Integrated Nested Laplace Approximations

Regina, Tian and Taddele

McMaster University

October 26, 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 I

1. Linear regression:
   $y \sim N(\mu, \sigma^2)$
   $\boldsymbol{\mu} = \mathbf{X}_i^T \beta + \epsilon$
2. Generalized Linear model: $y \sim Exponetialfamily$
   - Generalized linear model (a unified framework): Random effects, Hierarchical models, Missing variables, Nested and Non-nested models; all handled in the same framework.
   - 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
   $\boldsymbol{\mu}_i = E(\mathbf{Y_i})$

Review Linear Models II

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: $g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{X_i}^T \boldsymbol{\beta}$

3. Generalized Additive Model: $y \sim exponenetialfamily$
   $\boldsymbol{\mu} = E(\mathbf{y})$
   $g(\boldsymbol{\mu}) = \alpha_0 + f(x_1) + f(x_2) + ... + f(x_p)$
   i incorporate non-linear forms of the predictors
   ii Now the linear predictor incorporates smooth functions f(x) of at least some (possibly all) covariates

Review Linear Models III

iii maximize the quality of prediction of a dependent variable Y from various distributions, by estimating unspecific (non-parametric) functions of the predictor variables which are "connected" to the dependent variable via a link function.

4. Generalized Mixed Models:

   i Mixed models include both the usual fixed population effects and subject- or cluster-specific random effects in the linear predictor

   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 $\pi(\mathbf{x}|\theta)$.
- Some of the elements in the the random vector $\mathbf{x}$ are indirectly observed through the data $\mathbf{y}$
- 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 Gaussian models

- 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, $\pi(\mathbf{y}|\mathbf{x},\theta) = \prod_{i=1}^{n} \pi(y_i|\eta_i,\theta)$ ($\mathbf{y}$ 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 conditionalon 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 $\theta \sim \pi(\theta)$.

Review Bayesian

- Bayesian inference is based on computing the posterior distribution of a vector of the model parameters **x** conditioned on the vector of the observed data **y**
- Posterior distribution can be written as:

$$\pi(\mathbf{x}/\mathbf{y}) = \frac{\pi(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int \pi(\mathbf{y}|\mathbf{x})p(\mathbf{x})dx} \tag{1}$$

$$\pi(\mathbf{x}/\mathbf{y}) \propto \pi(\mathbf{y}/\mathbf{x})p(\mathbf{x}) \tag{2}$$

$\pi(\mathbf{y}/\mathbf{x})$: likelihood of the model
$p(\mathbf{x})$: prior distribution of the model parameters

- Usually $\pi(\mathbf{x}/\mathbf{y})$ 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

Review ...

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

1. Setting up of a full joint probability distribution for both observable, **y** and parameters, **x**;

$$\pi(\mathbf{y}|\mathbf{x}) = \pi(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

2. Update knowledge about the unknown parameters by conditioning this probability model on observed data, $\pi(\mathbf{x}|y)$
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, $p(y)$: The normalizing constant $p(y)$ in Bayes Theorem is a very important quantity defined by:

$$p(y) = \int \pi(\mathbf{y}, \mathbf{x})dx = \int \pi(\mathbf{y}|\mathbf{x})p(\mathbf{x})dx$$

  It represents the probability of observing the data that was observed before it was observed.

  ii Marginal effects of a subset of parameters in a multivariate model.
  Let $\mathbf{x} = (x_1, ...x_p)$ denote a p dimensional model. Suppose we are interested in $\pi(x_i|\mathbf{y})$, for some subset $x_i \epsilon \mathbf{x}$, then

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i, x_{-i}|\mathbf{y})dx_{-i} = \int \pi(x_i|x_{-i}, \mathbf{y})\pi(x_{-i}|\mathbf{y})dx_{-i}$$

  where $x_{-i} = x|x_i$ denotes the vector $\mathbf{x}$ with x_i removed.

  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:

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

Computational methods... I

1. Marckov Chain Monte Carlo Simulation (MCMC): the most widely used method to estimate posterior distribution
   - Can effectively be applied to any model
   - MCMC can provide (nearly) exact inference, given perfect convergence and MCMC error goes to 0.
   - 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,No U-turn sampling
     - Alternative methods of inference (eg Approximate Bayesian Computation (ABC), INLA)
2. Approximate Bayesian Computation (ABC)

Computational methods... II

3. Integrated Nested Laplace Approximation (INLA): computer performance of Bayesian inference for latent Gaussian models (LGM).

Examples of latent Gaussian models:

- Most of (generalised) linear models
- Smoothing spline models
- state space models,
- Semi-parametric regression,
- spatial and spatio-temporal models

....etc

Introduction to INLA

- Integrated nested Laplace approximation (INLA) is a computational approach to statistical inference for Latent Gaussian Markov Random field (GMRF), introduced by Rue and Martino (2007).
- It was proposed as an alternative to the usually time consuming MCMC methods.
- 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 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

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

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

- In general, we can partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ are the hyper parameter and $\boldsymbol{\theta}_2$ are the nuisance parameter and re-express the model as follows:

$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$	hyper prior
$\mathbf{x}|\boldsymbol{\theta} \sim p(\mathbf{x}|\boldsymbol{\theta}) = \mathrm{Normal}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_1))$	GMRF prior
$\mathbf{y}|\mathbf{x},\boldsymbol{\theta} \sim \prod_i \pi(y_i|x_i,\boldsymbol{\theta}_2)$	data model

LGMRF model as a general framework

- $\mathbf{y} = \{y_i\}_{i=1}^{n}$ is assumed to belong to an exponential family
- A very general way of formulation 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

Example: Dynamic models/ State space models

Dynamic models/ State space models

$$\begin{aligned} y_t &= F_t' x_t + v_t && (6) \\ x_t &= G_t' x_{t-1} + w_t && (7) \\ & && (8) \end{aligned}$$

where

$$\begin{aligned} v_t &\sim N(0, V_t) && (9) \\ w_t &\sim N(0, W_t) && (10) \\ & && (11) \end{aligned}$$

y_t is a time sequence of scaler 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 states 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 $\forall\ i$.

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j} \tag{12}$$

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \tag{13}$$

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

$$\tilde{\pi}(\theta_j|\mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-j} \tag{14}$$

$$\tilde{\pi}(x_i|\mathbf{y}) = \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \tag{15}$$

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

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
        - fast but computationally demanding
    - Simplified Laplace approximation
        - default in R-INLA
        - trade-off between speed and accuracy
3. Numerical integration detail

$$\tilde{\pi}_i(x_i|\mathbf{y}) = \sum_k \tilde{\pi}_i(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}_i(\boldsymbol{\theta}_k|\mathbf{y})\triangle_k \tag{16}$$

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

- Grid strategy
- Central composite design (CCD)

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

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

$$\begin{aligned}
\pi(\boldsymbol{\theta}|\mathbf{y}) &= \frac{\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (17)\\
&= \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\pi(\mathbf{y})}\frac{1}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (18)\\
&= \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{y})}\frac{1}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (19)\\
&\propto \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{x}|\boldsymbol{\theta}\mathbf{y})} && (20)\\
&\approx \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}\Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \doteqdot \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) && (21)
\end{aligned}$$

where

- $\mathbf{x} = \mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\mathbf{x}$, for a given $\boldsymbol{\theta}$

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

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 $\mathbf{z}$

$$\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z} \tag{22}$$

     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 H points $\{\theta_k^*\}$ associate with mass and area weights $\{\triangle_k\}$

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

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

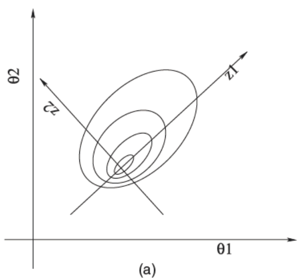

Fig. 1. Illustration of the exploration of the posterior marginal for θ: in (a) the mode is located and the Hessian and the co-ordinate system for **z** are computed; in (b) each co-ordinate direction is explored (•) until the log-density drops below a certain limit; finally the new points (•) are explored

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

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

The marginals for components x_i of the latent field

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \tag{23}$$

- 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

$$\pi(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}|Q|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}-\mu)'Q(\mathbf{x}-\mu)} \tag{24}$$

- based on the Cholesky decomposition of the precision matrix $Q(.) = LL'$
- fast but not very good due to the lack of skewness (Rue and Martino, 2007)

back

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

Laplace Approximation

$$
\begin{aligned}
\pi_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) &= \frac{\pi(\{x_j, \mathbf{x}_{-j}\}|\boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x}_{-j}|x_j, \boldsymbol{\theta}, \mathbf{y})} && (25)\\
&= \frac{\pi(\{x_j, \mathbf{x}_{-j}\}, \boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \frac{1}{\pi(\mathbf{x}_{-j}|x_j, \boldsymbol{\theta}, \mathbf{y})} && (26)\\
&\propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}_{-j}|x_j, \boldsymbol{\theta}, \mathbf{y})} \propto \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x})}{\pi(\mathbf{x}_{-j}, \mathbf{x}_j, \boldsymbol{\theta}, \mathbf{y})} && (27)\\
&\propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-j}=\mathbf{x}^*_{-j}(x_i,\boldsymbol{\theta})} && (28)\\
&\doteqdot \tilde{\pi}_{LA}(x_j|\boldsymbol{\theta}, \mathbf{y}) && (29)
\end{aligned}
$$

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

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

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

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

$$log\tilde{\pi}_{SLA}(x_i^s|\boldsymbol{\theta}, \mathbf{y}) = const. - \frac{1}{2}(x_i^s)^2 + \gamma_i^1(\boldsymbol{\theta})x_i^s + \frac{1}{6}(x_i^s)^3\gamma_i^3(\boldsymbol{\theta}) + \ldots \tag{30}$$

  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

$$\tilde{\pi}_i(x_i|\mathbf{y}) = \sum_k \tilde{\pi}_i(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}_i(\boldsymbol{\theta}_k|\mathbf{y})\triangle_k \quad (31)$$

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

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

1. Grid strategy
   - Find mode / Compute Hessian / Grid search
   - Note: accurate but may be time consuming if m=dim($\boldsymbol{\theta}$) is large
2. 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 by R-INLA

back

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 H 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

$$\tilde{\pi}_i(x_i|\mathbf{y}) = \sum_k \tilde{\pi}_i(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}_i(\boldsymbol{\theta}_k|\mathbf{y})\triangle_k \tag{32}$$

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

Integrated Nested Laplace Approximation (INLA)

Integrated Nested Laplace Approximation

- Laplace is the fundamental tool to estimate the unknown distributions
- Laplace approximation are nested within one another
- Numerical integration is used to obtain marginal posterior distributions

- Proposed formulated framework to fit common types of DM using INLA through simulated data sets
- Firstly, consider simple univariate dynamic linear model
- Next, state-space models that can be directly fitted using model options in the INLA library
- Models that could not be fitted using INLA's standard tools

A Toy Example (Ruiz et al (2010))

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

$$y_t = x_t + v_t, \quad v_t \sim N(0, V), \quad t = 1, ..., n \tag{33}$$

$$x_t = x_{t-1} + w_t, \quad w_t \sim N(0, W), \quad t = 2, ..., n \tag{34}$$

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 $\xi = \{x_1, ..., x_n\}$

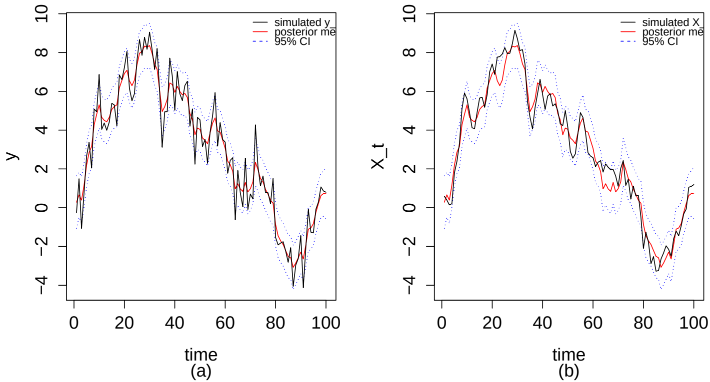

Figure : Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and states (b) in the toy example

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

$$(y_t|\mu_t) \sim Poisson(\mu_t) \tag{35}$$

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

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

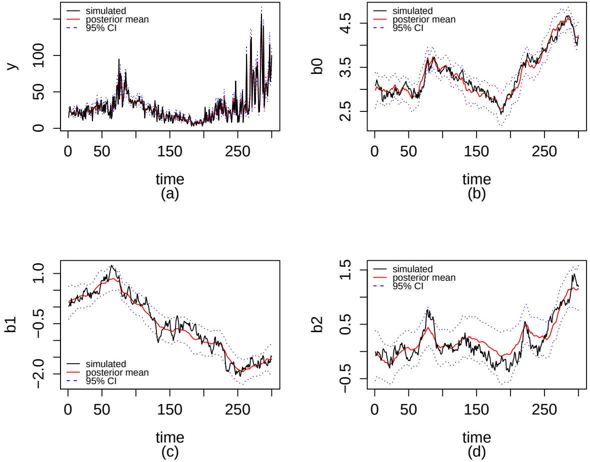

Figure : Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and regression coefficients, β_0, β_1 and β_2 (b-d) in the generalized dynamic regression example

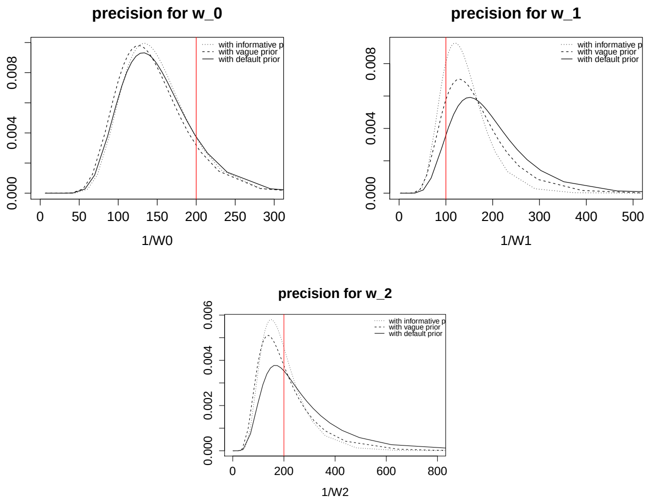

Figure : Posterior densities for the hyperparameters in the generalized dynamic regression example. Red lines indicate true simulated values

REFERNCES

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

Thank you!