NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

Simultaneous credible bands for latent Gaussian models

by

Sigrunn Holbek Sørbye & Håvard Rue

PREPRINT
STATISTICS NO. 4/2010

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL `http://www.math.ntnu.no/preprint/statistics/2010/S4-2010.pdf`
Håvard Rue has homepage: `http://www.math.ntnu.no/~hrue`
E-mail: `hrue@math.ntnu.no`
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Simultaneous credible bands for latent Gaussian models

Sigrunn Holbek Sørbye
Department of Mathematics and Statistics, University of Tromsø

Håvard Rue
Department of Mathematical Sciences, Norwegian University of Science and Technology

February 26, 2010

Abstract

Deterministic Bayesian inference for latent Gaussian models has recently become available using integrated nested Laplace approximations (INLA). Applying the INLA-methodology, marginal estimates for elements of the latent field can be computed efficiently, providing relevant summary statistics like posterior means, variances and pointwise credible intervals. In this paper, we extend the use of INLA to joint inference and present an algorithm to derive analytical simultaneous credible bands for subsets of the latent field. The algorithm is based on approximating the joint distribution of the subsets by multivariate Gaussian mixtures. Additionally, we present a saddlepoint approximation to compute Bayesian contour probabilities, representing the posterior support of fixed parameter vectors of interest. The given methods are applied to various examples from the literature.

Key words: contour probability, Gaussian mixtures, highest posterior density region, integrated nested Laplace approximation, simultaneous credible bands

1 Introduction

Latent Gaussian models represent a widely applicable class of Bayesian hierarchical models in which a Gaussian field is observed indirectly by conditionally independent observations. This class of models can be viewed as a subclass of structured additive regression models, in which the mean of observations belonging to an exponential family, is linked to a structured additive predictor. The predictor typically includes both linear and smooth effects of covariates and also unstructured random effects. By assigning Gaussian priors to the elements of the latent field, representing the random variables of the predictor, a latent Gaussian model is obtained. Parameters of the prior models for the latent field are referred to as hyperparameters, often being non-Gaussian. The resulting class of models is quite general and includes several familiar statistical models, like generalized linear or additive models, spline models, state-space models and various temporal, spatial or spatiotemporal models. Applications of latent Gaussian models are numerous, see for example Fahrmeier & Tutz (2001), Gelman et al. (2004), Rue et al. (2009) and Martino & Rue (2010).

In Rue et al. (2009), a unified framework for deterministic approximate Bayesian inference of latent Gaussian models was presented, using integrated nested Laplace approximations (INLA). Assuming the latent field to have Markov properties, efficient recursive algorithms based on sparse matrices, provide accurate estimates of posterior marginals for all elements of the latent field and additional hyperparameters. The INLA-methodology is available applying a user-friendly R-interface, see `www.r-inla.org`. For a specific model, relevant summaries of the marginals like posterior means, variances, quantiles and pointwise credible bands, are easily obtained. The main aim of this paper is to extend the use of INLA to deterministic joint inference, deriving analytical simultaneous credible bands. We also discuss Bayesian contour probabilities as a tool for model selection, see Held (2004) and Brezger & Lang (2008), calculating the posterior support of specific vectors of interest.

A simultaneous credible band is a visually appealing exploratory tool in joint inference, constructed as a collection of pointwise credible intervals having a specified coverage probability. Besag et al. (1995) presented an algorithm for calculating simultaneous credible bands based on order statistics of univariate samples obtained by Markov chain Monte Carlo (MCMC). The resulting credible bands were seen to be slightly conservative and due to Monte Carlo error, these credible bands might be unstable, especially for large coverage probabilities. Crainiceanu et al. (2007) and Krivobokova et al. (2009), derived simultaneous credible bands using MCMC for functions of penalized spline models.

The computational benefit of deterministic inference for latent Gaussian models using INLA, compared with sampling-based inference using MCMC, is huge, see Rue et al. (2009). Our algorithm to derive simultaneous credible bands makes use of the INLA-methodology, combined with multivariate Gaussian mixture approximations to joint distributions of subsets of the latent field. Gaussian mixtures represent a flexible class of distributional models, having a wide range of applications for high-dimensional data, see Chen & Tan (2009). In Rue & Martino (2007), Gaussian mixtures were considered as approximations to the marginals of the elements of a latent field. The resulting estimates showed to be quite accurate but slight errors in location and skewness were observerd. Rue et al. (2009) suggested that Gaussian mixtures, possibly including a correction for the mean, could be applied to approximate the joint distribution of small subsets of the latent field. We adopt this procedure also for larger dimensions of a subset, and present a simple iterative algorithm to compute simultaneous credible bands. The algorithm is based on finding individual highest posterior density (HPD) intervals for each component of the relevant subset, scaling the credible level of these intervals to obtain the correct simultaneous coverage probability.

In Held (2004), the use of simultaneous credible bands was questioned, as the resulting region corresponds to a hyper-rectangular credible region, possibly containing vectors that are not well-supported by the posterior distribution. Also, a vector in the tails of the posterior is not necessarily included in the credible band. A relevant alternative is to consider whether a fixed vector is within the HPD-region of a specified level or not, see Box and Tiao (1973, p. 125). The HPD-region is optimal in the sense of having the smallest volume of all credible regions. In Held (2004), the probability

$$p(x^* \mid y) = P(\pi(x \mid y) \leq \pi(x^* \mid y) \mid y), \tag{1}$$

is defined as the posterior contour probability of a fixed vector x^*, given observations y, where the posterior $\pi(x \mid y)$ is considered a random variable. The countour probability equals 1 minus the content of the HPD-region just covering x^*, and can basically be interpreted as p-values in checking the posterior support of specific vectors of interest. In general, the HPD-region is hard to obtain analytically, but the contour probabilities can be calculated using MCMC methods, see Held (2004). In our case, the contour probabilities can be evaluated straightforwardly using Monte Carlo estimation, sampling from the relevant Gaussian mixture. As an alternative, we present a method using the saddlepoint approximation of Lugannani & Rice (1980), in which only the involved fractional moments are approximated using Monte Carlo estimation.

The plan of this paper is as follows. In section 2, we present some background on the INLA-methodology and the Gaussian mixture assumption. The algorithms to calculate simultaneous credible bands and posterior contour probabilities are presented in section 3. In section 4, the given methodology is applied to various examples from the literature. Discussion and concluding remarks are given in section 5.

2 Finite Gaussian mixture approximations to joint distributions

A latent Gaussian model is a hierarchical model in which a latent field x is observed pointwise through conditionally independent observations $y = \{y_i : i \in \mathcal{I}\}$. Given hyperparameters θ, x is assumed to be a zero-mean Gaussian Markov random field, having a sparse precision (inverse covariance) matrix $Q(\theta)$. Generically, denote densities and conditional densities by $\pi(\cdot)$ and $\pi(\cdot \mid \cdot)$, respectively. The posterior distribution of the given model can then be expressed by

$$\pi(x, \theta \mid y) \propto \pi(\theta)\pi(x \mid \theta) \prod_{i \in \mathcal{I}} \pi(y_i \mid x_i, \theta).$$

In order to derive simultaneous credible bands for a subset $x_S \subset x$, we need to estimate the conditional posterior of x_S given y, expressed by

$$\pi(x_S \mid y) = \int \pi(x_S \mid \theta, y)\pi(\theta \mid y) d\theta. \tag{2}$$

The full conditional of the latent field,

$$\pi(x \mid \theta, y) \propto \pi(x \mid \theta) \prod_{i \in \mathcal{I}} \pi(y_i \mid x_i, \theta) \propto \exp\{-\frac{1}{2} x^T Q(\theta) x + \sum_i \log(\pi(y_i \mid x_i, \theta))\}, \tag{3}$$

can often be well approximated by a Gaussian distribution,

$$\tilde{\pi}_G(x \mid \theta, y) \propto \exp\{-\frac{1}{2}(x - \mu)^T Q^*(\theta)(x - \mu)\},$$

matching the mode and curvature by Taylor expansion, see Rue & Martino (2007) and Rue et al. (2009). Notice that in the case of having a Gaussian likelihood, such an approximation will be exact. The presented algorithms are based on applying a Gaussian approximation also to the conditional distribution of the subset x_S, denoted by $\tilde{\pi}_G(x_S \mid \theta, y)$.

Adopting the procedure in Rue & Martino (2007) and Rue et al. (2009), the marginal posterior of θ can be estimated by the Laplace approximation

$$\tilde{\pi}(\theta \mid y) \quad = \quad \left. \frac{\pi(x, \theta \mid y)}{\tilde{\pi}_G(x \mid \theta, y)} \right|_{x = x^*(\theta)}$$

where for each θ, $x^*(\theta)$ represents the mode of the Gaussian approximation. Applying numerical integration, we will approximate (2) by

$$\tilde{\pi}(x_S \mid y) = \sum_{j=1}^{k} w_j \tilde{\pi}_G(x_S; \mu_j, \Sigma_j \mid y), \tag{4}$$

in which $\sum_{j=1}^{k} w_j = \sum_{j=1}^{k} \tilde{\pi}(\theta_j \mid y)\Delta_j = 1$, where Δ_j denotes the area weight corresponding to θ_j. Further, $\tilde{\pi}_G(\cdot; \mu_j, \Sigma_j)$ denotes the multivariate Gaussian approximation with mean $\mu_j = \mu(\theta_j)$ and covariance matrix $\Sigma_j = \Sigma(\theta_j)$. Details concerning the numerical integration, are given in Rue et al. (2009). Basically, two different approaches for numerical integration are implemented, in which a grid approach can be applied to increase the accuracy. Applying the INLA-methodology, all weights, means (possible corrected, see Rue et al. (2009)) and covariance matrices in (4) can be estimated efficiently, facilitating algorithms to obtain simultaneous credible bands and posterior contour probabilities.

3 Simultaneous probability statements for Gaussian mixtures

In general, a $100(1-\alpha)\%$ credible region for a vector x, given observations y, is defined as the fixed subset R of the sample space of x in which the posterior probability

$$P(x \in R \mid y) = \int_R \pi(x \mid y)dx = 1 - \alpha.$$

As with frequentistic confidence regions, credible regions are not uniquely defined and different criteria can be applied in determining various types of credible regions. The simultaneous credible bands, to be derived in section 3.1, represent a visually appealing choice in joint inference. The contour probabilities, derived in section 3.2, give information concerning the optimal HPD-region.

3.1 Simultaneous credible bands

For a random vector $x = (x_1, \ldots, x_m)$, a $100(1-\alpha)\%$ simultaneous credible band can be defined as a hyper-rectangular region

$$R = \{x : \cap_{i=1}^{m}(x_i \in I_{i,\alpha_i})\}, \text{ where } P(x \in R \mid y) = 1 - \alpha,$$

and where $\{I_{i,\alpha_i}\}_{i=1}^{m}$ denotes a set of $100(1-\alpha_i)\%$ individual credible intervals for the elements of x. A natural idea to construct simultaneous credible bands for x, is to find a credible region with the correct coverage probability, also having a minimized volume. We propose to apply HPD-intervals for each component of x, all having the same credible level $1-\gamma$, in which $\gamma \in (\alpha/m, \alpha)$. The resulting simultaneous credible bands can be described by the region

$$R^*(\gamma) = \{x : \cap_{i=1}^{m}(x_i \in I_{i,\gamma})\}, \tag{5}$$

where $\{I_{i,\gamma}\}_{i=1}^{m}$ now denotes a set of $100(1-\gamma)\%$ HPD-intervals and γ is chosen to ensure the correct coverage probability, i.e.

$$P(x \in R^*(\gamma) \mid y) = 1 - \alpha. \tag{6}$$

Based on (4), the individual HPD-intervals for a given value γ, are calculated by a straightforward numerical approach using the marginals

$$\tilde{\pi}(x_i \mid y) \quad = \quad \int \tilde{\pi}(x \mid y)dx_{-i} = \sum_{j=1}^{k} w_j \tilde{\pi}_G(x_i; \mu_{ij}, \sigma_{ij} \mid y),$$

where x_{-i} denotes the vector x without the ith element. Further, μ_{ij} and σ_{ij} denote the jth term mean and variance of x_i, $i = 1, \ldots, m$, $j = 1, \ldots, k$. The correct individual level $1 - \gamma$ can be found by a few iterations, solving (6) with respect to γ, where

$$P(x \in R^*(\gamma) \mid y) = \sum_{j=1}^{k} w_j \int_{R^*(\gamma)} \tilde{\pi}_G(x; \mu_j, \Sigma_j \mid y) dx. \tag{7}$$

The multi-Gaussian probabilities in (7), which is the probability for x being in the hyper-rectangular region $R^*(\gamma)$, can be calculated in `R` using the package `mvtnorm`. This package implements the clever numerical integration algorithm developed for this purpose only, by Genz (1992, 1993). In cases where the dimension of x is large, the computation time in finding the simultaneous credible bands can be significantly reduced using a Gaussian copula approximation in (7). The copula is constructed by

$$C(F_1(x_1), \ldots, F_m(x_m)) = \Phi_m(\Phi^{-1}(F_1(x_1)), \ldots, \Phi^{-1}(F_m(x_m))), \tag{8}$$

where $\Phi_m(\cdot)$ denotes the ordinary cumulative distribution function for a m-dimensional Gaussian vector. The functions $F(\cdot)$ denote the marignal cumulative distribution functions for the Gaussian mixture, i.e.

$$F_i(x_i) \quad = \quad \sum_{j=1}^{k} w_j \int_{-\infty}^{x_i} \tilde{\pi}_G(x_i; \mu_{ij}, \sigma_{ij} \mid y) dx_i, \quad i = 1, \ldots, m.$$

As a result, (7) is evaluated as a single multi-Gaussian probability, in which the marginals equal the original marginals of the Gaussian mixture in (4). In the examples later, the copula approximation is seen to give very accurate results.

In constructing simultaneous credible bands, a Bonferroni correction which only adjusts the credible level according to the dimension of a vector x, is known to be very conservative. The given approach also adjusts for dependency and represents a computationally efficient alternative to the approach in Besag et al. (1995), in which individual intervals of the same level were found using MCMC. Their procedure was based on order statistics, in which the intervals were assumed symmetric and the credible level was defined as the proportion of samples which fell simultaneously in all intervals.

3.2 Posterior contour probabilities

According to Held (2004), simultaneous credible bands should be interpreted with care. Constructed as a projection of a hyper-rectangular region, credible bands might give a misleading impression concerning the support of the posterior, both containing vectors that are not supported or by underestimating the support of vectors located in the tail. In contrast, HPD-regions contain the subset of a random vector x, which is most probable.

Formally, a $100(1 - \alpha)\%$ HPD-region is defined as the subset $R(c_\alpha)$, in which

$$\begin{aligned} R(c_\alpha) &= \{x : \pi(x \mid y) \geq c_\alpha\} \\ P(x \in R(c_\alpha) \mid y) &= 1 - \alpha, \end{aligned}$$

and c_α denotes a relevant constant. The contour probabilities in (1), can be expressed by

$$p(x^* \mid y) = 1 - P(x \in R(c_\alpha) \mid y), \text{ where } c_\alpha = \pi(x^* \mid y),$$

providing information concerning the credible level in which the corresponding HPD-region covers the fixed vector x^*. A special case in which contour probabilities can be found analytically includes the Gaussian case, where the HPD-regions are ellipsoids. Consequently, the contour probabilities can easily be calculated by

$$p(x^* \mid y) \quad = \quad P(\pi_G(x;\mu,\Sigma) \leq \pi_G(x^*;\mu,\Sigma) \mid y) = P(\chi^2_m > (x^*-\mu)^T\Sigma^{-1}(x^*-\mu) \mid y),$$

where χ^2_m denotes a chi-squared variable with m degrees of freedom.

Generally, in cases where the functional form of the posterior is unknown, estimates of contour probabilities can be based on a MCMC-approach, combining Rao-Blackwellization and Monte Carlo estimation, see Held (2004). In our case, the posterior is approximated by the Gaussian mixture (4), and independent samples from this distribution, denoted by $\{x^{(t)}\}_{t=1}^T$, can be obtained efficiently. The contour probabilities can then be calculated unbiasedly using the Monte Carlo estimator

$$\hat{p}_{\mathrm{MC}}(x^* \mid y) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{\tilde{\pi}(x^{(t)} \mid y) \leq \tilde{\pi}(x^* \mid y)\}, \tag{9}$$

originally proposed by Wei & Tanner (1990).

As an alternative to (9), we propose to estimate contour probabilities using the saddlepoint approximation of Lugannani & Rice (1980), combined with Monte Carlo estimation of the involved expectation. To describe our approach, let x denote a given subset of the latent field and regard the density $\tilde{\pi}(x \mid y)$ in (4) as a random variable. The contour probability of a fixed vector x^* is given by

$$p(x^* \mid y) = P(\log\tilde{\pi}(x \mid y) < \log\tilde{\pi}(x^* \mid y) \mid y).$$

Let $K(u)$ denote the cumulant generating function of $\log\tilde{\pi}(x \mid y)$, which equals

$$K(u) = \log(E(e^{u\log\tilde{\pi}(x|y)})) = \log E(\tilde{\pi}(x \mid y)^u). \tag{10}$$

The saddlepoint approximation of the contour probabilities can be expressed by

$$\hat{p}_{\mathrm{SA}}(x^* \mid y) = \Phi(w) + \phi(w)(1/w - 1/v), \tag{11}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard Gaussian density and cumulative distribution functions, resepctively. Further,

$$\begin{aligned} v &= \hat{\psi}(K''(\hat{\psi}))^{1/2} \\ w &= sgn(\hat{\psi})(2(\hat{\psi}\log\pi(x^*) - K(\hat{\psi})))^{1/2}, \end{aligned}$$

in which the saddlepoint $\hat{\psi}$ is found numerically as the solution of $K'(\psi) = \log\tilde{\pi}(x^* \mid y)$.

The saddlepoint approximation of Lugannani & Rice (1980) is known to be very accurate, especially in estimating tail probabilities. In our experience, the new approach requires fewer Monte Carlo samples compared to (9), as the samples are used only to estimate the fractional moments $E(\tilde{\pi}(x \mid y)^u)$ in (10). The derivatives of the cumulant generating function are found numerically. Preferably, we would have liked to calculate the fractional moments analytically. A possible idea to evaluate these expectations is to apply a Taylor series expansion, as the given moments can be calculated exactly for positive integers $u = 1, 2, \ldots$, see the appendix. However, the cost of evaluation of the positive-integer moments grows exponentially in the order, hence this approach is not computationally feasible.

4 Examples

In this section, we reanalyse various examples from the literature using INLA, combined with the given algorithms to calculate simultaneous credible bands and contour probabilities. For the given examples, application of the copula approximation (8) in evaluating (7), visually gives the same estimates for the credible bands as the exact approach. For fixed parameter vectors, the empirical approach in (9) and the saddlepoint approximation in (11), have shown to give quite similar estimates for the contour probabilities. Differences between the two approaches are investigated further in section 4.1.

In exploring different models, the simultaneous credible bands provide a visual impression of the functional form of a given covariate. A relevant question in model selection is whether an assumed nonparametric covariate effect can be modeled for example as a constant or linear function, giving a more parsimonious model. Using the INLA-methodology, commonly applied critera for model selection can be assessed, like the Deviance Information Critera (DIC) of Spiegelhalter et al (2002), the predictive log-score given in Gneiting & Raftery (2007) and the PIT-histograms given in Czado et al. (2009), see Martino & Rue (2010) for several case studies. The posterior contour probabilities provide an additional tool in model selection, basically having the same interpretation as p-values, see Held (2004) and Brezger & Lang (2008) for various applications.

4.1 Childhoood undernutrition in Zambia

We first consider the data set given by Kandala et al. (2001), applying spatial analysis in studying undernutrition among children in a total of 57 regions of Zambia. This example is included as a case study in Martino & Rue (2010) and at the website `www.r-inla.org` and has also been analysed in Kneib et al. (2004) and Brezger & Lang (2008), applying the MCMC-based software `BayesX`.

Undernutrition among children can be measured by stunting, referring to insufficient height for age. In the given data set, stunting is measured by a Z-score

$$Z_i = (AI_i - MAI)/\sigma,$$

where AI refers to a child's anthropometric indicator (the childs height at a certain age), MAI refers to the median of the reference population and σ denotes the standard deviation of the population. Lower values indicate poorer nutritional status. The data set consists of 4846 observations, and the Z-scores are assumed to be conditionally independent Gaussian random variables with mean η_i for child i. The mean is assumed to be linked to a structured additive predictor,

$$\eta_i = \mu + f_1(bmi_i) + f_2(agc_i) + f_s(s_i) + f_u(s_i) + \beta' z, \tag{12}$$

where the functions $f_1(\cdot)$ and $f_2(\cdot)$, represent smooth effects of the mother's body mass index (*bmi*) and the age of the child measured in months (*agc*). These functions are modeled as second-order random walk processes (RW2). The functions $f_s(\cdot)$ and $f_u(\cdot)$ represent structured and unstructured effects of the district s_i, in which a child lives, where the structured effect is modeled as an intrinsic Gaussian Markov random field (IGMRF), specified as the Besag-model (see Ch. 3 in Rue & Held (2005) and `www.r-inla.org` for further details). The model also includes linear effects of categorical covariates z, like gender of the child, the mothers educational level and type of living area.

In exploring more parsimonious model than (12), we consider whether the covariates *bmi* and *agc* can be modeled as constants or linear functions. Having a Gaussian likelihood, the full conditional distribution of the latent field in (3) is a Gaussian distribution. Consequently, applying numerical integration the approximation in (4) where $x_S = (f_1(\cdot), f_2(\cdot))$, is an exact finite Gaussian mixture. The different integration strategies which can be applied using the INLA-methodology are described thorougly in Rue et al. (2009). In calculating simultaneous credible bands, the integration strategy used has not been seen to have any significant impact on the results. Applying the default integration method to the given example, the finite Gaussian mixture consists of $k = 27$ terms. The resulting credible bands, illustrating nutritional status as a function of *bmi* and *agc*, respectively, are given in Fig. 1.

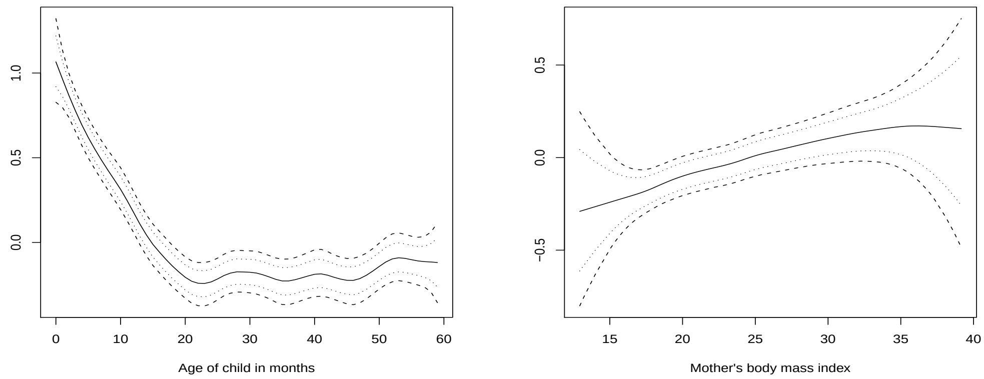

Fig. 1: Estimated effect of *agc* (left) and *bmi* (right) including 95% pointwise and simultaneous credible intervals.

For the age effect, we notice that the nutritional status of children decreases until the age of 20 months, remaining at an almost constant level thereafter. Visually, this effect seems clearly nonlinear. This is supported by the fact that the posterior contour probability is approximately 0 for any linear function, both applying the empirical approach in (9) and the saddlepoint approximation in (11). For fixed vectors, the two approaches to calculate the contour probabilities have been seen to give quite similar results. To illustrate the difference, we have evaluated the tail contour probabilities using (9) and (11) as a function of the argument $\log \tilde{\pi}(x^* \mid y)$, in which the empirical approach is seen to give more wiggly results. This is illustrated for the age effect in Fig. 2, using 2000 independent samples from (4). Increasing the number of samples to 10000, the empirical approach still gives a wiggly result.

A child's nutritional status has been presumed to have an inverse U-shape as a function of the mother's bmi, indicating that both mothers with very low and high bmi, have poorly nourished children, see Kandala et al. (2001). Fig. 1 illustrates that the effect of *bmi* seems almost linear and a wide range of linear functions fit within the given simultaneous credible bands. Visually, the *bmi* effect seems significantly different from 0, as the null-vector is not entirely within the 95% simultaneous credible limits. However, in applying (11), we find that the contour probability for the null-vector is approximately equal to 1, indicating that the *bmi* effect is non-significant.

Table 1 presents the DIC and log-score values for a constant, linear or nonlinear effect of *bmi*. The DIC-value of Spiegelhalter et al (2002), reflects a trade-off between the complexity and the fit of a given model, while the log-score criteria in Gneiting & Raftery (2007),

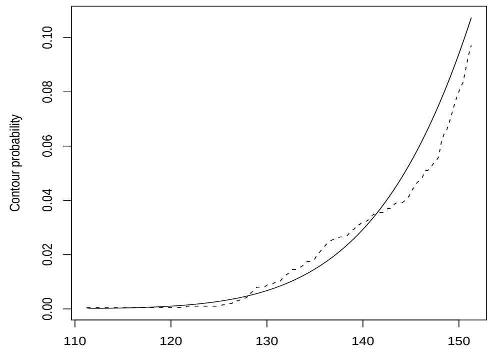

Fig. 2: The estimated contour probability for *agc* as a function of the argument $\log \tilde{\pi}(x^* \mid y)$ using the saddlepoint approximation $\hat{p}_{\mathrm{SA}}(x^* \mid y)$ (solid line) and the direct empirical approach $\hat{p}_{\mathrm{MC}}(x^* \mid y)$ (dashed line). The calculations are based on 2000 independent samples from the Gaussian mixture distribution in (4).

measures the predictive abilities of a given model. For both quantities, lower values are preferable and evaluated by these criteria, the linear model is the best choice, see also Martino & Rue (2010) and Brezger & Lang (2008).

Brezger & Lang (2008) applied a P-spline model and the procedure given in Held (2004) to calculate contour probabilities. For the given example, they estimated the contour probability for a constant effect of *bmi* to be in a medium range (defined as values between 0.1 – 0.4), allowing no clear decision. In comparing the contour probabilities and DIC-values as model selection criteria, Brezger & Lang (2008) noticed that the contour probabilities were more conservative than DIC. Performing a simulation study, conclusions based on DIC were seen to give too complex models in a considerable number of cases.

Model for *bmi*	DIC	log-score
Constant	12758	1.3162
Linear	12735	1.3139
Nonlinear	12740	1.3143

Table 1: The DIC and log-score as model selection criteria for the *bmi* effect.

4.2 Larynx cancer risk in Germany

The next example we consider, concerns spatial variation of larynx cancer risk in relation to a proxy variable for smoking. The data set consists of male larynx cancer mortality counts in $n = 544$ districts of Germany from 1986–1990 and has been analysed both in Natario & Knorr-Held (2003) and Rue & Martino (2007).

The observations are assumed to be conditionally independent Poisson variables,

$$y_i \mid \eta_i \sim \mathrm{Poisson}(E_i \exp(\eta_i)), \quad i = 1, \ldots, n,$$

in which E_i denotes a fixed district effect and η_i is the log-relative risk for district i. The log-relative risk is modeled as sum

$$\eta_i = u_i + v_i + f(c_i),$$

where u_i and v_i denote spatially structured and unstructered terms, respectively. As a covariate, lung cancer rate is used as a proxy for smoking consumption, having value c_i for district i. The function $f(\cdot)$ is modeled as a RW2, and parametrized as unknown values scaled to the interval $[1, 100]$. In addition, the constraint $\sum_i u_i = 0$ is imposed to separate the effect of the covariate, see Rue & Martino (2007). For further details, see the "Bym-example" at `www.r-inla.org`.

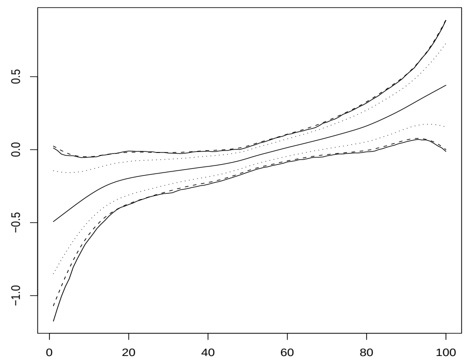

Fig. 3: Estimated effect of a proxy for smoking consumption, including 95% pointwise credible intervals (dotted) and 95% simultaneous credible bands derived analytically (dashed) and by MCMC-sampling (outer solid lines).

In Rue & Martino (2007), estimated quantiles using a Gaussian mixture assumption to the marginals were seen to be very accurate for the given example, but slight errors in skewness were observed. Natario & Knorr-Held (2003) applied MCMC and the software `BayesX`, to derive simultaneous credible bands for the covariate effect, using the method of Besag et al. (1995). In Fig. 3, the 95% credible bands obtained by the presented analytical algorithm are compared with the results using the algorithm of Besag et al. (1995) (outer solid lines), based on a total of 11719 MCMC-samples. The analytical credible bands are seen to be slightly narrower than the limits obtained by sampling and the estimated proportion of samples within the analytically derived credible band is 93.4%. We have applied the default numerical integration strategy in INLA, in which the mixture in (4) consists of $k = 15$ terms for this example.

Model for covariate	DIC	log-score
Constant	2810	2.6121
Linear	2780	2.5731
Nonlinear	2785	2.5837

Table 2: The DIC and log-score as model selection criteria for the proxy covariate of smoking

As for the *bmi* effect in the previous example, we notice that a wide range of linear functions for the covariate effect, fit within the given simultaneous credible bands. Again, the null-vector is not entirely within the limits, but the estimated posterior contour probability is approximately 1, implying that the given covariate is non-significant. Evaluated by the DIC and log-score values, a linear model seems to be the best choice, see Table 2.

4.3 Leukemia survival data in Northwest England

In the last example, we review a data set analysed in Henderson et al. (2002) concerning spatial variation in survival of adult acute myeloid leukemia (AML) patients in the northwestern part of England. Data are registered for $n = 1043$ patients, being diagnosed between 1982 and 1998. Henderson et al. (2002) applied a multivariate frailty approach, including possible spatial variation based on 24 districts. In Henderson et al. (2002), linear predictors were used for the covariate effects, while Kneib & Fahrmeir (2007) analysed the given data set using a mixed model approach, modeling covariates as penalized splines.

We consider the following model in which the survival time of the patients are linked to a predictor

$$\eta_i(t) = \mu + f_0(t) + f_1(wbc_i) + f_2(tpi_i) + f_s(s_i) + \beta' z, \quad i = 1, \ldots, n,$$

where $f_0(t)$ denotes a log-baseline function. The function $f_1(\cdot)$ accounts for the effect of the patients white blood cell counts (*wbc*). The effect of an index named the Townsend deprivation index (*tpi*), is given by $f_2(\cdot)$, in which higher and lower values indicate poorer and richer regions, respectively. The functions $f_1(\cdot)$ and $f_2(\cdot)$ are modeled as first and second-order random processes. The function $f_s(\cdot)$ accounts for the spatial effects of different districts s_i and is specified by the Besag-model. In addition, the patients gender and age are included as linear effects and all of the smooth functions are constrained to sum up to zero.

The numerical integration using the default integration strategy, gives in this case $k = 25$ terms in (4). Fig. 4 illustrates the effect of the different districts and the simultaneous credible bands for the log-baseline function, and the effects of *wbc* and *tpi*. The DIC value for the given model is 5252 and the log-score is equal to 0.9744. Kneib & Fahrmeir (2007) noted that some of the pointwise credible intervals for the districts are strictly positive or negative. However, simultaneously we find that the district covariate is non-significant as the contour probability for the null-vector is approximately 1. In the case of the log-baseline function, we notice that the null-vector is not within the 95% simultaneous credible band and the posterior contour probability of this vector is approximately 0.

	wbc			*tpi*		
Model for covariate	DIC	log-score	$\hat{p}_{\mathrm{SA}}(x^* \mid y)$	DIC	log-score	$\hat{p}_{\mathrm{SA}}(x^* \mid y)$
Constant	5297	0.9821	0.241	5265	0.9768	0.996
Linear	5250	0.9738	1	5255	0.9750	1

Table 3: The DIC and log-score as model selection criteria for the covariates *wbc* and *tpi* including estimated contour probabilities $\hat{p}_{\mathrm{SA}}(x^* \mid y)$.

More parsimonious models can be explored applying constant or linear models for the *wbc* and *tpi* covariates. Table 3 displays the DIC and the log-score values when either *wbc* or *tpi* is modeled as a constant or linear function, also including the estimated posterior contour probabilities $\hat{p}_{\mathrm{SA}}(x^* \mid y)$ for these cases. The contour probabilities indicate that the *tpi* effect is non-significant, while a linear effect for the *wbc* is clearly sufficient. Evaluated by the DIC and log-score criteria, a linear model for *wbc* and a nonparametric model for *tpi* give the lowest values. If both covariates are modeled by linear functions, the DIC-value equals 5254 and the log-score equals 0.9745, being about the same results as using nonparametric functions for both effects.

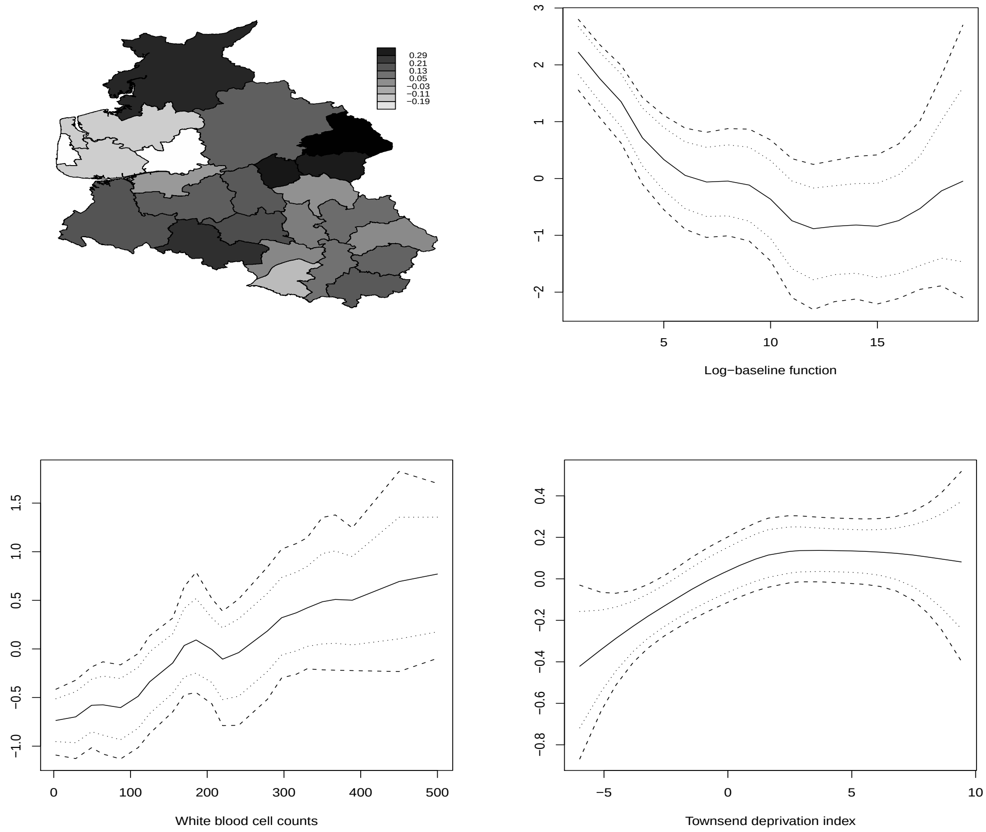

Fig. 4: Estimated effect of the 24 different districts is shown in the upper left panel. The other panels illustrate the effects of the log-baseline function (upper right), the white blood cell counts (lower left) and the Townsend deprivation index (lower right), including 95% pointwise and simultaneous credible bands.

5 Discussion and concluding remarks

Deterministic Bayesian inference using the INLA-methodology of Rue et al. (2009), has become a preferable choice in analysing latent Gaussian models, outperforming MCMC-based techniques in terms of computational speed and accuracy. Applying INLA, the given paper illustrates how simultaneous credible bands can be derived analytically, representing an alternative to the traditional MCMC sample-based technique of Besag et al. (1995). The algorithm makes use of the Laplace approximation in estimating the marginals of hyperparameters and the numerical integration scheme used in INLA. However, to make the resulting algorithm analytically tractable, Gaussian approximations (possible corrected) to the joint marginals of subsets of the latent field, are used. As a result, simultaneous credible bands can be estimated applying a two-step procedure: First, pointwise HPD-intervals are calculated using a common credible level. Secondly, the pointwise level is adjusted to obtain the correct simultaneous coverage probability. The resulting algorithm converges in a few iterations.

In Rue et al. (2009), the Laplace approximation was used to estimate the marginals of the latent field, also suggesting a computationally faster simplified version. The Laplace approximation or its simplified version, could be used also to estimate the joint marginal for a subset of the latent field. Alternatively, Rue et al. (2009) suggest a Gaussian copula approximation to the marginals of subsets with small dimensions. The practicalities in implementing these approaches to construct simultaneous credible bands, seem rather tedious and the fact that simultaneous credible bands are mainly used as a visual exploratory tool, emphasizes the need for a computationally efficient algorithm.

Ideally, we would like to also obtain the posterior contour probabilities analytically but we have not succeeded doing so. However, applying the saddlepoint approximation, we use Monte Carlo estimates of the involved fractional moments using independent samples from the Gaussian mixture distribution. In our experience, the suggested method requires fewer samples than a direct empirical approach and gives a more stable estimate for the contour probability as a function of its argument. For model selection, decisions based on the estimated contour probabilities have been seen to be conservative. Fixed vectors that are not included in the simultaneous credible bands, are seen to be supported by the posterior. Also, decisions based on contour probabilities result in more parsimonious models, than basing the decision on the DIC and log-score criteria.

References

Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* **10**, 3–66.

Brezger, A. & Lang, S. (2008). Simultaneous probability statements for Bayesian P-splines. *Stat. Model.* **8**, 141–168.

Chen, J. & Tan, X. (2009). Inference for multivariate normal mixtures. *J. Multivariate Anal.* **100**, 1367–1383.

Crainiceanu, C. M., Ruppert, C., Carroll, J., Joshi, A. & Goodner, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *J. Comput. Graph. Statist.* **16**, 265–288.

Czado, C., Gneiting, T. & Held, L. (2009). Predictive model assessment for count data. *Biometrics* **65**, 1254–1261.

Fahrmeier, L. & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*, 2nd edn. Springer, Berlin.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edn. Chapman & Hall, Boca Raton.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.* **1**, 141–150.

Genz, A. (1993). Comparison of methods for computation of multivariate normal probabilities. *Computing Science and Statistics* **25**, 400–405.

Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *J. Amer. Statist. Assoc.* **102**, 359–378.

Held, L. (2004). Simultaneous posterior probability statements from Monte Carlo output. *J. Comput. Graph. Statist.* **13**, 20–35.

Henderson, R., Shimakura, S. & Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *J. Amer. Statist. Assoc.* **97**, 965–972.

Kandala, N. B., Lang, S., Klasen, S. & Fahrmeier, L. (2001). Semiparametric analysis of the socio-demographic and spatial determinants of undernutrition in two African countries. *Research in Official Statistics* **1**, 81–100.

Kneib, T., Lang, S. & Brezger, A. (2004). Bayesian semiparametric regression based on MCMC techniques: A tutorial.

Kneib, T. & Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scand. J. Statist.* **34**, 207–228.

Krivobokova, T., Kneib, T. & Claeskens, G. (2009). Simultaneous confidence bands for penalized spline estimators. *Technical report.* University of Göttingen.

Lugannani, R. & Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. in Appl. Probab.* **12**, 475–490.

Martino, S. & Rue, H. (2010). Case studies in Bayesian computation using INLA. In *Complex data modeling and computationally intensive statistical methods* (eds P. Mantovan & P. Secchi). Springer-Verlag.

Natario, I. & Knorr-Held, L. (2003). Non-parametric ecological regression and spatial variation. *Biom. J.* **45**, 670–688.

Rue, H. & Held, L. (2005). *Gaussian Markov random fields: Theory and applications*, Chapman & Hall, London.

Rue, H. & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *J. Statist. Plann. Inference* **137**, 3177–3192.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. Roy. Statist. Soc. Ser. B* **71**, 319–392.

Spiegelhalter, D., Best, N., Bradley, P. & van der Linde, A. (2002). Bayesian measure of model complexity and fit (with discussion). *J. Roy. Statist. Soc. Ser. B* **64**, 583–639.

Wei, G. C. G. & Tanner, M. A. (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika* **77**, 649–652.

Corresponding author: Sigrunn Holbek Sørbye, Department of Mathematics and Statistics, University of Tromsø, N-9037 Tromsø, Norway. E-mail: sigrunn.sorbye@uit.no

Appendix

The positive integer-moments of a random variable $\pi(x) = \sum_{i=1}^{k} w_i \pi_G(x; \mu_i, \Sigma_i)$ can be evaluated applying the following identities for the multi-Gaussian distribution:

$$\begin{aligned}
\pi_G(x;\mu_i,\Sigma_i)\pi_G(x;\mu_j,\Sigma_j) &= \pi_G(0;\mu_i-\mu_j,\Sigma_i+\Sigma_j)\pi_G(x;\mu_{ij},\Sigma_{ij}) \\
\int \pi_G(x;\mu_i,\Sigma_i)\pi_G(x;\mu_j,\Sigma_j)dx &= \pi_G(0;\mu_i-\mu_j,\Sigma_i+\Sigma_j)
\end{aligned}$$

where

$$\begin{aligned}
\mu_{ij} &= \Sigma_j(\Sigma_i+\Sigma_j)^{-1}\mu_i + \Sigma_i(\Sigma_i+\Sigma_j)^{-1}\mu_j \\
\Sigma_{ij} &= \Sigma_i(\Sigma_i+\Sigma_j)^{-1}\Sigma_j.
\end{aligned}$$

A general expression for the mth moment is then given by

$$\begin{aligned}
E(\pi(x)^m) &= \sum_{\alpha_1=1}^{k} w_{\alpha_1} \dots \sum_{\alpha_{m+1}=1}^{k} w_{\alpha_{m+1}} \int \pi_G(x;\mu_{\alpha_1},\Sigma_{\alpha_1}) \cdots \pi_G(x;\mu_{\alpha_{m+1}},\Sigma_{\alpha_{m+1}})dx \\
&= \sum_{\alpha_1=1}^{k} w_{\alpha_1} \sum_{\alpha_2=1}^{k} w_{\alpha_2} f_{\alpha_1,\alpha_2} \sum_{\alpha_3=1}^{k} w_{\alpha_3} f_{\alpha_1,\dots,\alpha_3} \cdots \sum_{\alpha_{m+1}=1}^{k} w_{\alpha_{m+1}} f_{\alpha_1,\dots,\alpha_{m+1}}
\end{aligned}$$

where

$$f_{\alpha_1,\dots,\alpha_j} = \pi_G(0, \mu_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}} - \mu_{\alpha_j}, \Sigma_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}} + \Sigma_{\alpha_j}), \quad j > 2.$$

The expressions for the means and covariance matrices, for $j > 2$, can be calculated by

$$\begin{aligned}
\mu_{(\alpha_1,\dots,\alpha_{j-1}),\alpha_j} &= \Sigma_{\alpha_j}(\Sigma_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}} + \Sigma_{\alpha_j})^{-1}\mu_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}} \\
&\quad + \Sigma_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}}(\Sigma_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}} + \Sigma_{\alpha_j})^{-1}\mu_{\alpha_j} \\
\Sigma_{(\alpha_1,\dots,\alpha_{j-1}),\alpha_j} &= \Sigma_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}}(\Sigma_{(\alpha_1,\dots,\alpha_{j-2}),\alpha_{j-1}} + \Sigma_{\alpha_j})^{-1}\Sigma_{\alpha_j}.
\end{aligned}$$

Alternatively, the expression for the moments can be summarized by

$$E(\pi(x)^m) = \sum_{\alpha_i=1}^{k} w_{\alpha_1}\pi_G(0;\mu_{\alpha_1},\Sigma_{\alpha_1}) \dots \sum_{\alpha_{m+1}=1}^{k} w_{\alpha_{m+1}}\pi_G(0;\mu_{\alpha_{m+1}},\Sigma_{\alpha_{m+1}}) \frac{1}{\pi_G(0,\mu_\alpha,\Sigma_\alpha)}$$

where

$$\alpha = \{(\alpha_1,\dots,\alpha_{m+1})\}, \quad \Sigma_\alpha = (\sum_{i\in\alpha}\Sigma_i^{-1})^{-1}, \quad \mu_\alpha^T = (\sum_{i\in\alpha}\mu_i^T\Sigma_i^{-1})\Sigma_\alpha.$$