Integrated Nested Laplace Approximations

Regina, Tian and Taddele

McMaster University

November 4, 2015

Contents

Introduction

- In almost all uses of statistics,
  1. major interest centres on inference about characteristics of a population through observations made in a representative sample from that population.
     - Estimation
     - Hypothesis testing
     - Prediction
  2. Two approaches: Frequentist and Bayesian.
     - Frequentists: An orthodox view that sampling is infinite and decision rules can be sharp.
     - Bayesians: Unknown quantities are treated probabilistically and the state of the world can always be updated

Review Linear Models

1. Linear regression: GAM
   $y \sim N(\mu, \sigma^2)$
   $\boldsymbol{\mu} = \mathbf{X}_i^T \beta + \epsilon$
2. Generalized Linear model: $y \sim Exponetialfamily$
   - The three components of generalized linear model:

   i Distributions of response variables:
   $f(y_i; \theta_i) = exp\{y_i b(\theta_i) + c(\theta_i) + d(y_i)\}$ and denoting
   $\boldsymbol{\mu}_i = E(\mathbf{Y_i})$

   ii A linear predictor:

   $$g(\boldsymbol{\mu}) = \boldsymbol{\eta}_i = \mathbf{X_i}^T \boldsymbol{\beta}, \text{ where } \mathbf{X} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} \text{ and } \boldsymbol{\beta} = (\beta_1, ..., \beta_p)$$

   iii A monotonic link function: $g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{X_i}^T \boldsymbol{\beta}$

Review...

3. Generalized Additive Model: $y \sim exponenetialfamily$
   $\boldsymbol{\mu} = E(\mathbf{y})$
   $g(\boldsymbol{\mu}) = \alpha_0 + f_1(x_1) + f_2(x_2) + ... + f_p(x_p)$
   i. incorporate non-linear forms of the predictors
   ii. the linear predictor incorporates smooth functions f(x) of at least some (possibly all) covariates
   iii. maximize the quality of prediction of a dependent variable Y from various distributions, by estimating unspecific functions of the predictor variables which are "connected" to the dependent variable via a link function.
4. Generalized Linear Mixed Models:
   i. Mixed models include both the usual fixed effects and random effects in the linear predictor. That is
      $\boldsymbol{\eta} = \mathbf{X}\beta + \mathbf{Z}\gamma$
   ii. the likelihood function under a GLMM typically involves integrals with no analytic expressions, and therefore is difficult to evaluate.

Latent Gaussian Models

- Core idea: Unobserved multivariate Gaussian random variable $\mathbf{x}$, with density $\pi(\mathbf{x}|\theta)$.
- The observed data are assumed to be conditionally independent given the latent field $\mathbf{x}$, ie. $\pi(\mathbf{x}, \theta|\mathbf{y}) \propto \pi(\theta)\pi(\mathbf{x}|\theta) \prod \pi(y_i|x_i)$.
- Main interest for inference: posterior marginals for x_i and possibly, posterior marginals of θ or some θ_i.
- A wide range of models well known from the literature can be formulated as special cases of latent Gaussian models, for example: generalised additive models, generalised additive mixed models, geoadditive models, univariate and multivariate stochastic volatility models.

Latent...

- Hierarchical models are used when the data are structured in groups. e.g. demographically, temporally, spatially
- Latent Gaussian can be represented by a hierarchical structure containing three stages.
  1. The first stage is formed by conditionally independent likelihood function. That is, $\pi(\mathbf{y}|\mathbf{x}, \theta) = \prod_{i=1}^{n} \pi(y_i|\eta_i, \theta)$, where $\mathbf{y}$ is vector of response variable, $\mathbf{x}$ is latent field, θ is hyper-parameter vector, and η is linear predictor).
  2. Second stage is formed by the latent Gaussian distribution with mean $\mu(\theta)$ and precision matrix $\mathbf{Q}(\theta)$ to the latent field conditional on the hyper-parameter. That is, $\mathbf{x}|\theta \sim N(\mu(\theta), \mathbf{Q}(\theta)^{-1})$.
  3. The third stage is formed by the posterior distribution assigned to the hyper-parameters. That is $\theta \sim \pi(\theta)$.

Latent...

Many, but not all, the latent Gaussian models are assumed to satisfy two basic properties:

- The dimension of $\mathbf{x}$ is very large, admits conditional independence properties, i.e., the precision matrix $\mathbf{Q}$ is a sparse matrix. $(10^2 \sim 10^5)$. (Rue and Held, 2005)
- $m = vec(\boldsymbol{\theta})$ is very small, say ≤ 6.

Both properties are required to produce fast inference (Eidsvik et al, 2008).

Review Bayesian

- Bayesian inference is based on computing the posterior distribution of a vector of the model parameters $\boldsymbol{\theta}$ conditional on the vector of the observed data $\mathbf{y}$
- Posterior distribution can be written as:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{1}$$

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2}$$

  $\pi(\mathbf{y}|\boldsymbol{\theta})$: likelihood of the model
  $\pi(\boldsymbol{\theta})$: prior distribution of the model parameters
- Usually $\pi(\boldsymbol{\theta}|\mathbf{y})$ is highly multivariate $\Rightarrow$ difficult to obtain. Rarely computed in closed form
- Here is where computational approaches are needed.
- Bayesian methods are becoming increasingly popular as techniques for modelling "systems" since the advent of simulation-based techniques (notably MCMC).

Review...

Broadly speaking, there are three general steps to Bayesian data analysis:

1. Setting up of a full joint probability distribution for both observable, $\mathbf{y}$ and parameters, $\boldsymbol{\theta}$;

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

2. Update knowledge about the unknown parameters by conditioning this probability model on observed data, $\pi(\boldsymbol{\theta}|\mathbf{y})$
3. Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assumptions.

Review...

- Three key quantities of interest are:
  - i Prior predictive, $\pi(y)$: The normalizing constant $\pi(y)$ in Bayes Theorem is a very important quantity defined by:

$$\pi(y) = \int \pi(\mathbf{y}, \boldsymbol{\theta}) d\theta = \int \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

  - ii Marginal effects of a subset of parameters.

$$\pi(\theta_i|\mathbf{y}) = \int \pi(\theta_i, \theta_{-i}|\mathbf{y}) d\boldsymbol{\theta}_{-i} = \int \pi(\theta_i|\theta_{-i}, \mathbf{y})\pi(\theta_{-i}|\mathbf{y}) d\boldsymbol{\theta}_{-i}$$

  - iii Posterior predictions: Let $\tilde{\mathbf{y}}$ denote some future unobserved response of the system, then the posterior predictive $\pi(\tilde{\mathbf{y}}, \mathbf{y})$ is:

$$\pi(\tilde{\mathbf{y}}, \mathbf{y}) = \int \pi(\tilde{\mathbf{y}}|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \int \pi(\tilde{\mathbf{y}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$$

    $(\tilde{\mathbf{y}}, \mathbf{y})$ are conditionally independent given $\boldsymbol{\theta}$; but clearly, $\pi(\tilde{\mathbf{y}}, \mathbf{y})$ are dependent.

Computational methods

Marckov Chain Monte Carlo Simulation (MCMC): the most widely used method to estimate posterior distribution

- Their efficiency could be limited by complexity of the model (eg hierarchical models).
- Possible solutions:
  - More complex model specification (eg Blocking)
  - More complex sampling schemes
    (eg Hamiltonian Monte Carlo)
  - Alternative methods of inference
    (eg Approximate Bayesian Computation (ABC), Integrated Nested Laplace Approximation (INLA))

Introduction to INLA

- Integrated nested Laplace approximation (INLA) is a computational approach to do statistical inference for Latent Gaussian Markov Random field (GMRF) models, which is introduced by Rue and Martino (2007).
- Issues of convergence and mixing that are inherent to MCMC are no more problems with INLA.
- Perform fast Bayesian inference in the broad class of latent Gaussian models.
- The concept of LGM is intended for the modelling stage in a unified way using algorithm and software tool.

latent Gaussian Markov Random field (LGMRF) models

- A latent GMRF model is a hierarchical model
  - First: assume a probability model for the observations $\mathbf{y}$ given some latent parameters $\mathbf{x}$ and some additional parameters $\boldsymbol{\theta}$

$$\mathbf{y}|\mathbf{x},\boldsymbol{\theta} \sim \pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = \prod_j \pi(y_j|x_j,\boldsymbol{\theta}) \tag{3}$$

  - Second: assume the hyper-parameters are described by a Gaussian Markov Random field

$$\mathbf{x}|\boldsymbol{\theta} \sim \mathrm{Normal}(\mathbf{0}, Q(\boldsymbol{\theta})) \tag{4}$$

$$\mathbf{x}_l \perp\!\!\!\perp \mathbf{x}_m|\mathbf{x}_{-lm} \tag{5}$$

  - Third: $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.
- In general, we can partition $\boldsymbol{\theta}$ as $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, and re-express the model as follows:

$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$	hyper prior
$\mathbf{x}|\boldsymbol{\theta} \sim \pi(\mathbf{x}|\boldsymbol{\theta}) = \mathrm{Normal}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_1))$	GMRF prior
$\mathbf{y}|\mathbf{x},\boldsymbol{\theta} \sim \prod_j \pi(y_j|x_j,\boldsymbol{\theta}_2)$	data model

LGMRF (general framework)

- $\mathbf{y} = \{y_i\}_{i=1}^{n}$ is assumed to belong to an exponential family
- A very general way of formulating this problem is by modelling the mean μ_i for the i-th unit by means of a structured additive linear predictor η_i through a link function $g(.)$, $g(\mu_i) = \eta_i$

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

  where,
  - α is the intercept
  - $f^{(j)}$ is a set of unknown functions of the covariates $\mathbf{u}$
  - β_k's represent the linear effect of variates $\mathbf{z}$
  - ϵ_i's are unstructured terms
- Denote the vector of all latent Gaussian variables to be $\mathbf{x} = (\{\eta_i\}, \alpha, \{f^{(j)}\}, \{\beta_k\}) \sim \mathrm{Gaussian}(\mathbf{0}, Q(\boldsymbol{\theta}_1))$
  - where θ_1 is vector of hyper-parameters.

list of all GMRF model for reference

Dynamic models/ State space models

$$y_t = F_t x_t + v_t \tag{6}$$

$$x_t = G_t x_{t-1} + w_t \tag{7}$$

$$\tag{8}$$

where

$$v_t \sim N(0, V_t) \tag{9}$$

$$w_t \sim N(0, W_t) \tag{10}$$

y_t is a time sequence of scalar observations and x_t is a sequence of state (latent) parameters describing locally the system. F_t is a vector of explanatory variables, while G_t represents a matrix describing the state evolution.

Aim

- The INLA approach provides a fast way to do Bayesian inference using accurate approximations to $\pi(x_i|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$ for $\forall$ i, and j.

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j} \tag{11}$$

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \tag{12}$$

  where $\boldsymbol{\theta}_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots)$

- The key feature is to use this form to construct nested approximations

$$\tilde{\pi}(\theta_j|\mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j} \tag{13}$$

$$\tilde{\pi}(x_i|\mathbf{y}) = \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \tag{14}$$

  where $\tilde{\pi}(.|.)$ is an approximated conditional density of its arguments

Steps in the INLA project

1. Find a Laplace approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$. detail ▸ LA
2. Find an approximation to $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ detail
   - Gaussian approximation
     - fast but inaccurate
   - Laplace approximation
     - accurate but computationally demanding
   - Simplified Laplace approximation
     - default in R-INLA
     - trade-off between speed and accuracy
3. Numerical integration to $\tilde{\pi}_i(x_i|\mathbf{y})$ detail

Step 1: Find a Laplace approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$

- $\pi(\boldsymbol{\theta}|\mathbf{y})$ can be easily obtained by

$$
\begin{aligned}
\pi(\boldsymbol{\theta}|\mathbf{y}) &= \frac{\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (15)\\
&= \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\pi(\mathbf{y})}\frac{1}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (16)\\
&= \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{y})}\frac{1}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (17)\\
&\propto \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (18)\\
&\approx \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}\bigg|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \doteq \tilde{\pi}_{LA}(\boldsymbol{\theta}|\mathbf{y}) && (19)
\end{aligned}
$$

where

- $\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\mathbf{x}$, for a given $\boldsymbol{\theta}$
- $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$ is the Gaussian approximation of $\tilde{\pi}(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$

back

Step 2: Find an approximation to $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$

The marginals for components x_i of the latent field

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \tag{20}$$

- this step is a bit complex than step 1 because in general $\mathbf{x}$ contain more elements than $\boldsymbol{\theta}$.

Gaussian Approximation

- directly using normal distribution

$$\pi(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}|Q|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}-\mu)'Q(\mathbf{x}-\mu)} \tag{21}$$

- based on the Cholesky decomposition of the precision matrix $Q(.) = LL'$
- fast but inaccurate due to lack of skewness (Rue and Martino, 2007)

back

Step 2...

Laplace Approximation

$$
\begin{aligned}
\pi_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) &= \frac{\pi(\{x_i, \mathbf{x}_{-i}\}|\boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} && (22) \\
&= \frac{\pi(\{x_i, \mathbf{x}_{-i}\}, \boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \frac{1}{\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} && (23) \\
&\propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \propto \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{x}_{-i}, \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{y})} && (24) \\
&\propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(x_i,\boldsymbol{\theta})} && (25) \\
&\doteq \tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) && (26)
\end{aligned}
$$

where $\tilde{\pi}_G$ is the Gaussian approximation to $\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}$; and $(\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta}))$ is the modal configuration. Note that

- very accurate as $\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}$ is reasonably normal
- computationally expensive

Step 2...

Simplified Laplace Approximation.

- based on a Taylor's series expansion up to the third order of both numerator and denominator for $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$. i.e. calculate

$$log \tilde{\pi}_{SLA}(x_i^s|\boldsymbol{\theta}, \mathbf{y}) = const. - \frac{1}{2}(x_i^s)^2 + \gamma_i^1(\boldsymbol{\theta}) x_i^s + \frac{1}{6}(x_i^s)^3 \gamma_i^3(\boldsymbol{\theta}) + \ldots \tag{27}$$

  And fit a skew normal density
- this effectively corrects the Gaussian approximation for location and skewness to increase the fit to the required distributions
- implemented by default by R-INLA

back

Step 3: Numerical integration

Integration with respect to $\boldsymbol{\theta}$

$$\tilde{\pi}_i(x_i|\mathbf{y}) = \int \pi_i(x_i|\boldsymbol{\theta}_k, \mathbf{y})\pi_i(\boldsymbol{\theta}_k|\mathbf{y})d\boldsymbol{\theta} \quad (28)$$

$$= \sum_k \tilde{\pi}_i(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}_i(\boldsymbol{\theta}_k|\mathbf{y})\triangle_k \quad (29)$$

where the sum is over all values of $\boldsymbol{\theta}$ with area-weights $\triangle_k$.

Find suitable set of integration point for the numerical integration.

Grid strategy

1. find the mode $\tilde{\boldsymbol{\theta}}$ by optimising $\log\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$
   - Newton-like algorithm
2. Compute the Hessian at the modal configuration $\boldsymbol{\theta}^*$
3. Explore $\log\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$ using $\mathbf{z}$-parametrisation
   - Define $\boldsymbol{\theta}$ through $\mathbf{z}$ using $\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$. If $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is a Gaussian density, then $\mathbf{z}$ is $\mathrm{N}(\mathbf{0}, \mathbf{I})$; and $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$ be the eigen-decomposition of $\boldsymbol{\Sigma}$.
   - produce a grid of points $\{\theta_k^*\}$ associated with area weights $\{\triangle_k\}$

Example: The procedure when $\log \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is unimodal

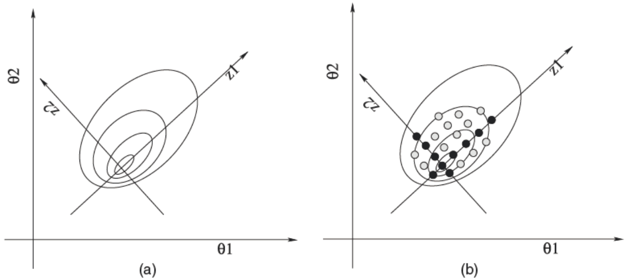

Fig. Illustration of the exploration of the posterior marginal for θ: in (a) the mode is located and the Hessian and the co-ordinate system for **z** are computed; in (b) each co-ordinate direction is explored (•) until the log-density drops below a certain limit; finally the new points (∘) are explored

This re-parametrisation corrects for scale and rotation as well as simplifies numerical integration.

Step 3...

Find suitable set of integration point for the numerical integration.

Central composite design

- Use small amount of support points in the m-dimensional space of $\boldsymbol{\theta}$
- Augment each center point with a group of points used to estimate the curvature of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$
- implemented by default R-INLA

INLA algorithm (operationally)

1. marginal joint posterior for the hyper-parameters $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$
   - find the mode $\tilde{\boldsymbol{\theta}}$ by optimising $\log \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$
     - Newton-like algorithm
   - compute the Hessian at $\boldsymbol{\theta}^*$
   - explore $\log \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$ using **z**-parametrisation
     - produce a grid of points $\{\theta_k^*\}$ associate with mass and area weights $\{\triangle_k\}$
2. for each element $\{\theta_k^*\}$ in the grid
   - find the marginal posterior $\tilde{\pi}(\theta_k^*|\mathbf{y})$
   - evaluate the conditional posterior $\tilde{\pi}(x_j|\theta_k^*, \mathbf{y})$
3. obtain the marginal posterior $\tilde{\pi}_i(x_i|\mathbf{y})$ using numerical integration

$$\tilde{\pi}_i(x_i|\mathbf{y}) = \sum_k \tilde{\pi}_i(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}_i(\boldsymbol{\theta}_k|\mathbf{y})\triangle_k \tag{30}$$

where the sum is over values of $\boldsymbol{\theta}$ with area-weights $\triangle_k$.

A Toy Example (Ruiz et al (2010))

- Simulated example: first order univariate dynamic linear model
- The observational and the system equations respectively of the model is given as:

$$y_t = x_t + v_t, \quad v_t \sim N(0, V), \quad t = 1, ..., n \tag{31}$$

$$x_t = x_{t-1} + w_t, \quad w_t \sim N(0, W), \quad t = 2, ..., n \tag{32}$$

  - assuming $F_t = G_t = 1$, $V_t = V$ and $W_t = W$, for all t
  - vector of hyperparameters is given by $\theta = \{V, W\}$
  - latent field corresponds to $\mathbf{x} = \{x_1, ..., x_n\}$
- Generic approach consists in equating to zero the system equation

$$0 = x_t - x_{t-1} + w_t, \quad w_t \sim N(0, W), \quad t = 2, ..., n \tag{33}$$

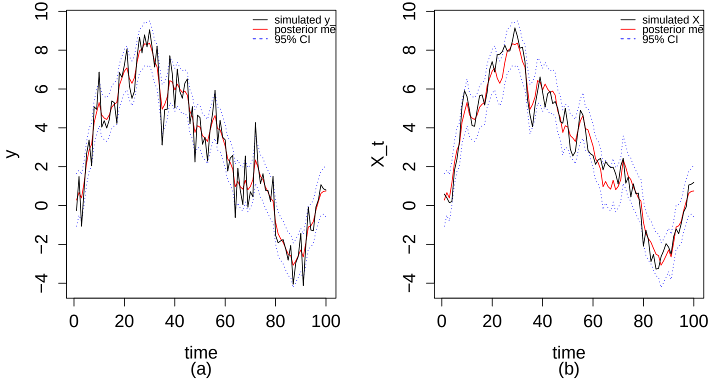

Figure 1 : Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and states (b) in the toy example

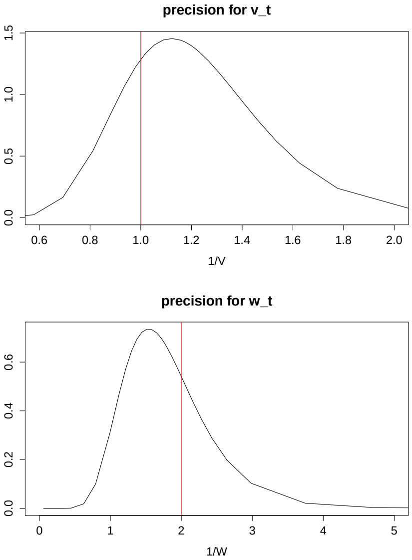

Figure 2 : Posterior densities for the hyperparameters in the toy example. Red lines indicate true simulated values

Example (Ruiz et al (2010))

- Simulated data from a multiple Poisson model with two regressors, Z_{1t} and Z_{2t}
- The model has the following observational and System equations

$$(y_t|\mu_t) \sim Poisson(\mu_t) \tag{34}$$

$$\begin{aligned}
log(\mu_t) &= \lambda_t = \beta_{0_t} + \beta_{1_t} Z_1 + \beta_{2_t} Z_2, \;\; t = 1, ..., n \\
\beta_{0t} &= \beta_{0,t-1} + \omega_{0t}, \;\; \omega_{0t} \sim N(0, W_0), t = 2, ..., n \\
\beta_{1t} &= \beta_{1,t-1} + \omega_{1t}, \;\; \omega_{1t} \sim N(0, W_1), \;\; t = 2, ..., n \\
\beta_{2t} &= \beta_{2,t-1} + \omega_{2t}, \;\; \omega_{2t} \sim N(0, W_2), \;\; t = 2, ..., n
\end{aligned}$$

The linear predictor is given by $\lambda_t = F_t x_t$, where $F_t = (1, Z_{1t}, Z_{2t})$ and the regression coefficients $x_t = (\beta_{0t}, \beta_{1t}, \beta_{2t})$

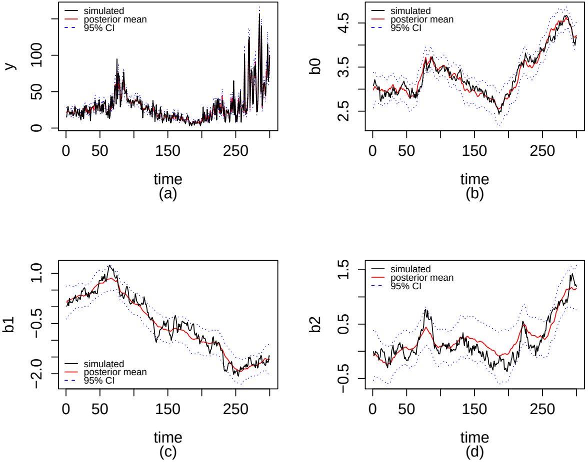

Figure 3 : Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and regression coefficients, β_0, β_1 and β_2 (b-d) in the generalized dynamic regression example

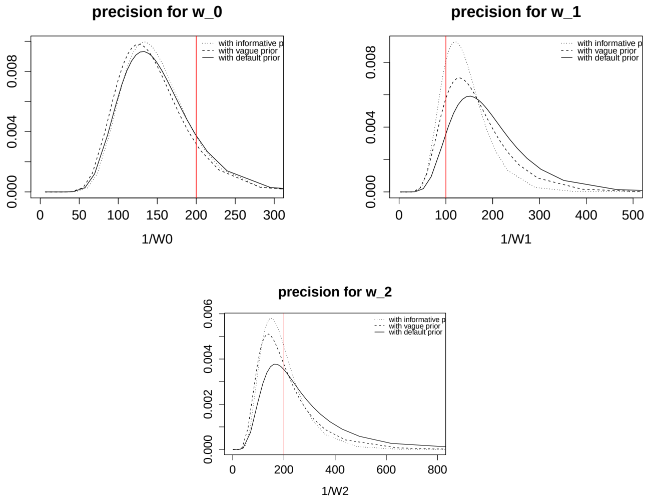

Figure 4 : Posterior densities for the hyperparameters in the generalized dynamic regression example. Red lines indicate true simulated values

REFERENCES

- Ruiz-Cárdenas, R., Krainski, E. T., and Rue, H. (2010). Fitting dynamic models using integrated nested Laplace approximations-INLA. *Preprint Statistics*, **(12)**.
- 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**.

Thank you!