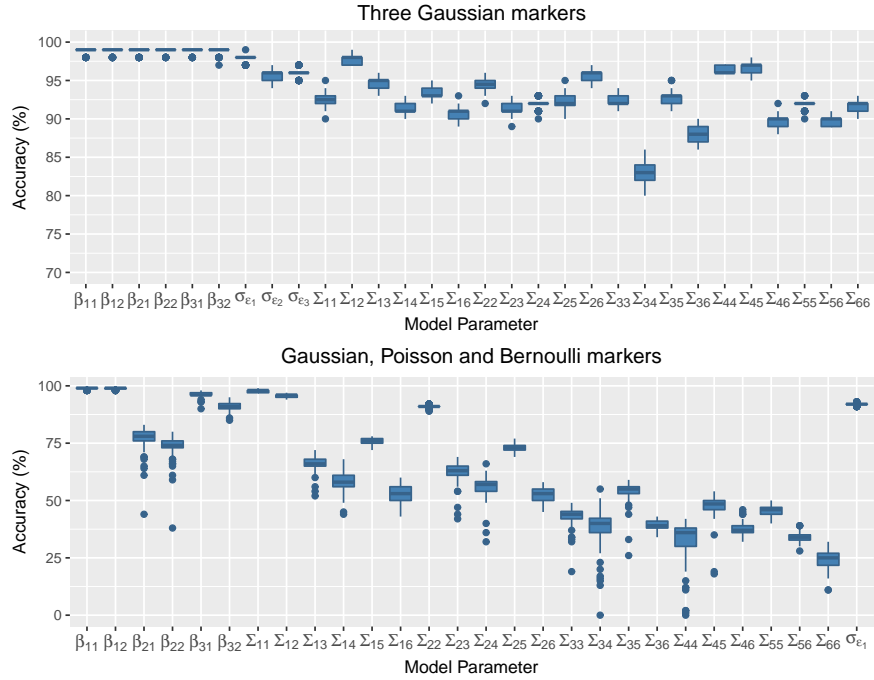


Fig. 1. Accuracy scores for mean field variational Bayes compared to MCMC for simulated datasets with 3 continuous longitudinal markers (top panel) and 3 types of markers (bottom panel) in the simulation with $m = 1000$ individuals

ROHDE, DAVID AND WAND, MATT P. (2016). Semiparametric mean field variational bayes:

General principles and numerical issues. *The Journal of Machine Learning Research* **17**(1), 5975–6021.

WAND, MP. (2002). Vector differential calculus in statistics. *The American Statistician* **56**(1),

55–62.

[Received August 1, 2010; revised October 1, 2010; accepted for publication November 1, 2010]

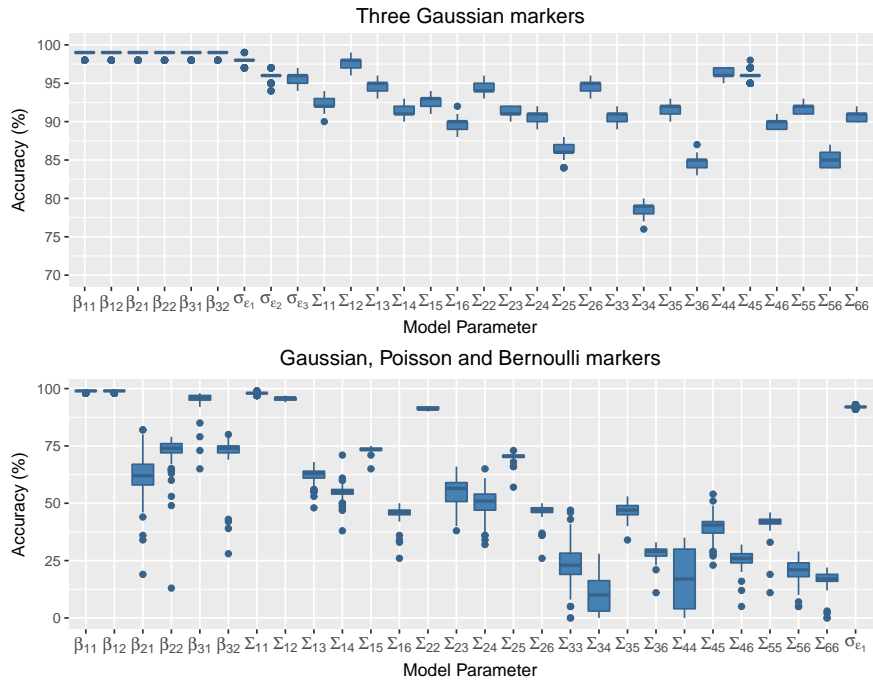


Fig. 2. Accuracy scores for mean field variational Bayes compared to MCMC for simulated datasets with 3 continuous longitudinal markers (top panel) and 3 types of markers (bottom panel) in the simulation with $m = 10000$ individuals

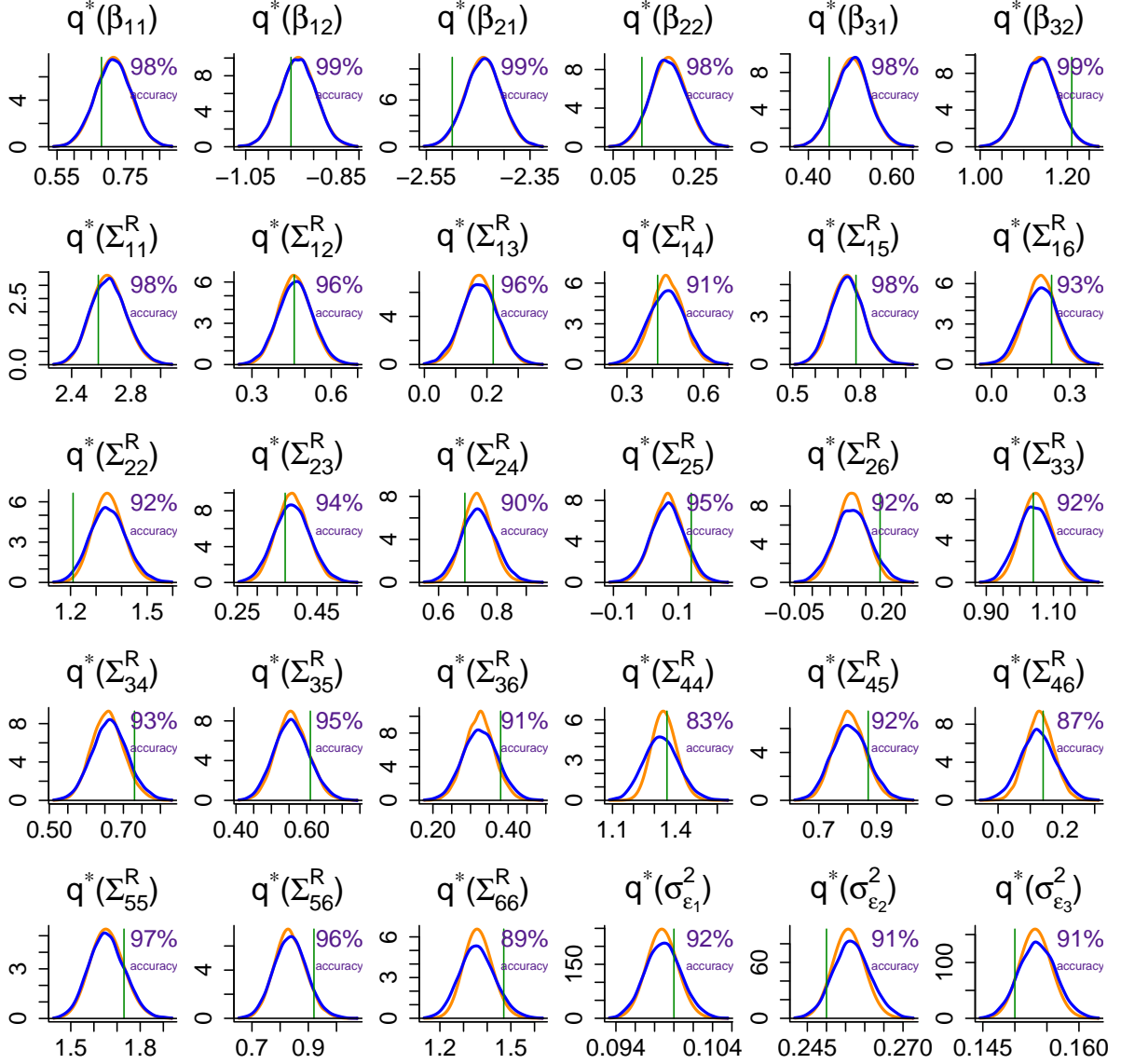


Fig. 3. Accuracy scores for mean field variational Bayes compared to MCMC for simulated datasets with 3 continuous longitudinal markers and $m = 1000$ individuals. The MCMC posteriors are shown by the blue densities, whilst the MFVB posterior densities are in orange. True values are shown by the green vertical line.

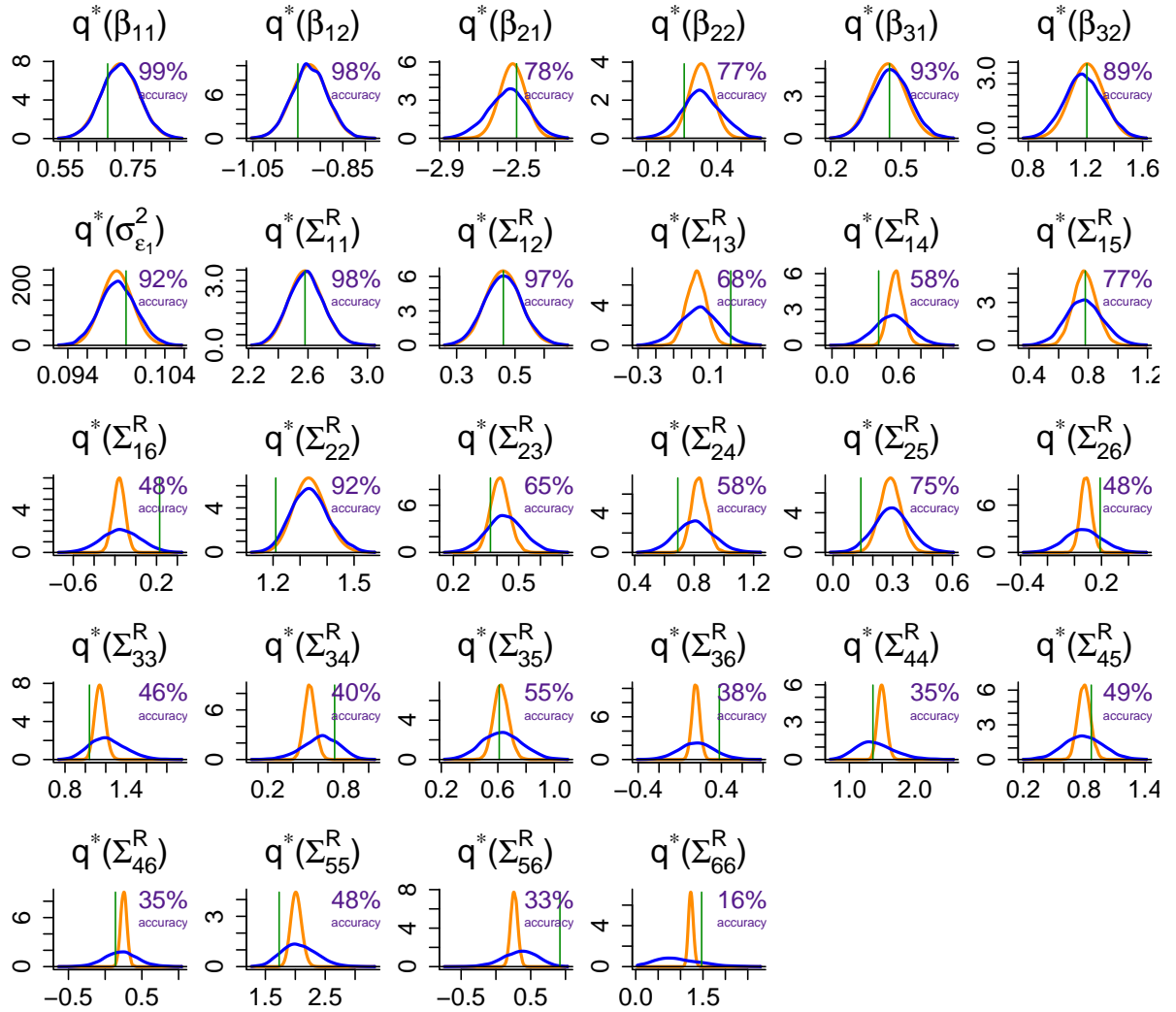


Fig. 4. Accuracy scores for mean field variational Bayes compared to MCMC for simulated datasets with one continuous, one binary and one Poisson longitudinal marker and $m = 1000$ individuals. The MCMC posteriors are shown by the blue densities, whilst the MFVB posterior densities are in orange. True values are shown by the green vertical line.

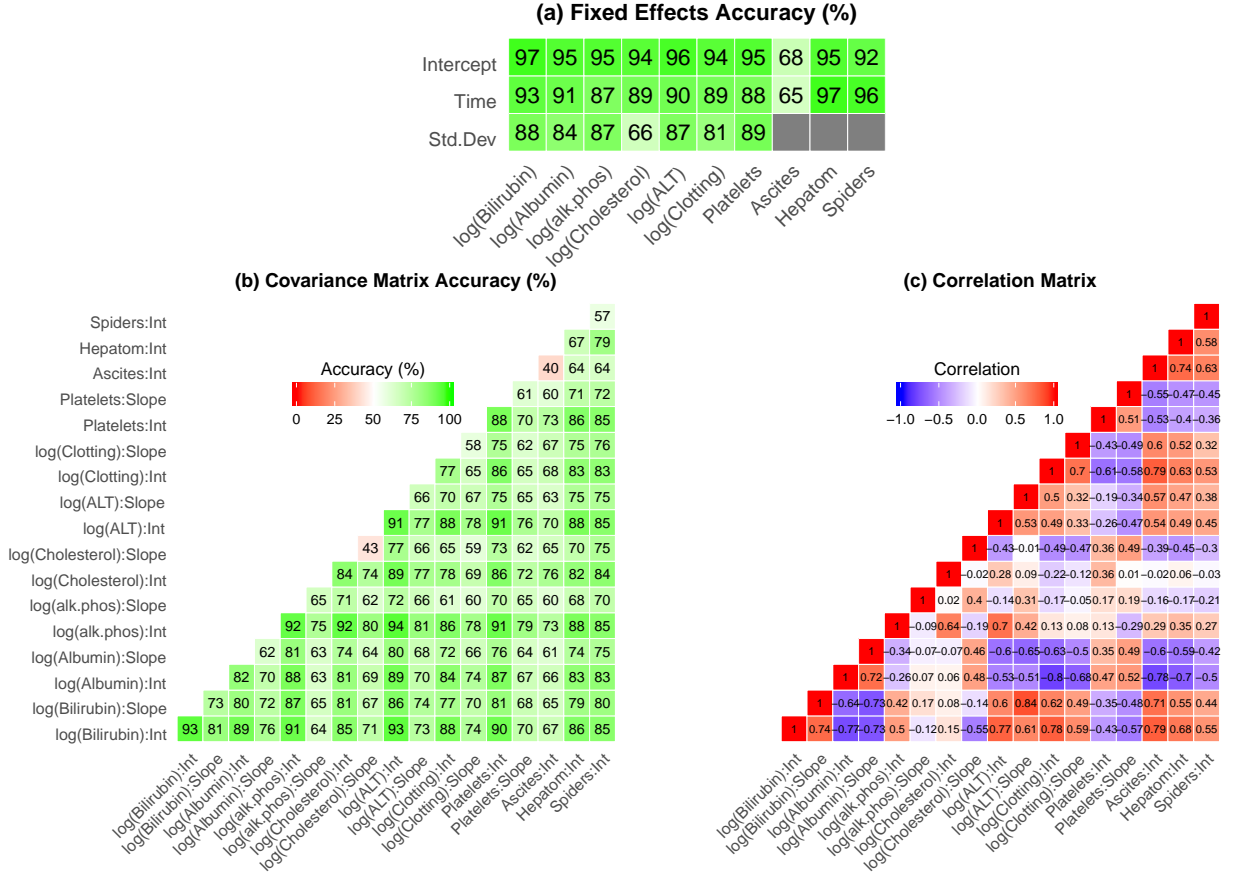


Fig. 5. Model results for a 10 marker multivariate generalised linear mixed model in the primary biliary cirrhosis data. Panel (a) shows the accuracy heat maps of the MFVB fixed effects estimates and residual standard deviations (compared to the MCMC estimates), (b) shows the accuracy of the MFVB random effects covariance matrix (compared to the MCMC estimates), and (c) shows the implied matrix of correlations between the 10 longitudinal markers calculated using MFVB.