Integrated nested Laplace approximations for extended latent Gaussian models with application to the Naomi HIV model

Waterloo SAS Student Seminar Series

Adam Howes

Imperial College London

November 2022

Motivation

- Surveillance of the HIV epidemic in sub-Saharan Africa
- Want to estimate indicators used for monitoring and response, including:
  - Prevalence ρ: the proportion of people who are HIV positive
  - Incidence λ: the proportion of people newly infected
  - Treatment coverage α: the proportion of PLHIV on treatment
- We would like to provide them at a local, district-level

This is a challenging task! Data is noisy, sparse and biased. $\implies$ compelling case for thoughtful Bayesian modelling.

A simple small-area model for prevalence

- Consider small-areas $i = 1, \ldots, n$ like the districts of a country
- Simple random sample household-survey of size m_i^{HS} in each area
- The number of people testing positive for HIV is y_i^{HS}
- You could calculate direct estimates of prevalence by $y_i^{\text{HS}} / m_i^{\text{HS}}$ but because the survey is powered at a national-level, the sample sizes are small and these estimates would be noisy

A simple small-area model for prevalence

- We can use a binomial logistic regression of the form:

$$y_i^{\mathrm{HS}} \sim \mathrm{Bin}(m_i^{\mathrm{HS}}, \rho_i^{\mathrm{HS}}),$$
$$\mathrm{logit}(\rho_i^{\mathrm{HS}}) \sim g(\vartheta^{\mathrm{HS}}), \quad i = 1, \dots, n,$$

- We usually set up g as a Gaussian spatial smoother
- This allows for pooling of information between districts

Latent Gaussian models

- Three-stage Bayesian hierarchical model

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

  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 $\mathbf{y}$
- If the middle layer is Gaussian, then it's a latent Gaussian model

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

- Covers most of the models commonly used in spatiotemporal statistics
- Latent field is typically indexed by spatiotemporal location, such that $n > m$

Limitations of household surveys

- A typical household survey takes around $UNKNOWN to run
- This means that they don't happen very often
- For this reason, the information might be quite out of date, and difficult to base policy on

Adding ANC surveillance

- Pregnant women attending antenatal care clinics are routinely tested for HIV, to avoid mother-to-child transmission
- This data source is more real-time than household surveys, but it's also more biased, because attendees are unlikely to be as representative of the population
- But perhaps this bias is consistent, in which case we can still make use of the ANC data to supplement our model!

Adding ANC surveillance

- Suppose of m_i^{ANC} women attending ANC, y_i^{ANC} are HIV positive, then we can use another binomial logistic regression:

$$
\begin{aligned}
y_i^{\text{ANC}} &\sim \text{Bin}(m_i^{\text{ANC}}, \rho_i^{\text{ANC}}), \\
\text{logit}(\rho_i^{\text{ANC}}) &= \text{logit}(\rho_i^{\text{HS}}) + b_i, \\
b_i &\sim \mathcal{N}(\beta_b, \sigma_b^2),
\end{aligned}
$$

- This is similar to using ρ_i^{ANC} as a covariate in the model for household survey prevalence, but this way takes into account sampling variation

Adding ART coverage

- We're also interested in what proportion α_i of people living with HIV (PLHIV) are receiving treatment
- Suppose we record A_i attendees from a known population of N_i in each district
- We can use another logistic regression model

$$
\begin{aligned}
A_i &\sim \text{Bin}(N_i, \rho_i^{\text{HS}} \alpha_i), \\
\text{logit}(\alpha_i) &\sim \mathcal{N}(\beta_\alpha, \sigma_\alpha^2).
\end{aligned}
$$

Naomi evidence synthesis model

- Combining these three modules is the basis of the Naomi evidence synthesis model
- Used by countries to produce HIV estimates in a yearly process supported by UNAIDS
- Can't run long MCMC in this setting, so we require fast, accurate, approximations
- It's a complicated model, and requires something more flexible than R-INLA
- Currently using a package called Template Model Builder TMB

Figure 1: A supermodel

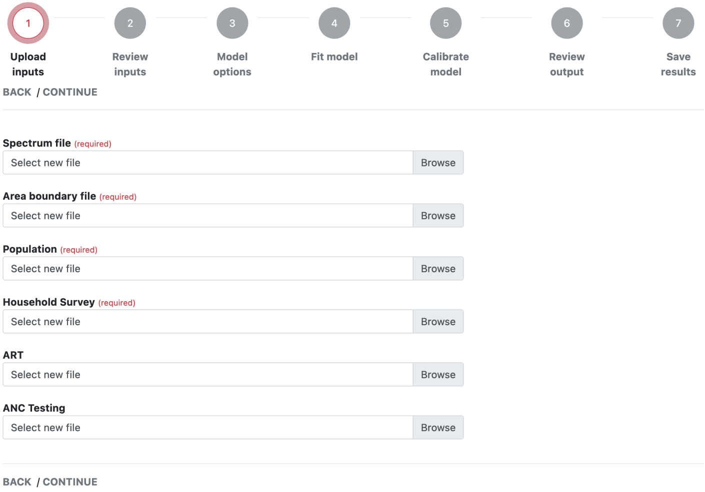

Figure 2: Example of the user interface from https://naomi.unaids.org/

Template Model Builder

- `TMB` (Kristensen et al. 2015) is an R package which implements the Laplace approximation for latent variable models
- To get started, write an objective function $f(\mathbf{x}, \boldsymbol{\theta})$ in `TMB` C++ syntax
- As pseudo-Bayesians, we choose the log-posterior

$$f(\mathbf{x}, \boldsymbol{\theta}) = -\log p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

Template Model Builder

For example, for the model

$$\mathbf{y} \sim \mathcal{N}(\mu, 1)$$

with $p(\mu) \propto 1$ then the $\texttt{TMB}$ user template looks like...

```
#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

- We can use TMB to obtain the Laplace approximation

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

- Integrate out a Gaussian approximation $\tilde{p}_{\mathsf{G}}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$ to the latent field
- TMB uses automatic differentiation (Griewank and Walther 2008) via CppAD

Integrated Nested Laplace Approximation

- Integrated nested Laplace approximation (INLA) (Rue, Martino, and Chopin 2009; Blangiardo and Cameletti 2015) is an approach to approximate inference which builds on the Laplace approximation
- Goal is to approximate posterior marginals $\{\tilde{p}(x_i \mid \mathbf{y})\}_{i=1}^n$ and $\{\tilde{p}(\theta_j \mid \mathbf{y})\}_{j=1}^m$

$$p(x_i \mid \mathbf{y}) = \int p(x_i, \boldsymbol{\theta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\theta} = \int p(x_i \mid \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\theta}, \quad i = 1, \ldots, n, \tag{1}$$

$$p(\theta_j \mid \mathbf{y}) = \int p(\boldsymbol{\theta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\theta}_{-j} \quad j = 1, \ldots, m. \tag{2}$$

- To do so, we require the approximations $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$ and $\tilde{p}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$
- There are four steps as to how the method works (bare with me!)

Step 1)

1) First Laplace approximate hyperparameter posterior

$$\tilde{p}_{\mathsf{LA}}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \left. \frac{p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})}{\tilde{p}_{\mathsf{G}}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x} = \boldsymbol{\mu}^\star(\boldsymbol{\theta})} \tag{3}$$

which can be marginalised to get $\tilde{p}(\theta_j \mid \mathbf{y})$

- Notice that this is the same object we had been working with in `TMB`
- We use this approximation nested within integrals – hence the name INLA

Step 2)

2) In both Equations (1) and (2) we want to integrate w.r.t. $\boldsymbol{\theta}$, so choose integration nodes and weights $\{\boldsymbol{\theta}(\mathbf{z}), \omega(\mathbf{z})\}_{\mathbf{z} \in \mathcal{Z}}$

- For low m `R-INLA` uses a grid-strategy
- For larger m this becomes too expensive and `R-INLA` uses a CCD design
- We plan to use adaptive Gaussian Hermite quadrature (AGHQ), which has recently been shown to have theoretical guarantees (Bilodeau, Stringer, and Tang 2021)

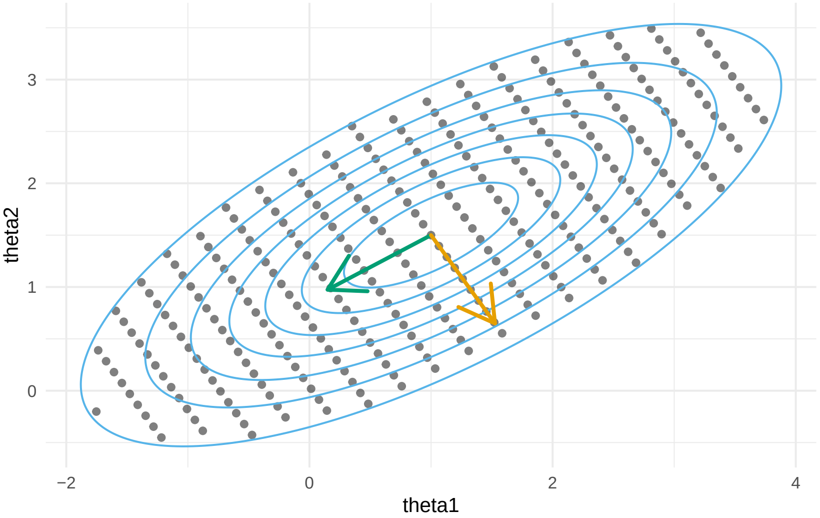

Figure 3: An illustration of the R-INLA grid method for selecting integration nodes using a toy bivariate Gaussian distribution for $\boldsymbol{\theta}$. Start at the mode and work outwards along the eigenvectors until the density drops sufficiently low.

Adaptive Gaussian Hermite Quadrature

- Gauss-Hermite quadrature is a way of picking nodes and weights, and is based on the theory of polynomial interpolation
- The adaptive part means that it uses the location (mode) and curvature (Hessian) of the target (posterior) to automatically choose the node locations
  - Does not require manual tuning!
- Works particularly well when the integrand is pretty Gaussian
- Use k quadrature nodes per dimension, for example if $k = 3$ then 3^m total nodes
- Implemented in the `aghq` R package. See vignette Stringer (2021)

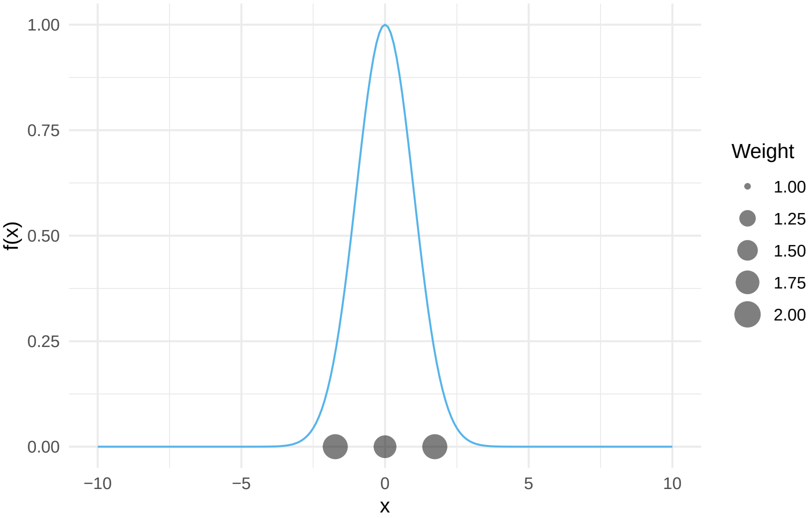

Figure 4: One dimensional example of AGHQ.

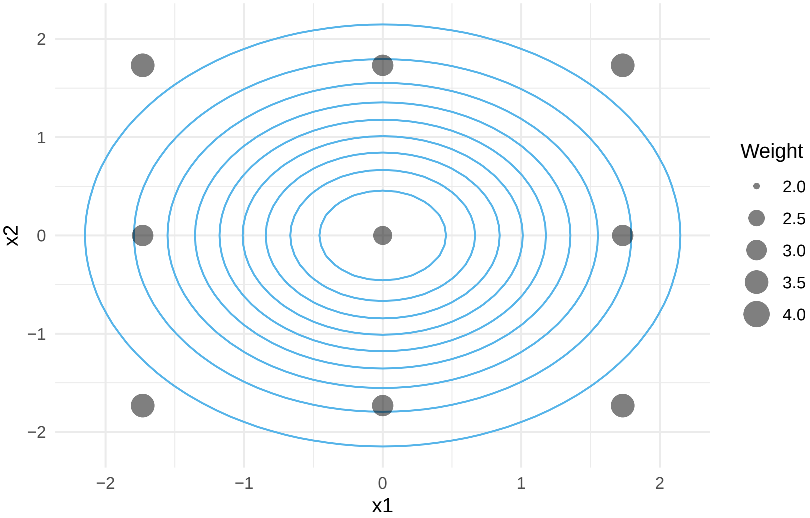

Figure 5: Two dimensional example of AGHQ.

Step 3)

3) Choose approximation for $\tilde{p}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$

- Simplest version (Rue and Martino 2007) is to marginalise $\tilde{p}_{\mathrm{G}}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$

$$\tilde{p}_{\mathrm{G}}(x_i \mid \boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(x_i \mid \mu_i^\star(\boldsymbol{\theta}), 1/q_i^\star(\boldsymbol{\theta})) \tag{4}$$

- In R-INLA, the above is referred to as `method = "gaussian"`
- There are two better, more complex approximations, confusingly called `"simplified.laplace"` and `"laplace"`
- Uses sparsity properties of $\boldsymbol{Q}(\boldsymbol{\theta})$, i.e. if $\mathbf{x}$ is a Gaussian Markov random field (GMRF)

Step 4)

4) Finally, use quadrature to combine

- our approximation $\tilde{p}_{\mathrm{LA}}(\boldsymbol{\theta} \mid \mathbf{y})$ from step 1),
- some choice of integration nodes and weights $\{\boldsymbol{\theta}(\mathbf{z}), \omega(\mathbf{z})\}$ from step 2),
- some choice of approximation $\tilde{p}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$ from step 3) to give

$$\tilde{p}(x_i \mid \mathbf{y}) = \sum_{\mathbf{z} \in \mathcal{Z}} \tilde{p}(x_i \mid \boldsymbol{\theta}(\mathbf{z}), \mathbf{y}) \times \tilde{p}_{\mathrm{LA}}(\boldsymbol{\theta}(\mathbf{z}) \mid \mathbf{y}) \times \omega(\mathbf{z}) \tag{5}$$

Experiments

- We wrote a simplified version of the Naomi model up in `TMB`
- This allowed us to test three inference methods **all using precisely the same model and C++ code**

1. A direct Gaussian approximation via `TMB`
2. Adaptive Gaussian Hermite quadrature via `aghq`
3. No-U-Turn Sampling (NUTS – a type of Hamiltonian Monte Carlo) via `tmbstan`

- Note: using different software it is usually very difficult to ensure the model is precisely the same, so we're very fortunate here

Comparison approach

- You could look at the summaries like the mean and standard deviation of each of the posterior marginals
  - Any approximation method should be pretty good at getting the mean right
  - Gaussian approximations should be good at getting the second moment right
- It's probably better to compare the whole posterior distributions
- One way to do this is via Kolmogorov-Smirnov statistics, which give the maximum difference between two empirical CDFs

References I

Bilodeau, Blair, Alex Stringer, and Yanbo Tang. 2021. "Stochastic Convergence Rates and Applications of Adaptive Quadrature in Bayesian Inference." https://arxiv.org/abs/2102.06801.

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 Brad Bell. 2015. "TMB: automatic differentiation and Laplace approximation." *arXiv Preprint arXiv:1509.00660*.

Rue, Håvard, and Sara Martino. 2007. "Approximate Bayesian inference for hierarchical Gaussian Markov random field models." *Journal of Statistical Planning and Inference* 137 (10): 3177–92.

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.

Stringer, Alex. 2021. "Implementing Approximate Bayesian Inference Using Adaptive Quadrature: The Aghq Package." https://arxiv.org/abs/2101.04468.