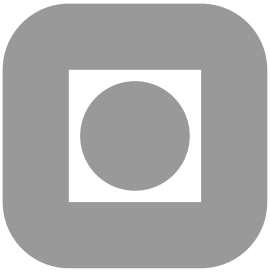
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Extending INLA to a class of near-Gaussian latent models

by

Thiago G. Martins and Håvard Rue

PREPRINT
STATISTICS NO. 6/2012



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2012/S6-2012.pdf
Thiago G. Martins has homepage: http://www.math.ntnu.no/∼guerrera
E-mail: guerrera@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Abstract

This work extends the Integrated Nested Laplace Approximation (INLA) method to latent models outside the scope of latent Gaussian models, where independent components of the latent field can have a near-Gaussian distribution. Two important class of models that can be addressed with our proposed method are non-Gaussian random effects models and dynamic models with non-Gaussian error term for the observation and/or system equation. Our approach is applied to two examples and the results are compared with that obtained by Markov Chain Monte Carlo (MCMC), showing similar accuracy with only a small fraction of computational time. Implementation of the proposed extension is available in the R package `INLA`.

Keywords: Approximate Bayesian inference, INLA, MCMC, near-Gaussian latent models

1 Introduction

Integrated Nested Laplace Approximation (INLA) is an approach proposed by Rue et al. (2009) to perform approximate fully Bayesian inference on the class of latent Gaussian models. It was demonstrated in the original paper that, when compared with the more usual Markov Chain Monte Carlo (MCMC) schemes (Robert and Casella, 2004; Gamerman and Lopes, 2006), INLA outperforms the latter both in terms of accuracy and speed. Monte Carlo averages are characterized by additive $O_p(N^{-1/2})$ errors, where N is the simulated sample size, meaning that we need 100 times more computational time to improve our estimates by one digit. Besides that, due to the additive nature of Monte Carlo estimates, it is even harder to accurately estimate tail probabilities with MCMC. On the other hand, INLA bypass the need for stochastic simulation by an extensive use of simple and fast Gaussian approximations in a clever way to take advantage of the properties of latent Gaussian models, where for most real problems and data sets, the conditional posterior of the latent field is typically well behaved, being close to a Gaussian. As opposed to MCMC, INLA has relative error which allow for more accurate estimates of small quantities, as for example the estimation of tail probabilities.

INLA is not meant to be a replacement of MCMC in applied statistics, but it is a specific tailored algorithm that works extremely well in the broad class of latent Gaussian models, and thus offer a better option in this context. This paper proposes an extension that allows INLA to be applied to models where some independent components of the latent field have a near-Gaussian distribution (see Section 3.1). Interest for such models arise often in the literature but the lack of user friendly software able to handle them in a fast and accurate way might lead someone to stay with the more standard models, even though it might not be the best for their applications. Examples of such models are the survival models with gamma frailty (Ibrahim et al., 2001) where the random effects has a log gamma distribution, more robust mixed effect models (Pinheiro et al., 2001), where the distribution of the random effects is assumed to be Student's t rather than Gaussian, thus allowing for outlier identification and accommodation and non-Gaussian state space models (Kitagawa, 1987) where the distribution of the noise in the state space evolution equation is assumed to be non-Gaussian. The list is, of course, much larger than that, and

a reliable method to fit this large class of models efficiently will provide the right tools for the applied scientist to practice a more flexible and realistic data analysis.

This extension is not straightforward given the central role that the latent Gaussian field plays in the INLA methodology. In this paper, we take advantage of the types of approximations performed and the way they are combined in INLA to propose a new way to look at this latent (near-Gaussian) models of interest and show how to adapt INLA to fit this more complex class of models. The paper is organized as following: Section 2 will describe the latent Gaussian models and the INLA methodology, highlighting the importance of the Gaussian latent field to the success of the method. Section 3 define the sub-class of latent models of interest in this paper and describe our proposed extension to fit this set of models. Section 4 illustrate our approach with two examples and Section 5 concludes with some remarks and point to future research involving the topics discussed in this paper. The proposed extension is already implemented as part of the R package `INLA`, and its use is illustrated in Appendix A, where the R code from the example in Section 4.1 is displayed. [1]

2 Integrated Nested Laplace Approximation

This section contains a brief description of latent Gaussian models and a review of the INLA method proposed by Rue et al. (2009). Latent Gaussian models has a wide range of applications and includes, for example, regression models, dynamic models, spatial and spatiotemporal models. In Section 2.1 we define the class of latent Gaussian models and its hierarchical representation that will make the exposition of the approximation methods described in this paper easier to read. The Gaussian approximation to conditional distributions of the latent Gaussian field, which is the core of INLA is described in Section 2.2. The INLA method applied to latent Gaussian models is described in Section 2.3, while the importance of the Gaussian prior on the latent field to the success of INLA is made explicit is Section 2.4.

2.1 Latent Gaussian models

The INLA framework was designed to deal with latent Gaussian models where the observation (or response) variable y_i is assumed to belong to a distribution family (not necessarily part of the exponential family) where the mean μ_i is linked to a structured additive predictor η_i through a link function $g(\cdot)$, so that $g(\mu_i) = \eta_i$. The structured additive predictor η_i accounts for effects of various covariates in an additive way:

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{\eta_\beta} \beta_k z_{ki} + \epsilon_i, \tag{1}$$

where $\{f^{(j)}(\cdot)\}$'s are unknown functions of the covariates $\boldsymbol{u}$, used for example to relax linear relationship of covariates and to model temporal and/or spatial dependence, the $\{\beta_k\}$'s represent the linear effect of covariates $\boldsymbol{z}$ and the $\{\epsilon_i\}$'s are unstructured terms. Then a Gaussian prior is assigned to α, $\{f^{(j)}(\cdot)\}$, $\{\beta_k\}$ and $\{\epsilon_i\}$.

[1] Please visit `http://www.r-inla.org/` for more information about the R package `INLA`.

We can also write the model described above using a hierarchical structure, where the first stage is formed by the likelihood function with conditional independence properties given the latent field $\boldsymbol{x} = (\boldsymbol{\eta}, \alpha, \boldsymbol{f}, \boldsymbol{\beta})$ and possible hyperparameters $\boldsymbol{\theta}_1$, where each data point $\{y_i, i = 1, ..., n_d\}$ is connected to one element in the latent field x_i. Assuming that the elements of the latent field connected to the data points are positioned on the first n_d elements of $\boldsymbol{x}$, we have

Stage 1. $\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1 \sim \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}_1)$.

The conditional distribution of the latent field $\boldsymbol{x}$ given some possible hyperparameters $\boldsymbol{\theta}_2$ forms the second stage of the model and has a joint Gaussian distribution,

Stage 2. $\boldsymbol{x}|\boldsymbol{\theta}_2 \sim \pi(\boldsymbol{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2))$,

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ denotes a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and a precision matrix $\boldsymbol{Q}$. In most applications, the latent Gaussian field have conditional independence properties, which translates into a sparse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_2)$, which is of extreme importance for the numerical algorithms that will follow. The latent field $\boldsymbol{x}$ may have additional linear constraints of the form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{e}$ for an $k \times n$ matrix $\boldsymbol{A}$ of rank k, where k is the number of constraints and n the size of the latent field. The hierarchical model is then completed with an appropriate prior distribution for the hyperparameters of the model $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$

Stage 3. $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

2.2 The Gaussian approximation

The Gaussian approximation to densities of the form

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto \exp\left\{ -\frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i \in I} g_i(x_i) \right\}, \tag{2}$$

plays a important role in INLA, where $g_i(x_i)$ is a function of x_i that may depend on y_i and $\boldsymbol{\theta}$, and I is an index set. Hence in this section we describe one of the many possible ways to approximate (2) by a Gaussian distribution.

We can perform a Taylor expansion up to second order of $g_i(x_i)$ around an initial guess $\mu_i^{(0)}$

$$g_i(x_i) \approx g_i(\mu_i^{(0)}) + b_i x_i - \frac{1}{2} c_i x_i^2,$$

where b_i and c_i depend on $\mu_i^{(0)}$, and then a Gaussian approximation is obtained with precision matrix $\boldsymbol{Q} + \text{diag}(\boldsymbol{c})$ and mode given by the solution of $\{\boldsymbol{Q} + \text{diag}(\boldsymbol{c})\}\boldsymbol{\mu}^{(1)} = \boldsymbol{b}$, where $\boldsymbol{b}$ and $\boldsymbol{c}$ are vectors formed by $b_i's$ and $c_i's$ respectively. This process is repeated until it converges to a Gaussian distribution with, say, mean $\boldsymbol{x}^*$ and precision matrix $\boldsymbol{Q}^* = \boldsymbol{Q} + \text{diag}(\boldsymbol{c}^*)$, where $\boldsymbol{c}^* = \boldsymbol{c}(\boldsymbol{\mu}^*)$, which we denote hereafter by $\pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$.

2.3 The INLA method

For the hierarchical model described in Section 2.1 the joint posterior distribution of the unknowns then reads

$$\begin{aligned}\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\prod_{i=1}^{n_d}\pi(y_i|x_i,\boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta})|\boldsymbol{Q}(\boldsymbol{\theta})|^{n/2}\exp\left[-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x}+\sum_{i=1}^{n_d}\log\{\pi(y_i|x_i,\boldsymbol{\theta})\}\right]\end{aligned}$$

The approximated posterior marginals of interest $\tilde{\pi}(x_i|\boldsymbol{y})$, $i = 1, .., n$ and $\tilde{\pi}(\theta_j|\boldsymbol{y})$, $j = 1, ..., m$ returned by INLA has the following form

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}^{(k)},\boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\boldsymbol{y})\ \Delta\boldsymbol{\theta}^{(k)} \tag{3}$$

$$\tilde{\pi}(\theta_j|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_j \tag{4}$$

where $\{\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\boldsymbol{y})\}$ are the density values computed during a grid exploration on $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$. Since we don't have $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ evaluated at all points required to compute the integral in Eq. (4) we construct an interpolation $I(\boldsymbol{\theta}|\boldsymbol{y})$ using the density values $\{\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\boldsymbol{y})\}$ computed during the grid exploration on $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ and approximate (4) by

$$\tilde{\pi}(\theta_j|\boldsymbol{y}) = \int I(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-j}. \tag{5}$$

Looking at [(3)-(5)] we can see that the method can be divided into three main tasks, (1) propose an approximation $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ to the joint posterior of the hyperparameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, (2) propose an approximation $\tilde{\pi}(x_i|\boldsymbol{\theta},\boldsymbol{y})$ to the marginals of the conditional distribution of the latent field given the data and the hyperparameters $\pi(x_i|\boldsymbol{\theta},\boldsymbol{y})$ and (3) explore $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ on a grid and use it to integrate out $\boldsymbol{\theta}$ in Eq. (3) and $\boldsymbol{\theta}_{-j}$ in Eq. (5).

The approximation used for the joint posterior of the hyperparameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\pi_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}*(\boldsymbol{\theta})} \tag{6}$$

where $\pi_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ is the Gaussian approximation (see Section 2.2) to the full conditional of $\boldsymbol{x}$, and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\boldsymbol{x}$, for a given $\boldsymbol{\theta}$. Expression (6) is equivalent to Tierney and Kadane (1986) Laplace approximation of a marginal posterior distribution, and it is exact if $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ is a Gaussian.

For $\pi(x_i|\boldsymbol{\theta},\boldsymbol{y})$, three options are available, and they vary in terms of speed and accuracy. The fastest option, $\pi_G(x_i|\boldsymbol{\theta},\boldsymbol{y})$, is to use the marginals of the Gaussian approximation $\pi_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ already computed when evaluating expression (6). The only extra cost to obtain $\pi_G(x_i|\boldsymbol{\theta},\boldsymbol{y})$ is to compute the marginal variances from the sparse precision matrix of $\pi_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$. The Gaussian approximation often gives reasonable results, but there can be errors in the location and/or errors due to the lack of skewness

(Rue and Martino, 2007). The more accurate approach would be to do again a Laplace approximation, denoted by $\pi_{LA}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, with a form similar to expression (6)

$$\pi_{LA}(x_i|\boldsymbol{\theta}, \boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_{GG}(\boldsymbol{x}_i|x_i, \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}_{-i}=\boldsymbol{x}_{-i}*(x_i,\boldsymbol{\theta})}, \tag{7}$$

where $\pi_{GG}(\boldsymbol{x}_i|x_i, \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation to $\boldsymbol{x}_i|x_i, \boldsymbol{\theta}, \boldsymbol{y}$ and $\boldsymbol{x}_{-i} * (x_i, \boldsymbol{\theta})$ is the modal configuration. A third option $\pi_{SLA}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, called simplified Laplace approximation, is obtained by doing a Taylor expansion on the numerator and denominator of expression (7) up to third order, thus correcting the Gaussian approximation for location and skewness with a much lower cost when compared to $\pi_{LA}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. We refer to Rue et al. (2009) for a detailed description of the Gaussian, Laplace and simplified Laplace approximations to $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$.

2.4 INLA and the importance of the Gaussian field

The main challenge in applying INLA to latent models is that the method depends heavily on the latent Gaussian prior assumption to work properly, both from the computational point of view and from the choice of approximations used as described in Section 2.3.

For the full conditional of $\boldsymbol{x}$ to be well approximated by a Gaussian distribution in equations (6) and (7), we need it to be well behaved and close to a Gaussian. This is basically ensured by the latent Gaussian prior that is assigned to $\boldsymbol{x}$ (see Stage 2 of Section 2.1) in latent Gaussian models, which has a non-negligible effect on the posterior, especially in terms of dependence between the components of $\boldsymbol{x}$.

Another important issue is that the conditional independence properties often encountered in the latent field translates into a sparse precision matrix when it is Gaussian distributed. This imply a huge decrease in computational time when performing the Gaussian approximation, which is extremely important since a Gaussian approximation needs to be computed for each value $\boldsymbol{\theta}^{(k)}$ used on the grid for the numerical integration in Eq. (3).

3 Extension to near-Gaussian latent models

In this section we show how to extend the scope of INLA to include models similar in structure to the latent Gaussian models described in Section 2.1 but where the prior for some components of the latent field can have a near-Gaussian distribution. Section 3.1 will define these latent models in general while Section 3.2 will present an specific example, namely a survival model with Gamma frailty.

3.1 Near-Gaussian latent models

The models we are interested in this paper has the same structure as the latent Gaussian models described in Section 2.1 with the exception that the latent field has some independent non-Gaussian components. We redefine stage 2 of the hierarchical model of Section 2.1 as

Stage 2^{new}. $\underbrace{(\boldsymbol{x}_G, \boldsymbol{x}_{NG})}_{\boldsymbol{x}}|\boldsymbol{\theta}_2 \sim \pi(\boldsymbol{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\boldsymbol{x}_G; \boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_i \pi(\boldsymbol{x}_{NGi}|\boldsymbol{\theta}_2)$,

where $\boldsymbol{x}_G$ and $\boldsymbol{x}_{NG}$ represent the Gaussian and non-Gaussian terms of the latent field, respectively. In addition we assume that $\boldsymbol{x}_{NG}$ is formed by independent random variables. As a result, the distribution of the latent field is not Gaussian anymore, which precludes the use of INLA to fit this class of models.

The term *near-Gaussian* latent models refer to the restrictions we impose on the non-Gaussian components of the latent field. We aim for non-Gaussian distributions that are not to different from a Gaussian one, and are so that they could be well enough approximated by a Gaussian density (at least locally). Unfortunately, it is not very useful to make a precise definition of this property, as what matters in the end is how our new approximation scheme performs in practice. We hope that it all becomes more clear and intuitive through out the rest of Section 3.

A first approach to fit these models with INLA would be to include the non-Gaussian components $\boldsymbol{x}_{NG}$ in the hyperparameters $\boldsymbol{\theta}$ of the model. However, this is not a good idea in practice since the size of $\boldsymbol{x}_{NG}$ is usually large and typically increase as the number of data points. This naive approach would lead to accurate results but the cost would be a large increase in computational time due to the grid exploration necessary to compute Eqs. (3)-(5). Our approach will deliver accurate results without the burden in computational time.

3.2 Survival model - A first example

Consider the following exponential model that can be used to analyze survival data that comes from subjects of the same group who are related to each other or from multiple recurrence times of a event for the same individual. The likelihood

$$t_{ij} \sim \exp(\lambda_{ij}), \quad i = 1, ..., I \text{ and } j = 1, ..., J \tag{8}$$

is given by independent exponential distributions given $\boldsymbol{\lambda} = \{\lambda_{ij},\ i = 1, ..., I,\ j = 1, ..., J\}$, where $\lambda_{ij} = 1/\mu_{ij}$ and μ_{ij} is the mean of t_{ij}. The index I could be interpreted as the number of groups in the data, while J would be the number of individuals in each group. This is a case of balanced data-set, but the unbalanced case could be treated just as easy by our method.

It is expected that individuals belonging to the same group are correlated with each other, this can be included in the model through the addition of random effects $\boldsymbol{w} = \{w_1, ..., w_I\}$ to account for variation between groups,

$$\eta_{ij} = \log(\lambda_{ij}) = \beta_0 + \beta_1 x_{ij} + \log w_i. \tag{9}$$

Besides the random effects, it is common to include some covariate effects that in our case are represented by the fixed effects β_0 and β_1. In the survival analysis literature, the random effects are often called frailty and it is common to assume that they are Gamma distributed, $w_i \sim \text{Gamma}(\kappa, \kappa)$, with $E(w_i) = 1$ to avoid identifiability issues. Gaussian priors are assumed for the fixed effects and a Gamma prior is often used for the random-effect hyperparameter κ.

The latent field for this model is given by $\boldsymbol{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{b})$, with $\boldsymbol{b} = \log(\boldsymbol{w})$ and it is non Gaussian since

$\boldsymbol{b}$ is formed by independent log-Gamma random variables,

$$\pi(\boldsymbol{b}|\kappa) = \prod_{i=1}^{I} \pi(b_i|\kappa) = \prod_{i=1}^{I} \frac{\kappa^\kappa}{\Gamma(\kappa)} \exp\{\kappa(b_i - \exp(b_i))\}. \tag{10}$$

Such a model cannot be applied straightforwardly using INLA since the assumption in Stage 2 is violated. However, it fits the class of models in Section 3.1 and will be further analyzed in Section 4.1 with our approach that we now describe.

3.3 The extension

It was described in Section 2.4 some points that explain the importance of the latent Gaussian prior for INLA to run smoothly. With that in mind we propose to approximate the prior of the non-Gaussian components $\pi(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2)$ by a Gaussian distribution $\pi_G(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2)$ and correct this approximation by the correction term

$$CT = \pi(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2)/\pi_G(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2) \tag{11}$$

in the likelihood. This is, in fact, a way of writing a latent model of the form described in Section 3.1 into a latent Gaussian model, defined in Section 2.1.

The first stage is now formed by the original likelihood multiplied by the correction term

$$\prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}_1) \times \pi(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2)/\pi_G(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2).$$

Another way of writing this is to define an extended response vector $\boldsymbol{z}$, with $z_i = y_i$ if $i \leq n_d$ and $z_i = 0$ if $n_d < i \leq n_d + k$, where k is the length of $\boldsymbol{x}_{NG}$ and write

Stage 1. $\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta} \sim \pi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d+k} \pi(z_i|x_i, \boldsymbol{\theta})$, where

$$\pi(z_i|x_i, \boldsymbol{\theta}) = \begin{cases} \pi(y_i|x_i, \boldsymbol{\theta}_1) & \text{for } 1 \leq i \leq n_d \\ \pi(x_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(x_{NG,i}|\boldsymbol{\theta}_2) & \text{for } n_d < i \leq n_d + k \end{cases} \tag{12}$$

It is important to emphasize that Stage 1 above is not the likelihood function, but expressing the model using this form will make the description and implementation of the algorithm that follows easier. The latent field has now a Gaussian approximation replacing the non-Gaussian distribution of $\boldsymbol{x}_{NG}$,

Stage 2. $\underbrace{(\boldsymbol{x}_G, \boldsymbol{x}_{NG})}_{\boldsymbol{x}}|\boldsymbol{\theta}_2 \sim \pi(\boldsymbol{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\boldsymbol{x}_G; \boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)) \times \pi_G(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2)$,

which means that now $\pi(\boldsymbol{x}|\boldsymbol{\theta}_2)$ is Gaussian distributed. The third stage is once again formed by the prior distribution on the hyperparameters,

Stage 3. $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

Independent of the Gaussian approximation $\pi_G(\boldsymbol{x}_{NG}|\boldsymbol{\theta}_2)$ used, the hierarchical model above is equivalent to the model described in Section 3.1. Considerations on how to chose this Gaussian approximation and how this model formulation will help us to perform inference will be presented soon, but first lets rewrite the survival model of Section 3.2 in this new formulation.

Survival model (Cont.)

For the survival model defined in Section 3.2, we have that the original likelihood function, defined in Eq. (8), is an exponential distribution

$$\log \pi(t_{ij}|\eta_{ij}) = \eta_{ij} - \exp(\eta_{ij} t_{ij}),$$

where η_{ij} is the linear predictor defined in Eq. (9). Based on Eq. (10) the correction term (see Eq. (11)) is given by

$$\begin{aligned} CT &= \prod_{i=n_d+1}^{n_d+I} \pi(\boldsymbol{x}_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(\boldsymbol{x}_{NG,i}|\boldsymbol{\theta}_2) \\ &= \prod_{i=1}^{I} \pi(b_i|\kappa)/\pi_G(b_i|\mu_b, \tau_b) = CT_i, \end{aligned}$$

with

$$\log CT_i = \kappa(b_i - \exp(b_i)) + \frac{\tau_b(\kappa)}{2}(b_i - \mu_b(\kappa))^2 + \text{const}, \tag{13}$$

where $\mu_b(\kappa)$ and $\tau_b(\kappa)$ are the mean and precision parameter of the Gaussian approximation to the log-Gamma random effects $\boldsymbol{b}$ and const is a constant that does not depend on $\boldsymbol{b}$. The latent field $\boldsymbol{x} = (\boldsymbol{\eta}, \boldsymbol{b}, \boldsymbol{\beta})$ is now Gaussian since $\pi(\boldsymbol{b}|\kappa)$ is approximated by $\pi_G(\boldsymbol{b}|\mu_b(\kappa), \tau_b(\kappa)) = \mathcal{N}(\boldsymbol{b}; \mu_b(\kappa)\mathbf{1}_I, \tau_b(\kappa)^{-1}\boldsymbol{I}_I)$, where $\mathbf{1}_n$ is a vector of ones with dimension n and $\boldsymbol{I}_n$ is an $n \times n$ identity matrix.

□

We now have our latent (non-Gaussian) model of interest written as a latent Gaussian model and can adapt INLA to perform inference on such models. The main change is on the Gaussian approximation to the full conditional of the latent field (see Section 2.2), that now takes the form

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto \exp\left\{ -\frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i=1}^{n_d} g_i(x_i) + \sum_{n_d+1}^{n_d+k} h_i(x_i) \right\}, \tag{14}$$

where $g_i(x_i) = \log \pi(y_i|x_i, \boldsymbol{\theta})$ as before and

$$h_i(x_i) = \log CT_i = \log \pi(\boldsymbol{x}_{NG,i}|\boldsymbol{\theta}_2) - \log \pi_G(\boldsymbol{x}_{NG,i}|\boldsymbol{\theta}_2).$$

It was shown in Section 2.2 that a Gaussian approximation is obtained by approximating the non-quadratic functions, in this case $g_i(x_i)$ and $h_i(x_i)$, by quadratic functions using Taylor expansion up to second order. Once we know we are dealing with a well behaved log likelihood function $g_i(x_i)$ as, for example, those belonging to the exponential family, the success of a Gaussian approximation to Eq. (14) will depend heavily on the shape of $h_i(x_i)$. For instance, it is desirable to have a bounded correction term

$$\pi_{NG}(\cdot|\boldsymbol{\theta})/\pi_G(\cdot|\boldsymbol{\theta}) < \infty$$

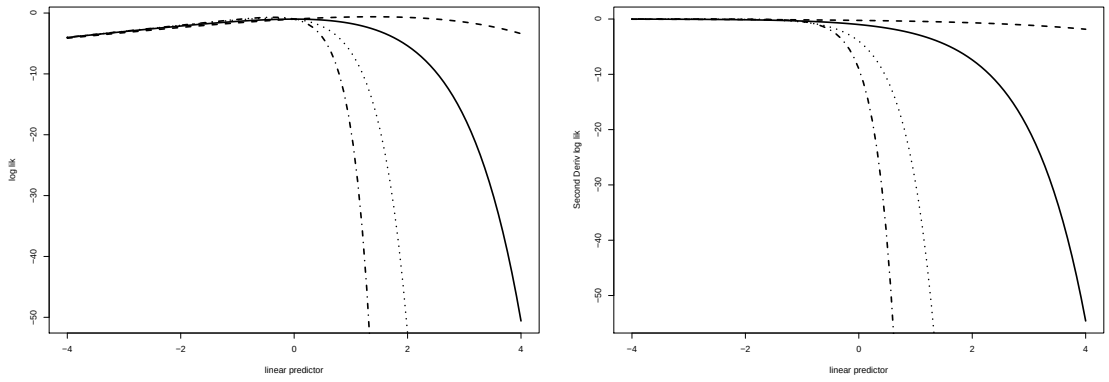
for a quadratic form approximation to $h_i(x_i)$ to make sense. This imply that the Gaussian approximation $\pi_G(\boldsymbol{x}_{NG}|\boldsymbol{\theta})$ should ideally dominate $\pi_{NG}(\boldsymbol{x}_{NG}|\boldsymbol{\theta})$ in the sense that it should have thicker tails than the

non-Gaussian distribution it is trying to approximate. But in practice, it is sufficient to have a bounded correction term on the region that concentrates the bulk of probability mass since we can afford a bigger approximation error on the region that doesn't contribute much to the density (14).

Although not necessary, it is also desirable to have a log-concave correction term, at least on a neighborhood of the mode of Eq. (14), since it gives more stability to the optimization step required to find the mode of $\pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, which is necessary to compute Eq. (6). In our examples, we have chosen $\pi_G(\cdot|\boldsymbol{\theta})$ to be a Gaussian distribution with zero mean and low precision to approximate the distribution of the non-Gaussian components $\pi_{NG}(\cdot|\boldsymbol{\theta})$.

Survival model (Cont.)

For the survival model, we have a log-concave likelihood function as we can see in Figure 1, where we have the plots of the log likelihood and the second derivative of the log likelihood for a given data point, assuming different values for the data point.



(a) (b)

Figure 1: Plot of the log likelihood (a) and of the second derivative of the log likelihood (b) against the linear predictor for different values of data points for the survival model. It was used the following values for the data point: 0.5 (Dashed), 1 (Solid), 2 (Dotted) and 3 (Dot-dash)

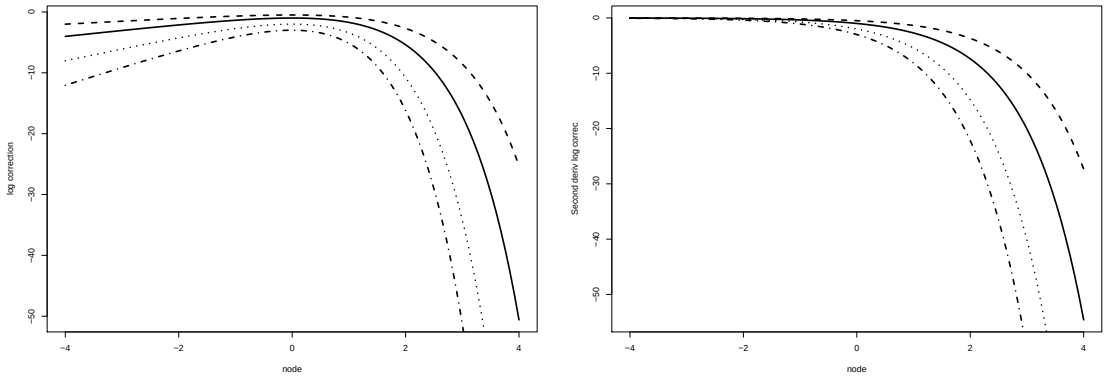
If we use a zero mean and low precision Gaussian distribution ($\mu_b = 0$ and $\tau_b \to 0$) in Eq. (13) we also attain a log-concave correction term,

$$\log CT_i = \kappa(b_i - \exp(b_i)) + \text{const}$$

illustrated in Figure 2.

□

Our approach approximates the latent field by a latent Gaussian field to take advantage of the benefits described in Section 2.4 at the cost of having a more complicated "likelihood" as displayed in Eq. (12),



(a) (b)

Figure 2: Plot of the log correction term (a) and of the second derivative of the log correction term (b) against b_i for different values of κ for the survival model. It was used the following values for κ: 0.5 (Dashed), 1 (Solid), 2 (Dotted) and 3 (Dot-dash)

which will basically affect the Gaussian approximation in Eqs. (6) and (7). Performance considerations of this approach will be given along the examples of Section 4 and in the conclusion of the paper in Section 5.

We now proceed to two examples where we apply the methodology proposed in this paper and compare the results with that obtained by MCMC.

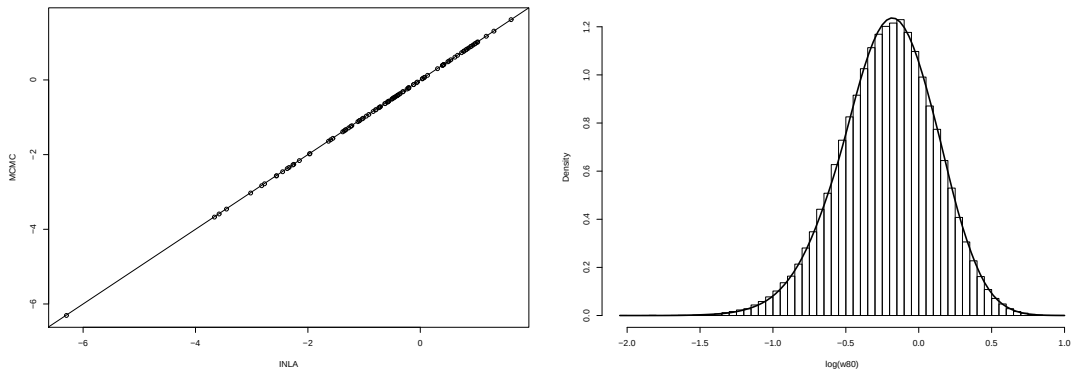
4 Examples

4.1 Survival analysis with Gamma frailty

Here we apply our proposed extension to fit the model defined in Section 3.2 in a simulated data-set and compare the results with that obtained by MCMC. For the experiment reported we have simulated 100 groups, each of which with 10 individuals. The covariates $\{x_{ij}\}$ in Eq. (9) were simulated from a uniform distribution on the interval $(0, 1)$ while the frailties came from a Gamma distribution with both the shape and rate parameters equal to 1. The fixed effects β_0 and β_1 were chosen to be 1.

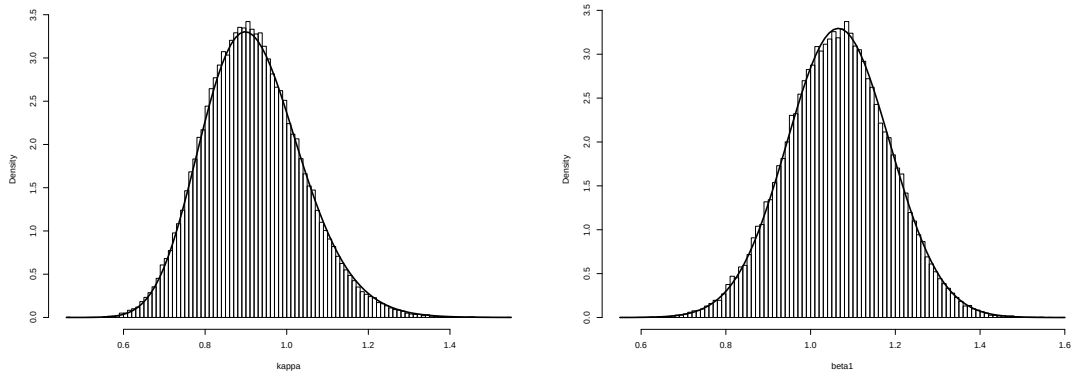
We use OpenBUGS (Lunn et al., 2009) to generate samples from the posterior distribution. Figure 3(a) show the posterior mean of the log frailties $\{b_i : b_i = \log w_i,\ i = 1, ..., 100\}$ obtained by INLA (x-axis) and by MCMC (y-axis). An identity function is also plotted in order to help visualize the strong agreement between both methods. Figure 3(b) display the histogram formed by the samples of $\pi(b_{80}|\boldsymbol{y})$ returned by OpenBUGS and the line is the approximated posterior computed using our extension. This specific component was chosen at random, since similar accuracy was obtained for all log frailties in our simulation study. Figure 4 show similar pictures for β_1 and κ to show that the excellent results are also valid for the fixed effects and for the hyperparameter κ of our model. The R code used to run INLA in

this example is available in Appendix A.



(a) (b)

Figure 3: Comparison between INLA and MCMC for the exponential gamma frailty example: (a) Plot of the posterior mean of the log frailties returned by INLA (x-axis) vs. MCMC (y-axis). (b) Approximate posterior density for $\log w_{80}$ obtained by INLA (solid line) and by MCMC (histogram).



(a) (b)

Figure 4: Comparison between INLA and MCMC for the exponential gamma frailty example: (a) Approximate posterior density for κ obtained by INLA (solid line) and by MCMC (histogram) (b) Approximate posterior density for β_1 obtained by INLA (solid line) and by MCMC (histogram)

It is important to note that the comparisons with MCMC were made with millions of samples, taking minutes to run, since for short to medium number of samples it was possible to visually detect errors in the MCMC estimates when compared to INLA, that took a little bit more than 1 second to run on a Intel Core i5 with 2.67GHz. One can argue that the number of samples (and time) necessary by a MCMC scheme to attain the desired accuracy of our application could be reduced if more time was spent designing a specific MCMC scheme for this particular application instead of using the general purpose OpenBUGS. Although this is true in theory, we are here comparing two general purpose tools for the

class of latent models into consideration, and even with a tailored MCMC scheme, we believe that the difference in time will still be in orders of magnitude, not to mention the time necessary to develop specific algorithms for each new model belonging to this same class.

4.2 Robust mixed-effects models using Student-t distribution

In this example, we show our method applied to a Bayesian random effects model where the random effects has an Student-t distribution. Assume we have data $\{\boldsymbol{y}_i;\ i = 1, ..., n\}$ recorded for n groups each having k_i individuals. Lets assume that $\boldsymbol{y}_i's$ are independent Gaussian random vector described by the following standard mixed effect model (Laird and Ware, 1982), useful to analyze repeated measures or grouped data,

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + b_i + \boldsymbol{e}_i, \tag{15}$$

where both the random effect b_i and the error term $\boldsymbol{e}_i$ have a Gaussian distribution, $b_i \sim N(0, \sigma_b^2)$ and $e_i \sim N(0, \sigma_e^2\boldsymbol{I})$ with variances σ_b^2 and σ_e^2 respectively, being $\boldsymbol{I}$ a $(k_i \times k_i)$ identity matrix. X_i represent the $(k_i \times p)$ design matrix for group i and β is a $(p \times 1)$ vector of fixed effects.

Statistical inference based on the Gaussian distribution is known to be vulnerable to outliers. One approach to more robust modeling is to replace the Gaussian distribution by Student-t distribution in the model. In the context of linear mixed effects model, Pinheiro et al. (2001) suggested to follow the robust statistical modeling approach described by Lange et al. (1989) in which the Gaussian distributions of b_i and e_i are replaced by t-distributions,

$$b_i \sim t(0, \psi_b^2, \nu), \quad \boldsymbol{e}_i \sim t(0, \psi_e^2\boldsymbol{I}, \nu), \tag{16}$$

where ψ_b^2 and ψ_e^2 are the scale parameters and ν is the common degree of freedom parameter. They also noted that in mixed effects models the outlier may occur either at the level of within-group error $\boldsymbol{e}_i$, called $\boldsymbol{e}$-outliers, or at the level of random effects b_i, called $\boldsymbol{b}$-outliers. This approach can be regarded as outlier accommodation although it provides useful information for outlier identification. For the simulation experiment performed later in this Section, we have used a Gamma prior with shape and rate parameters given by 1 and 0.1 respectively for the inverse scale parameters, a Gaussian distribution with mean zero and low precision (10^{-4}) for the fixed effects and a Gaussian distribution with mean 3 and variance 1 for $\nu^* = \log(\nu - 5)$, so that the bulk of prior probability mass is between 7 and 150 for the degree of freedom parameter ν. Note that we have defined ν^* so that $\nu > 5$ in order to get a well defined first four moments of the Student-t distribution.

The model (15)-(16) has the likelihood function formed by a t distribution

$$y_{ij}|\boldsymbol{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n \text{ and } j = 1, ..., k_i,$$

which does not belong to the exponential family, where η_{ij} is the linear predictor

$$\eta_{ij} = \boldsymbol{X}_i\boldsymbol{\beta} + b_i.$$

The latent field is then formed by $\boldsymbol{x} = (\boldsymbol{\eta}, \boldsymbol{b}, \boldsymbol{\beta})$, where $\boldsymbol{b} = \{b_i;\ i = 1, ..., n\}$ is formed by independent t distributed random variables and has therefore a non-Gaussian distribution given by

$$\pi(\boldsymbol{b}|\psi_b^2, \nu) = \prod_{i=1}^{n} \pi(b_i|\psi_b^2, \nu) = \prod_{i=1}^{n} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (\psi_b^2 \pi\nu)^{-1/2} \left[1 + \frac{b_i^2}{\psi_b^2 \nu}\right]^{-(\nu+1)/2} \tag{17}$$

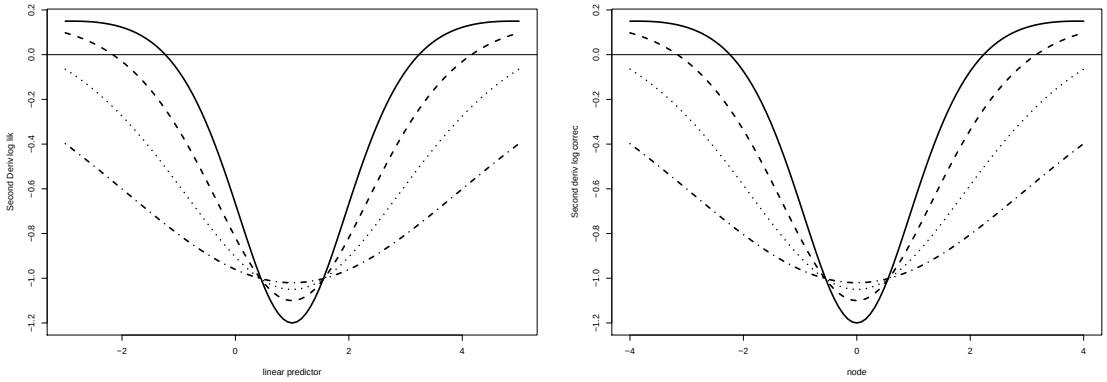
If we use Eq. (11) and (17) we get the following log correction term

$$\begin{aligned} \log CT_i &= \log \pi(b_i|\psi_b^2, \nu) - \log \pi_G(b_i|\mu_b, \tau_b) \\ &= -\frac{(\nu+1)}{2} \log\left\{1 + \frac{b_i^2}{\psi_b^2 \nu}\right\} + \frac{\tau_b}{2}(b_i - \mu_b)^2 + \text{const.} \end{aligned}$$

Again, if we assume a zero mean and low precision Gaussian distribution ($\mu_b = 0$ and $\tau_b \to 0$) we end up with

$$\log CT_i = -\frac{(\nu+1)}{2} \log\left\{1 + \frac{b_i^2}{\psi_b^2 \nu}\right\} + \text{const.}$$

Figure 5 show plots of the second derivative of the likelihood (Figure 5(a)) and of the log correction (Figure 5(b)) term assuming a data point $y = 1$, variances $\psi_e^2 = 1$, $\psi_b^2 = 1$ and different values for ν ($\nu = 5, 10, 20, 50$).



(a) (b)

Figure 5: (a) Plot of the second derivative of the log likelihood against linear predictor for $\psi_e^2 = 1$, $y = 1$ and different values of ν and (b) plot of the second derivative of the log correction term against b_i for $\psi_b^2 = 1$ and different values of ν for the t- mixed effect model. It was used the following values for ν: 5 (Solid), 10 (Dashed), 20 (Dotted) and 50 (Dot-dash).

We have here an example where both the likelihood and the correction term are not log-concave, specially for low values of degree of freedom parameter ν. We can still apply our extension to such models but one has to be careful when designing the optimization algorithm to optimize Eq. (14) on such cases.

Now we can proceed to a contamination study similar to that performed in Pinheiro et al. (2001). Here we have data simulated from

$$y_i = X\beta + b_i + e_i, \quad i = 1, ..., 27, \quad X = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \tag{18}$$

with the following mixture of Gaussian models being used to contaminate the distributions of the b_i and the e_i.

$$b_i \overset{ind}{\sim} (1 - p_b)\mathcal{N}(0, \sigma_b^2) + p_b f\mathcal{N}(0, \sigma_b^2)$$
$$e_i \overset{ind}{\sim} (1 - p_e)\mathcal{N}(0, \sigma_e^2) + p_e f\mathcal{N}(0, \sigma_e^2), \quad i = 1, ..., 27, \quad j = 1, ..., 4$$

where p_b and p_e denote, respectively, the expected percentage of $\boldsymbol{b}$- and $\boldsymbol{e}$-outliers in the data and f denotes the contamination factor. The true parameters for the uncontaminated distributions are $\sigma_b^2 = 3$ and $\sigma_e^2 = 2$, while the true values for the fixed effects are $\beta = (12, 1)^T$.

All 32 combinations of $p_b, p_e = 0, .05, .1, .25$, and $f = 2, 4$ were used in the simulation study. The $f = 2$ case corresponds to a close contamination pattern, while $f = 4$ illustrates a more distant contamination pattern. A total of 500 Monte Carlo replications were obtained for each $(p_b,\ p_e,\ f)$ combination.

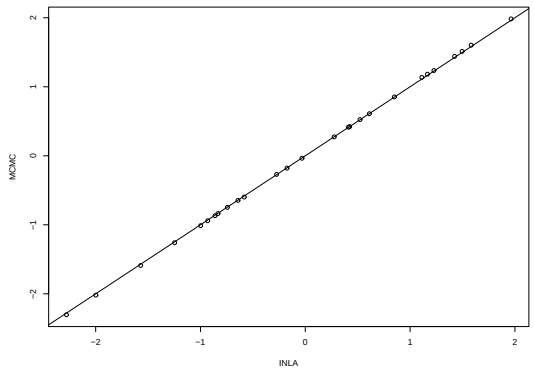
Let θ denote a parameter of interest, with target value $\theta_0 \neq 0$, estimated by $\hat{\theta}$, which in our case will be the posterior mean of θ. The efficiency of the Gaussian estimator $\hat{\theta}_G$ relative to the multivariate t estimator $\hat{\theta}_T$ is defined as the ratio of the respective mean square errors,

$$E(\hat{\theta}_G - \theta_0)^2 / E(\hat{\theta}_T - \theta_0)^2, \tag{19}$$

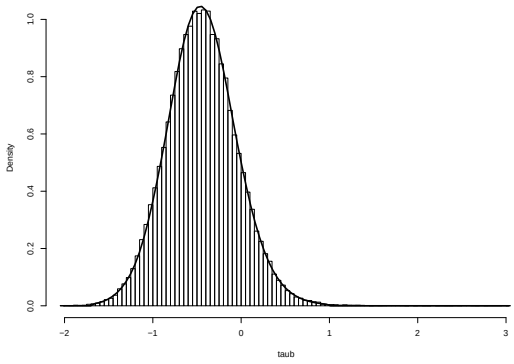
where expectations are taken with respect to the simulation distribution, that is $\widehat{E(\hat{\theta} - \theta_0)^2} = \sum_{i=1}^{500}(\hat{\theta}_i - \theta_0)^2/500$.

We have chosen some data-sets out of the $32 \times 500 = 16000$ used in this contamination study and fitted the model using both MCMC and INLA to make sure that INLA is doing at least as good as MCMC in terms of accuracy. After that we proceed with the contamination study with INLA as the only estimation method as it would be impractical to fit all 16000 data-sets with MCMC. Figures 6 and 7 illustrate this comparison for one of the data-sets, where Figure 6(a) display the log random effects returned by INLA (x-axis) vs. MCMC (y-axis), while Figures 6(b) and 7 display the approximate posterior densities for $\log \tau_b = \log 1/\psi_b^2$ and for the log fixed effects $\log \beta_0$ and $\log \beta_1$ respectively, obtained by INLA (solid line) and by MCMC (histogram).

Figures 8 and 9 plots the relative efficiencies, defined in Eq. (19), between the posterior means of the t model over the Gaussian linear mixed effects model. Based on the plots the conclusion of our simulation study are, as expected, similar to that obtained by Pinheiro et al. (2001). There are substantial gains in efficiency for all parameters under the more distant contamination pattern ($f = 4$) and moderate gains under the close contamination pattern ($f = 2$). The efficiency gains are bigger for the precision of the random effects and the non-monotonic behavior of the efficiency gains suggest that the t model is

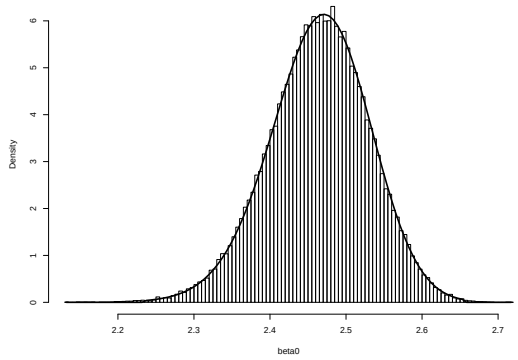


(a) Plot of the posterior mean of the log random effects returned by INLA (x-axis) vs. MCMC (y-axis).

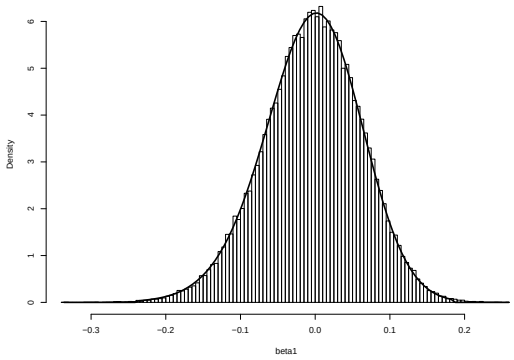


(b) Approximate posterior density for $\log \tau_b = \log 1/\psi_b^2$ obtained by INLA (solid line) and by MCMC (histogram)

Figure 6: Comparison between INLA and MCMC for the robust mixed effect model.



(a) Approximate posterior density for $\log \beta_0$ obtained by INLA (solid line) and by MCMC (histogram)



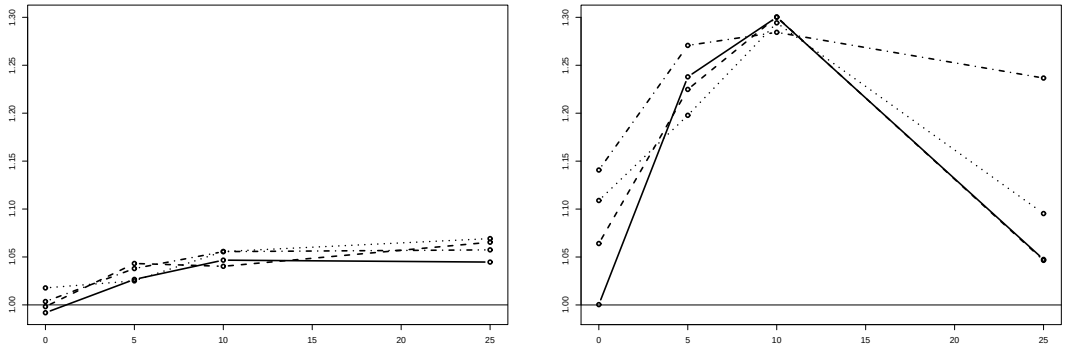
(b) Approximate posterior density for $\log \beta_1$ obtained by INLA (solid line) and by MCMC (histogram)

Figure 7: Comparison between INLA and MCMC for the robust mixed effect model.

more robust than the Gaussian model especially for moderate percentage $(5 - 10\%)$ of outliers. The two methods have about the same efficiency under the no-contamination case.

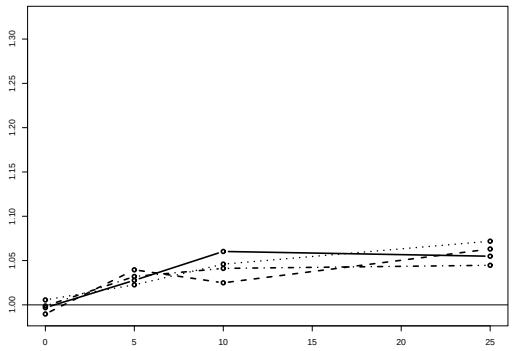
5 Conclusion

This paper extends the INLA method to latent models outside the scope of latent Gaussian models, where independent components of the latent field can have a near-Gaussian distribution. It provides two examples of interest and compare the results obtained by our method with that returned by MCMC, showing similar accuracy with only a small fraction of computational time. We strongly believe that this extension will have a significant impact for the applied community, allowing them to check, for example,

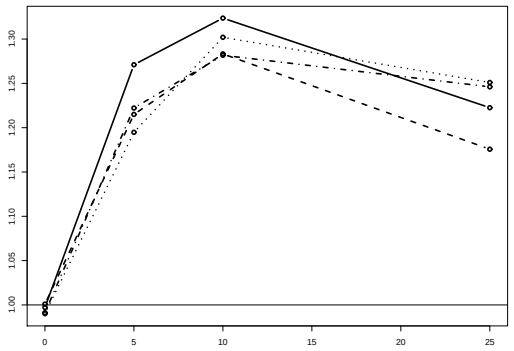


(a) β_0 under close contamination pattern ($f = 2$)

(b) β_0 under distant contamination pattern ($f = 4$)



(c) β_1 under close contamination pattern ($f = 2$)



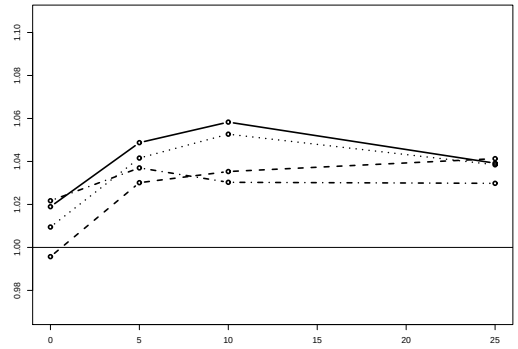
(d) β_1 under distant contamination pattern ($f = 4$)

Figure 8: Relative efficiencies (see Eq. (19)) of the t posterior mean with respect to the Gaussian posterior mean for the fixed effects in the linear mixed effect example. It plots the efficiency on the y-axis and p_e on the x-axis. The meaning for the different types of lines are: (Solid line) $p_b = 0\%$, (Dashed line) $p_b = 5\%$, (Dotted line) $p_b = 10\%$ and (Dot-sash line) $p_b = 25\%$.

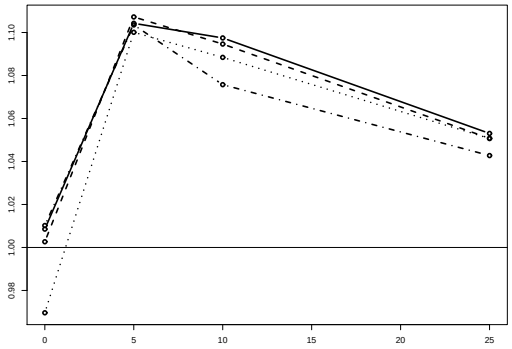
normality assumption that are usually taken for granted due to the lack of user friendly tools to fit more complex models accurately and in a reasonable amount of time.

Further work is required to evaluate which class of distributions can be used in the framework described in this paper with the same level of accuracy as the one displayed in our examples for the log-gamma and Student t distribution. It was our experience that the method gives very accurate results when the non-Gaussian components have distributions not so far from the Gaussian, hence the name near-Gaussian distributions. Distributions that have properties far from a Gaussian, such as those not defined on the real line (e.g. the exponential distribution) are not expected to be well approximated by the method described here.

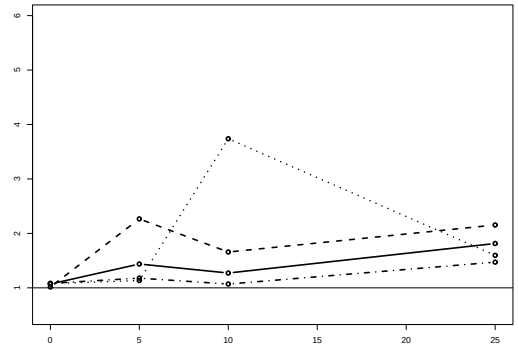
As mentioned in Section 3, it was not our interest to give a precise and restrictive definition of near-Gaussian distributions, and there are mainly two reasons for that. First, it is not easy, if at all possible,



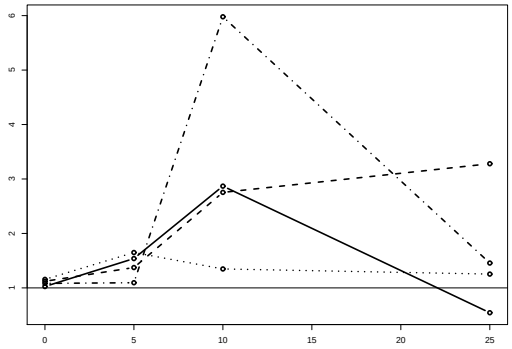
(a) $\tau_e = 1/\sigma_e^2$ under close contamination pattern ($f = 2$)



(b) $\tau_e = 1/\sigma_e^2$ under distant contamination pattern ($f = 4$)



(c) $\tau_b = 1/\sigma_b^2$ under close contamination pattern ($f = 2$)



(d) $\tau_b = 1/\sigma_b^2$ under distant contamination pattern ($f = 4$)

Figure 9: Relative efficiencies (see Eq. (19)) of the t posterior mean with respect to the Gaussian posterior mean for the precision parameters in the linear mixed effect example. It plots the efficiency on the y-axis and p_e on the x-axis. The meaning for the different types of lines are: (Solid line) $p_b = 0\%$, (Dashed line) $p_b = 5\%$, (Dotted line) $p_b = 10\%$ and (Dot-sash line) $p_b = 25\%$.

to come up with a formal definition where our proposed extension would be guaranteed to work for every family of distribution that is contained in such definition, specially because the INLA approach is know to fail in some cases, even without allowing for near-Gaussian components; see the response from the authors in Rue et al. (2009). Second, our main interest is to see how our approximation scheme performs in practice, and for that purpose we believe that a possibly restrict definition here would do more harm than good.

Another set of useful models that can be addressed by the approach described here and deserves further investigation are the class of dynamic models with non-Gaussian error term for the observation and/or system equation, as in Kitagawa (1987) for example. An ongoing project is to use the methodology

described in this paper to develop a systematic approach to model selection that take advantage of the speed of our method to test standard assumptions of a given model. For example, this could be used by applied researchers that routinely use random effects models to verify the standard assumption of Gaussian distributed random effects.

A R code - Survival model

Following is the INLA code used in example 4.1. Please visit `http://www.r-inla.org/` for more information about the R package INLA.

```
# Simulate a dataset
#--------------------
n = 100 # number of groups
m = 10  # number of individuals in the same group
z = runif(n*m) # simulate covariate
eta = 1 + z    # linear predictor
frailty = rgamma(n, 1, 1) # simulate frailties
y = rexp(n*m, rate = rep(frailty, each = m) * exp(eta)) # simulate data

# INLA code
#------------
## Construct an extended response vector Y
yy = inla.surv(c(y, rep(NA, n)),
               c(rep(1, n*m), rep(NA, n))) # Observation component
ff = c(rep(NA, n*m), rep(1, n)) # Frailty component,
                                # any observation will do, like '1'
Y = list(yy, ff) # extended response vector
## Construct extended covariates and frailties
intercept = c(rep(1, n*m), rep(NA, n))       # intercept
zz = c(z,  rep(NA, n))                       # covariate
loggamma.frailty = c(rep(1:n, each=m), 1:n)  # frailty
## Model formula
formula = Y ~ -1 + intercept + zz +
              f(loggamma.frailty, model="iid",
                hyper = list(prec = list(initial=-5, fixed=TRUE)))
## prior for the frailty
hyper.frailty = list(prec = list(param=c(1, 1)))
## Run inla function
rr = inla(formula,
          data = list(Y=Y, zz=zz, intercept=intercept,
                      loggamma.frailty= loggamma.frailty),
```

```
family = c("exponential",  "loggammafrailty"),
control.data = list(list(),  list(hyper = hyper.frailty)),
control.fixed = list(prec = list(default = 0.01)),
control.inla = list(strategy = "laplace")
)
```

References

Gamerman, D. and Lopes, H. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman & Hall/CRC.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian survival analysis*. Springer.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American statistical association*, 82(400):1032–1041.

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Lange, K., Little, R., and Taylor, J. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, pages 881–896.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The bugs project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067.

Pinheiro, J., Liu, C., and Wu, Y. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10(2):249–276.

Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Verlag.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, 71(2):319–392.

Tierney, L. and Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, pages 82–86.