Comment on:
Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations
by H. Rue, S. Martino and N. Chopin

BY J.T. ORMEROD & M. P. WAND, *University of Wollongong, Australia*

Full address for receiving proofs:

Professor M.P. Wand,
School of Mathematics and Applied Statistics,
University of Wollongong,
Northfields Avenue, Wollongong 2522,
Australia.
e-mail: mwand@uow.edu.au

18th December, 2008

We concur with the authors that good analytic approximations, as an accurate alternative to MCMC, are worth pursuing. These early results on INLAs are impressive and we look forward to seeing how this methodology progresses. In particular we are interested in the advertised interface from R and eventually giving INLAs a 'test drive'.

Our recent research has involved work in variational approximation for similar models. Most of the discussion in Section 1.6 pertains to a particular version of variational approximation where $q(\mathbf{x},\boldsymbol{\theta}) = q_{\mathbf{x}}(\mathbf{x})q_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. The phrase "the VB approach is not without potential problems" and subsequent discussion actually corresponds to this one type of variational approximation, even though $q(\mathbf{x},\boldsymbol{\theta})$ can be constrained in other ways. Indeed, some variational approximations, such as those developed in Jaakkola & Jordan (2000), do not involve Kullback-Leibler contrast. Lastly, the VB acronym gives the impression of variational approximation being specific to Bayesian approaches, which is not the case.

Recently, we have explored some other approaches to variational approximations that exhibit improved accuracy in our test examples. One approach involves applying the Jaakkola & Jordan (2000) tangent transform idea in a grid-wise fashion (Ormerod, 2008; Ormerod & Wand, 2008). Another takes the Kullback-Leibler contrast route, but restricts q to be in a parametric family, such as the Gaussian distribution. We close with some details on the latter approach, which we call *Gaussian variational approximation*, for frequentist Poisson mixed models with a single variance component:

$$y_{ij}|u_i \overset{\text{ind.}}{\sim} \text{Poisson}\{\exp(\boldsymbol{\beta}^T\mathbf{x}_{ij}+u_i)\}, \quad u_i \overset{\text{ind.}}{\sim} N(0,\sigma^2), \quad 1\le j\le n_i,\ 1\le i\le m. \tag{1}$$

The log-likelihood of $(\boldsymbol{\beta},\sigma^2)$ is

$$\begin{aligned}\ell(\boldsymbol{\beta},\sigma^2) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\{y_{ij}(\boldsymbol{\beta}^T\mathbf{x}_{ij})-\log(y_{ij}!)\}-\tfrac{m}{2}\log(2\pi\sigma^2)\\ &\quad+\sum_{i=1}^{m}\log\int_{-\infty}^{\infty}\exp\left(\sum_{j=1}^{n_i}y_{ij}u-e^{\boldsymbol{\beta}^T\mathbf{x}_{ij}+u}-\frac{u^2}{2\sigma^2}\right)du.\end{aligned}$$

A variational approach to handling the m intractable integrals is to multiply the integrand by the quotient of the $N(\mu_i,\lambda_i)$ density function with itself and invoke Jensen's inequality:

parameter	GVA	exact
β_0	$1.924\ (1.767, 2.081)$	$1.924\ (1.766, 2.082)$
β_{Base}	$0.165\ (-0.128, 0.458)$	$0.165\ (-0.128, 0.459)$
β_{Trt}	$0.842\ (0.013, 1.671)$	$0.842\ (0.014, 1.673)$
β_{BT}	$-0.366(-0.805, 0.072)$	$-0.366\ (-0.806, 0.073)$
β_{Age}	$-0.328(-1.072, 0.416)$	$-0.328\ (-1.074, 0.418)$
β_{V4}	$0.236\ (0.138, 0.333)$	$0.236\ (0.138, 0.333)$
$\tau_\varepsilon^{-1/2}$	$0.580\ (0.466, 0.723)$	$0.581\ (0.461, 0.700)$

Table 1: *Estimates and approximate 95% confidence intervals for Gaussian variational approximation (GVA) corresponding to the example in Section 5.2 of Rue et al. (2008) with the ν_{ij} term omitted. Exact answers (obtained via adaptive Gauss-Hermite quadrature) are given for comparison.*

$$\begin{aligned}
&\log \int_{-\infty}^{\infty} \exp\left(\sum_{j=1}^{n_i} y_{ij}u - e^{\boldsymbol{\beta}^T\mathbf{x}_{ij}+u} - \frac{u^2}{2\sigma^2}\right) \frac{e^{-(u-\mu_i)^2/(2\lambda_i)}/\sqrt{2\pi\lambda_i}}{e^{-(u-\mu_i)^2/(2\lambda_i)}/\sqrt{2\pi\lambda_i}}\,du \\
\geq\;& E_{U\sim N(\mu_i,\lambda_i)}\left\{\left(\sum_{j=1}^{n_i} y_{ij}U - e^{\boldsymbol{\beta}^T\mathbf{x}_{ij}+U} - \frac{U^2}{2\sigma^2}\right) + (U-\mu_i)^2/(2\lambda_i) + \tfrac{1}{2}\log(2\pi\lambda_i)\right\}.
\end{aligned}$$

After simplification we obtain the following lower bound on $\ell(\boldsymbol{\beta}, \sigma^2)$:

$$\begin{aligned}
\underline{\ell(\boldsymbol{\beta}, \sigma^2, \mu, \boldsymbol{\lambda})} \;=\;& \sum_{i=1}^{m}\sum_{j=1}^{n_i}\{y_{ij}\boldsymbol{\beta}^T\mathbf{x}_{ij} - \log(y_{ij}!)\} + \tfrac{m}{2}\{1-\log(\sigma^2)\} \\
&+ \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}\mu_i - e^{\boldsymbol{\beta}^T\mathbf{x}_{ij}+\mu_i+\frac{1}{2}\lambda_i}\right) + \tfrac{1}{2}\sum_{i=1}^{m}\left\{\log(\lambda_i) - \frac{\mu_i^2+\lambda_i}{\sigma^2}\right\}
\end{aligned}$$

for all values of the *variational* parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$. Maximising over these parameters narrows the gap between $\underline{\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})}$ and $\ell(\boldsymbol{\beta}, \sigma^2)$ and so sensible estimators of the model parameters are:

$$(\underline{\widehat{\boldsymbol{\beta}}}, \underline{\widehat{\sigma}}^2) = (\boldsymbol{\beta}, \sigma^2) \text{ component of } \underset{\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}}{\operatorname{argmax}}\, \underline{\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})}.$$

Table 1 conveys excellent performance of Gaussian variational approximation when (1) is applied to the data used in Section 5.2 Rue *et al.* (2008). Early theoretical exploration looks promising.

References

Jaakkola, T.S. & Jordan, M.I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, **10**, 25–37.

Ormerod, J.T. (2008). On Semiparametric Regression and Data Mining. *PhD Thesis*. School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia.

Ormerod, J.T. & Wand, M.P. (2008). Variational approximations for logistic mixed models. *Proceedings of the Ninth Iranian Statistics Conference, Department of Statistics, University of Isfahan, Isfahan, Iran*, pp. 450–467.