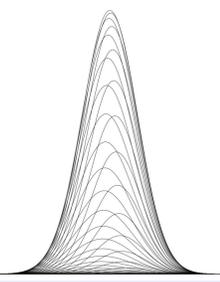


Validating the Joint Posterior Approximations provided by INLA



Cristian Chiuchio¹, Håvard Rue¹
¹King Abdullah University of Science and Technology
 Bayesian Computational Statistics & Modeling
 cristian.chiuchio@kaust.edu.sa



General Overview

- Bayesian Inference can be applied to any model through **MCMC** but many solutions still pay a lot in terms of speed
- **INLA** is able to provide faster approximate inference for a wide class of models known as Latent Gaussian Models (**LGMs**)
- It is even possible to use the internal approximations within **INLA** to provide independent corrected samples from the approximated joint posterior
- This approximation turns out to be enough accurate for any dimension and distribution

Joint Approximation

In the context of **LGMs**, the true joint posterior for a set of observations $\mathbf{y} = \{y_i, i \in I\}$ and a joint set $(\mathbf{x}, \boldsymbol{\theta})$ of latent field and hyperparameters respectively is given by

$$\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \prod_{i \in I} \pi(y_i | x_i, \boldsymbol{\theta})$$

The following posterior marginals are computed internally in **INLA**[4] using nested Laplace Approximations

$$\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \propto \frac{\pi(\mathbf{x}^*, \boldsymbol{\theta} | \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}^* | \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}^* = \boldsymbol{\mu}(\boldsymbol{\theta})}$$

$$\tilde{\pi}(\theta_j | \mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\theta_{-j}, \quad j = 1, \dots, p$$

$$\tilde{\pi}(x_i | \mathbf{y}) = \int \tilde{\pi}(x_i | \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad i = 1, \dots, n$$

The joint $\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$ can be approximated through numerical integration

$$\tilde{\pi}(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \approx \sum_k \tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \mathbb{1}_{[\boldsymbol{\theta} = \boldsymbol{\theta}_k]} \Delta_k$$

where k describes the number of points in the grid exploration for achieving $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$.

Methodology for the Joint

How we get the approximation $\tilde{\pi}(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$?

1. the hyperparameters are sampled from the integration points $\{\theta_i, i = 1, \dots, k\}$
2. for each sampled configuration point a sample is drawn from $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$

We can also extend the joint approximation $\tilde{\pi}(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$ exploiting Skew-Normal marginals $\tilde{\pi}_{SN}(x_i | \boldsymbol{\theta}, \mathbf{y})$ for the approximation $\tilde{\pi}(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$ through a Gaussian copula [3]. This leads to new transformed values

$$\tilde{x}_i = \sigma_i(\boldsymbol{\theta}) \tilde{F}_i^{-1} \left[\Phi \left(\frac{x_i - \mu_i(\boldsymbol{\theta})}{\sigma_i(\boldsymbol{\theta})} \right) \right] + \tilde{\mu}_i(\boldsymbol{\theta})$$

where $x_i \sim N(\mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta}))$. This describes a **Skewness Correction** within the internal function `inla.posterior.sample` which extends the default mean correction applying the Standard Skew-Normal cumulative function.

JAGS/INLA Comparison

Through the tail summary $E[\mathbf{x} | \mathbf{x} \notin (-\gamma_l, \gamma_r)]$ applied on the following simple model with an *iid* random effect \mathbf{u} based on $m = 10$ randomized groups and $n = 50$ observations

$$\begin{aligned} \mathbf{y} | \boldsymbol{\eta} &\sim \text{Poi}(\boldsymbol{\mu}) \\ \log(\boldsymbol{\mu}) &= \boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{u} \\ \mathbf{u} &\sim N_m(\mathbf{0}, \tau^{-1} \mathbf{I}) \\ \boldsymbol{\alpha} &\sim N(0, 1000) \\ \tau &\sim \text{Ga}(0.1, 0.1) \end{aligned}$$

we show the improvement coming from the new skewness correction

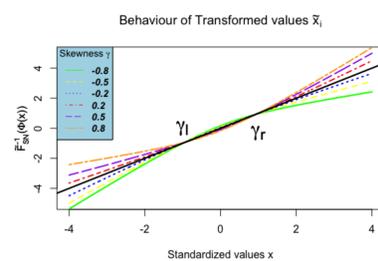


Figure 1: Standardized Gaussian values against Skewness corrected values \tilde{x}_i

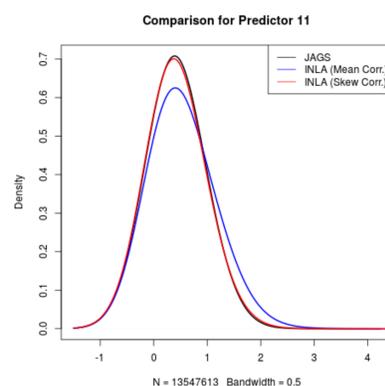


Figure 2: Marginal posterior plot for Predictor 11 after long run simulation using JAGS and INLA

Validate the Joint with SBC

Algorithm 1 SBC[1] applied on a subset \mathbf{x}_S

Input: The joint posterior density $\pi(\mathbf{x}, \boldsymbol{\theta}_k | \mathbf{y})$ with k fixed; number of replicated data sets M and dimension S for $\mathbf{x}_S = \{x_i, i \in S\}$;

Output: p-value for testing if the joint approximation is biased in the i^{th} margin of \mathbf{x}_S .

- 1: **for** $j = 1 : M$ **do**
- 2: Solve $\mathbf{Q}\boldsymbol{\Sigma} = \mathbf{I}$;
- 3: Extract $\boldsymbol{\Sigma}_{SS}$ and solve $\boldsymbol{\Sigma}_{SS}\mathbf{Q}_{SS} = \mathbf{I}_{SS}$;
- 4: Draw $\mathbf{x}_S^{(0)} \sim N(\boldsymbol{\mu}_S, \mathbf{Q}_{SS}^{-1})$;
- 5: Generate a data set \mathbf{y}_j from $\pi(\mathbf{y}_j | \mathbf{x}_S, \boldsymbol{\theta}_k)$;
- 6: Run the sampler on dataset \mathbf{y}_j and get an approximation for the joint subset $q_{S(j)}$;
- 7: **for** $i = 1 : S$ **do**
- 8: Label $x_{i(j)}^{(0)}$ as the i^{th} marginal component of $\mathbf{x}_{S(j)}^{(0)}$;
- 9: Label x_i^* as the i^{th} marginal component of \mathbf{x}_S^* ;
- 10: Get $p_{ij} = Pr(x_{i(j)}^{(0)} < x_i^* | \mathbf{x}_S^* \sim q_{S(j)})$
- 11: **end for**
- 12: **end for**
- 13: **for** $i = 1 : S$ **do**
- 14: Test the symmetry of the distribution of $\{p_{ij}\}_{j=1}^M$
- 15: If rejected the i^{th} marginal is biased
- 16: **end for**

Testing Uniformity

We can test if the univariate marginal posteriors for \mathbf{x}_S are biased through visual diagnostics on the p-values p_{ij} computed with SBC.

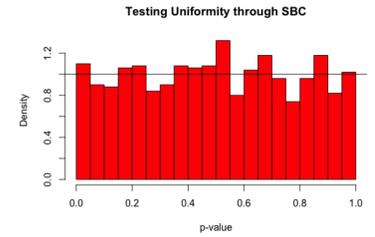


Figure 3: The histogram does not show any relevant deviation and the Kolmogorov-Smirnov test is higher than 0.9 confirming the bias absence

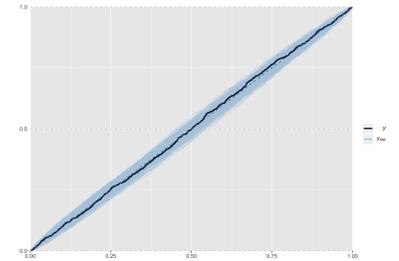


Figure 4: ECDF of the p-values does not show any deviation expected from a uniform distribution

HIV Prevalence in Kenya

As reported in [2], we can use **INLA** to fit a spatial model with GRF $S(\cdot)$

$$\begin{aligned} y_{ij} &\sim \text{Bin}(N_{ij}, P_{ij}) \quad i = 1, \dots, n, \quad j = 1, \dots, m_i \\ P_{ij} &= \text{logit}^{-1} \{ \beta_0 + S(x_i) + \epsilon_{ij} \} \end{aligned}$$

and explore the average HIV prevalence over sub regions A_k in Kenya using posterior summaries $T_k = \int_{A_k} \text{logit}^{-1} \{ \beta_0 + S(x) \} dx$

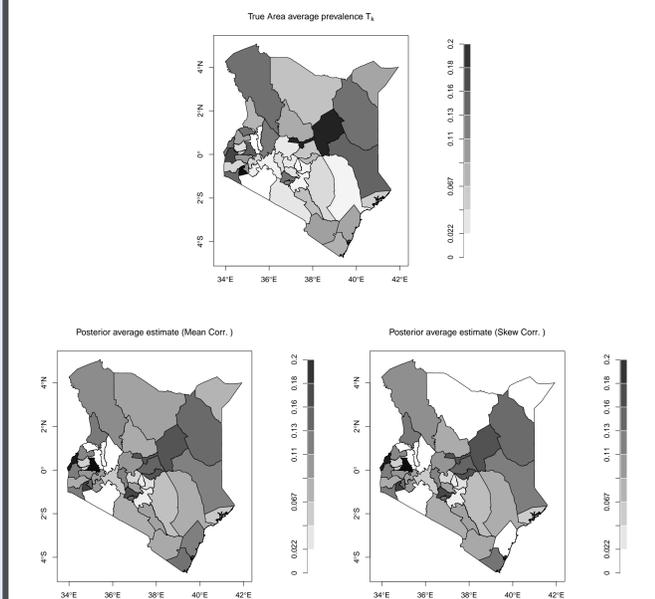


Figure 5: SPDE results for area average prevalence $\{T_k, k = 1, \dots, 47\}$ through INLA

References

- [1] Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman. Yes, but did it work?: Evaluating Variational Inference. *Proceedings of the 35th International Conference on Machine Learning*, 2018.
- [2] Jon Wakefield, Daniel Simpson and Jessica Godwin. 'Spatial Modeling, with Application to Complex Survey Data'. Discussion of 'Model-based Geostatistics for Prevalence Mapping in Low-Resource Settings', by Diggle and Giorgi. *arXiv:1608.03769*, 2016.
- [3] E. Ferkingstad and H. Rue. Improving the INLA approach for approximate Bayesian inference for latent gaussian models. *Electronic Journal of Statistics*, 2015.
- [4] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of Royal Statistical Society*, 2009.