David Nott (*National University of Singapore*) **and Robert Kohn** (*University of New South Wales, Sydney*)
We congratulate Rue, Martino and Chopin on their paper which addresses the important issue of effective computation for Bayesian inference. The authors demonstrate that the class of latent models that they consider makes fast computations possible and provides important insights and solutions. The results will be used by applied researchers and will also generate new research in Bayesian computation.

Our first comment concerns Section 6.1 where they suggest copula approximations for marginals of subsets of $\mathbf{x}$ based on the univariate marginals. We have also recently developed some approximate Bayesian computational methods using copulas, in particular for marginal likelihood computation in Bayesian model comparison. The starting point for this is the so-called candidate's formula, similar to expression (3). Writing the set of all unknowns including any latent variables now as simply θ,

$$p(\mathbf{y}) = p(\theta)\, p(\mathbf{y}|\theta)/p(\theta|\mathbf{y}),$$

which holds for any value of θ and clearly approximation of the posterior at a point allows approximation of the marginal likelihood. A subset of the parameters could be handled non-parametrically as in the present paper. Laplace approximation corresponds to use of a Gaussian approximation for $p(\theta|\mathbf{y})$ evaluated at the posterior mode. As an extension it is natural to approximate $p(\theta|\mathbf{y})$ with a Gaussian copula and this can be done both with and without simulation (Nott *et al.*, 2008). Various extensions such as the use of copulas with importance-sampling-based methods and the use of the t-copula instead of a Gaussian copula are possible. It would be interesting to apply copula approximations to posterior distributions based on methods for the current paper in models where those methods can be applied.

A second comment concerns the class of models that was considered. Although this class of models is quite broad, the use of a Gaussian latent variable, a small number of hyperparameters and perhaps most importantly the conditional independence assumptions for $\mathbf{y}$ given $\mathbf{x}$ are serious restrictions for many applications. Future developments of the work that is described in Section 6.5 are important, we feel, as are methods for combining the approximations that are proposed with simulation-based methods.

J. T. Ormerod and M. P. Wand (*University of Wollongong*)
We concur with the authors that good analytic approximations, as an accurate alternative to Markov chain Monte Carlo methods, are worth pursuing. These early results on integrated nested Laplace approximations 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 in giving integrated nested Laplace approximations 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 variational Bayes approach is not without potential problems' and subsequent discussion actually correspond 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 and Jordan (2000), do not involve Kullback–Leibler contrast. Lastly, the name 'variational Bayes' gives the impression of variational approximation being specific to Bayesian approaches, which is not so.

Recently, we have explored some other approaches to variational approximations that exhibit improved accuracy in our test examples. One approach involves applying the Jaakkola and Jordan (2000) tangent transform idea in a gridwise fashion (Ormerod, 2008; Ormerod and 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:

$$y_{ij}|u_i \overset{\text{ind}}{\sim} \text{Poisson}\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij} + u_i)\}, \qquad u_i \overset{\text{ind}}{\sim} N(0, \sigma^2), \quad 1 \leqslant j \leqslant n_i, \quad 1 \leqslant i \leqslant m. \tag{39}$$

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

$$l(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \{y_{ij}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij}) - \log(y_{ij}!)\} - \frac{m}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{m} \log\left[\int_{-\infty}^{\infty} \exp\left\{\sum_{j=1}^{n_i} y_{ij}u - \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij} + u) - \frac{u^2}{2\sigma^2}\right\} \mathrm{d}u\right].$$

Table 1. Estimates and approximate 95% confidence intervals for Gaussian variational approximation corresponding to the example in Section 5.2 with the ν_{ij}-term omitted†

Parameter	*Gaussian variational approximation*	*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 (−1.806, 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)

†Exact answers (obtained via adaptive Gauss–Hermite quadrature are given for comparison.

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 to invoke Jensen's inequality:

$$\log\left[\int_{-\infty}^{\infty} \exp\left\{\sum_{j=1}^{n_i} y_{ij}u - \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij}+u) - \frac{u^2}{2\sigma^2}\right\} \frac{\exp\{-(u-\mu_i)^2/2\lambda_i\}/\sqrt{(2\pi\lambda_i)}}{\exp\{-(u-\mu_i)^2/2\lambda_i\}/\sqrt{(2\pi\lambda_i)}}\,\mathrm{d}u\right]$$
$$\geqslant E_{U\sim N(\mu_i,\lambda_i)}\left[\left\{\sum_{j=1}^{n_i} y_{ij}U - \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij}+U) - \frac{U^2}{2\sigma^2}\right\} + \frac{(U-\mu_i)^2}{2\lambda_i} + \frac{1}{2}\log(2\pi\lambda_i)\right].$$

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

$$\underline{l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\{y_{ij}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij} - \log(y_{ij}!)\} + \frac{m}{2}\{1-\log(\sigma^2)\}$$
$$+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{y_{ij}\mu_i - \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij} + \mu_i + \frac{1}{2}\lambda_i\right)\right\} + \frac{1}{2}\sum_{i=1}^{m}\left\{\log(\lambda_i) - \frac{\mu_i^2+\lambda_i}{\sigma^2}\right\}$$

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

$$(\hat{\underline{\boldsymbol{\beta}}}, \hat{\underline{\sigma}}^2) = (\boldsymbol{\beta}, \sigma^2)\text{component of}\arg\max_{\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}}\{\underline{l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\lambda})}\}.$$

Table 1 conveys excellent performance of Gaussian variational approximation when expression (39) is applied to the data that were used in Section 5.2. Early theoretical exploration looks promising.

Carl Edward Rasmussen (*University of Cambridge*)

I congratulate Professor Rue and his colleagues for their contribution to developing efficient analytic approximation methods for a wide and practically important class of models.

I am concerned, however, about the extent to which the shortcomings of the Laplace approximation may have been treated too lightly when advocating it as a generally applicable tool. The Achilles heel of the Laplace approximation is expansion around the *mode* of the distribution. In high dimensions, for non-Gaussian, non-symmetric posterior distributions, the mode may not be typical of the distribution; for a skew distribution the majority of the mass may lie far to one side of the mode. This is true even for unimodal, log-concave and otherwise fairly harmless distributions. As the Laplace approximation is symmetric around the mode, this may seriously hamper its accuracy.

Gaussian latent variable models with a logistic likelihood is an example in point which has been studied carefully in the machine learning community, where it is known as Gaussian process classification (Rasmussen and Williams, 2006). Careful comparisons between the Laplace approximation and other analytical approximations as well as a Markov chain Monte Carlo gold standard (Kuss and Rasmussen,