Mini-project: Integrated nested Laplace Approximation with Automatic Differentiation

Adam Howes

Imperial College London

April 2020

Motivation

- R-INLA (Martins et al. 2013) only works for the particular models which have been implemented
- Alternative implementation based on automatic differentation (AD) would allow INLA to be used for a broader class of models
- For example, the HIV inference group at Imperial is working on a model just outside R-INLA's capacity

What is INLA, why do we want to use it, and why can't we currently?

Three-stage model

- Want to do Bayesian inference in spatiotemporal statistics
- Three-stage model covers most of the models used

$$\begin{aligned} &\text{(Observations)} && \mathbf{y} \sim p(\mathbf{y} \,|\, \mathbf{x}), \\ &\text{(Latent field)} && \mathbf{x} \sim p(\mathbf{x} \,|\, \boldsymbol{\theta}), \\ &\text{(Hyperparameters)} && \boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \end{aligned}$$

  where $\mathbf{y} = (y_1, \ldots, y_n)$, $\mathbf{x} = (x_1, \ldots, x_n)$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$
- Interested in learning both $(\boldsymbol{\theta}, \mathbf{x})$ from data $\mathbf{y}$

Have you tried MCMC?

- Markov chain Monte Carlo is slow for high dimensional correlated parameter spaces
- We have both of these problems:
  - If $\mathbf{x}$ represents spatiotemporal location then $\dim(\mathbf{x}) = n$ will be very large
  - Tobler's first law of geography "everything is related to everything else, but near things are more related than distant things" $\implies$ $\mathbf{x}$ has lots of correlation structure

Approximate Bayesian inference

- In applied statistics (at least in health and social science) we fit misspecified models to biased and incomplete data
- Is inferential exactness (as $n_{\text{sim}} \to \infty$ for chain of length n_{sim}) the scientific bottleneck?
- If not $\implies$ shouldn't be afraid of approximate methods
  - Approximate Bayesian computation (ABC)
  - Variational Bayes
  - Integrated nested Laplace approximation (INLA)

Integrated nested Laplace approximation (I)

- See Rue, Martino, and Chopin (2009) or Blangiardo and Cameletti (2015)
- Approximate Bayesian inference for latent Gaussian models (LGMs), which are three-stage models with middle layer

$$\text{(Latent field)} \qquad p(\mathbf{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}(\boldsymbol{\theta})^{-1}).$$

- Takes advantage of sparsity properties of $\boldsymbol{Q}(\boldsymbol{\theta})$, i.e. if $\mathbf{x}$ is a Gaussian Markov random field (GMRF)
- Gives approximate posterior marginals $\{\tilde{p}(x_i \mid \mathbf{y})\}_{i=1}^{n}$ and $\{\tilde{p}(\theta_j \mid \mathbf{y})\}_{j=1}^{m}$

Integrated nested Laplace approximation (II)

1) First Laplace approximate hyperparameter posterior

$$\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \left. \frac{p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\tilde{p}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}=\boldsymbol{\mu}^\star(\boldsymbol{\theta})} \tag{1}$$

   which can be marginalised to get $\tilde{p}(\theta_j \mid \mathbf{y})$
2) Choose integration points and weights $\{\boldsymbol{\theta}^{(k)}, \Delta^{(k)}\}$ to integrate w.r.t. (1)
3) Choose approximation for $\tilde{p}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$ (simplest version: Gaussian)
4) Finally use quadrature to get

$$\tilde{p}(x_i \mid \mathbf{y}) = \sum_{k=1}^{K} \tilde{p}(x_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}) \times \tilde{p}(\boldsymbol{\theta}^{(k)} \mid \mathbf{y}) \times \Delta^{(k)} \tag{2}$$

Naomi, evidence synthesis for HIV

- Combine HIV prevalence ρ_i and ART coverage α_i models together
- Model is close to, but not, a LGM
- Small non-linearities e.g. multiplying two latent Gaussian fields

$$A_i \sim \text{Bin}(\rho_i \alpha_i, N_i),$$

where A_i be the number observed on ART and N_i the population
- Need something more flexible than R-INLA

Figure 1: Supermodel

What do we do currently instead?

Template Model Builder (I)

- Currently we use `TMB` (Kristensen et al. 2016)
- R package which implements the Laplace approximation for latent variable models using AD (via `CppAD`)
  - For more about AD see e.g. Griewank and Walther (2008)
- Write an objective function $f(\mathbf{x}, \boldsymbol{\theta})$ in C++ ("user template")
  - We select $f(\mathbf{x}, \boldsymbol{\theta}) = -\log p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$

Template Model Builder (II)

```
#include <TMB.hpp>

template <class Type>
Type objective_function<Type>::operator()() {
  // Define data e.g.
  DATA_VECTOR(y);
  // Define parameters e.g.
  PARAMETER(mu);
  // Calculate negative log-likelihood e.g.
  nll = Type(0.0);
  nll -= dnorm(y, mu, 1, true).sum()
  return(nll);
}
```

Template Model Builder (III)

- Performs the Laplace approximation $L_f(\boldsymbol{\theta}) \approx L_f^\star(\boldsymbol{\theta})$ (step 1 of INLA) – use R to optimise this with respect to $\boldsymbol{\theta}$ to give $\hat{\boldsymbol{\theta}}$
- MAP estimate of $\mathbf{x}$ conditional on $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ (REML? Empirical Bayes?)
- Standard errors calculated using the δ-method (a Gaussian assumption)

What do you want to do in the future?

Aims

- Compare accuracy of `TMB` to `R-INLA`
- Implement the INLA method using AD via `TMB`
- Apply new method to models with different degrees of non-linearity
  - Small degree: Naomi.
  - Larger degree: ODE models e.g. SIR or other compartmental models

Thanks! Questions / comments / corrections?

References I

Blangiardo, Marta, and Michela Cameletti. 2015. *Spatial and Spatio-Temporal Bayesian Models with r-INLA*. John Wiley & Sons.

Griewank, Andreas, and Andrea Walther. 2008. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Vol. 105. Siam.

Kristensen, Kasper, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70 (5): 1–21. https://doi.org/10.18637/jss.v070.i05.

Martins, Thiago G, Daniel Simpson, Finn Lindgren, and Håvard Rue. 2013. "Bayesian Computing with INLA: New Features." *Computational Statistics & Data Analysis* 67: 68–83.

References II

Rue, Håvard, Sara Martino, and Nicolas Chopin. 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): 319–92.