Eventually an introduction to the `aghq` R package

Statistics and Epidemiology Group Meeting

Adam Howes

Imperial College London

July 2021

Motivation I

- People on ART A_i can be used as supplementary data for small-area estimation of HIV prevalence ρ_i

$$
\begin{aligned}
A_i &\sim \text{Bin}(N_i, \rho_i \alpha_i), \\
y_i &\sim \text{Bin}(m_i, \rho_i), \\
\text{logit}(\alpha_i) &\sim f(\vartheta_\alpha), \\
\text{logit}(\rho_i) &\sim g(\vartheta_\rho), \quad i = 1, \ldots, n,
\end{aligned}
$$

- If f and g are Gaussian then model is almost, but not quite, a latent Gaussian model by the definition of Rue, Martino, and Chopin (2009)
  - This is due to small non-linearities (multiplying two latent Gaussian fields)
  - Each observation depends on more than one element of the latent field

Motivation II

- Previous slide is a simplified version of the Naomi evidence synthesis model
- Countries to fit the model using their own data ("in production"?)
    - Can't run long MCMC on the cluster for weeks, as might be the case if this was one paper
- Can't use `R-INLA`, require something more flexible
- Currently using Template Model Builder `TMB` (Kristensen et al. 2015)

Figure 1: A supermodel

Aside: common theme I

- Combining flawed (sparse, aggregated) gold standard (measuring the thing we want) data with other correlated (more available, high resolution) data (measuring not exactly what we want)
- Consistently resulting in models with multiple outcomes (evidence synthesis, multi-output)
- I think a lot of these are going to be not quite LGMs

Aside: common theme II

- Examples include
  - Naomi model: DHS data is "gold standard", supported by ANC data from pregnant women
  - Sexual risk behaviour model: estimates of FSW population at national level, supported by DHS data
    - The national-level FSW estimates might be more like "bronze standard"
    - DHS approximately asks "have you received money or gifts in exchange for sex in past 12 months"
  - Loa loa prevalence and eyeworm history prevalence model: measuring eyeworm history is a cheap proxy for Loa loa (Amoah, Diggle, and Giorgi 2020)

Recap on latent Gaussian models

- Three-stage model

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

  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 $\mathbf{y}$
- Covers most of the models used in spatiotemporal statistics

Recap on Integrated Nested Laplace Approximation I

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

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

- R-INLA implementation takes advantage of sparsity properties of $\boldsymbol{Q}(\boldsymbol{\theta})$, i.e. if $\mathbf{x}$ is a Gaussian Markov random field (GMRF)

Recap on Integrated Nested Laplace Approximation II

- Gives approximate posterior marginals $\{\tilde{p}(x_i \mid \mathbf{y})\}_{i=1}^{n}$ and $\{\tilde{p}(\theta_j \mid \mathbf{y})\}_{j=1}^{m}$
- To approximate posterior marginals below requires $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$ and $\tilde{p}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$

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

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

Recap on Integrated Nested Laplace Approximation III

1) First Laplace approximate hyperparameter posterior

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

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

2) In both (1) and (2) we want to integrate w.r.t. (3), so choose integration points and weights $\{\boldsymbol{\theta}^{(k)}, \Delta^{(k)}\}$

- For low m INLA uses a grid-strategy which I illustrate in the next slide
- For larger m this becomes too expensive and a CCD design is used

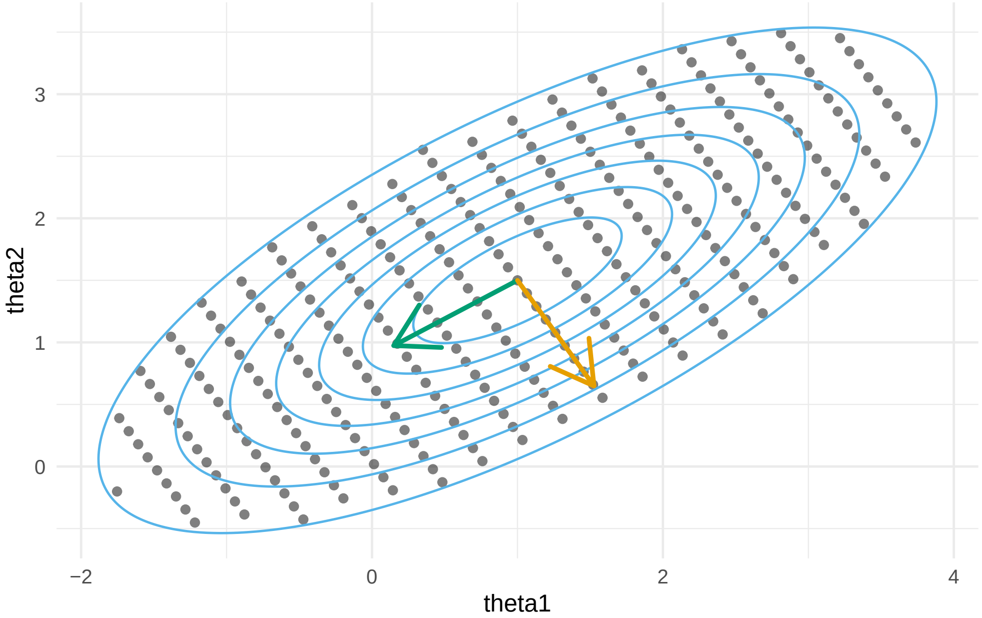

Figure 2: An illustration of the INLA grid method for selecting integration points using a toy

Recap on Integrated Nested Laplace Approximation IV

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

- Simplest version (Rue and Martino 2007) is to marginalise the $p_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$

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

- The above is referred to as "Gaussian" approximation, and confusingly there are two more complex ones called "simplified Laplace" and "Laplace"
- You can pick which one in `R-INLA` using the `method` option

4) Finally use quadrature to get

$$\tilde{p}(x_i \mid \mathbf{y}) = \sum_{k=1}^{K} \tilde{p}(x_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}) \times \tilde{p}(\boldsymbol{\theta}^{(k)} \mid \mathbf{y}) \times \Delta^{(k)} \tag{5}$$

Template Model Builder I

- 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)
  - Useful for getting the mode, Hessian
- 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 $L_f(\boldsymbol{\theta}) \approx L_f^{\star}(\boldsymbol{\theta})$ and use R to optimise this with respect to $\boldsymbol{\theta}$ to give $\hat{\boldsymbol{\theta}}$ (the central point in Figure 2)
  - This is done by specifying the `random` argument to be the parameters that you want to integrate out with a Laplace approximation (the latent field)
- MAP estimate of $\mathbf{x}$ conditional on $\hat{\boldsymbol{\theta}}$
- Standard errors calculated using the δ-method (a Gaussian assumption)

Adaptive Gaussian Hermite Quadrature

- Recent work by Alex Stringer and coauthors on AGHQ
  - `aghq` R package and vignette (Stringer 2021)
  - Theory paper (Bilodeau, Stringer, and Tang 2021)
- 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)
- Use k quadrature points
  - If k is odd then they include the mode
  - If $k = 1$ then it's a Laplace approximation
  - In the vignette $k = 3$ (for each dimension, so 3^m total) is chosen quite often

Epil example I

- Epilepsy example from Section 5.2. of Rue, Martino, and Chopin (2009) (previously from BUGS):
  - Patients $i = 1, \ldots, 59$ each either assigned treatment $\texttt{Trt}_i = 1$ or placebo $\texttt{Trt}_i = 0$ to help with seizures
  - Visits to clinics $j = 1, \ldots, 4$ times with y_{ij} the number of seizures of the ith person in the two weeks proceeding their jth visit to the clinic
  - Covariates age $\texttt{Age}_i$, baseline seizure counts $\texttt{Base}_i$ and an indicator for the final clinic visit $\texttt{V}_4$

Epil example II

This is what the model looks like (it's a Poisson GLMM):

$$
\begin{aligned}
y_{ij} &\sim \mathsf{Poisson}(\lambda_{ij}),\\
\lambda_{ij} &= e^{\eta_{ij}},\\
\eta_{ij} &= \beta_0 + \beta_{\mathtt{Base}} \log(\mathtt{Baseline}_j/4) + \beta_{\mathtt{Trt}} \mathtt{Trt}_i + \beta_{\mathtt{Trt}\times\mathtt{Base}} \mathtt{Trt}_i \times \log(\mathtt{Baseline}_j/4)\\
&\quad + \beta_{\mathtt{Age}} \log(\mathtt{Age}_i) + \beta_{\mathtt{V_4}} \mathtt{V_4}_j + \epsilon_i + \nu_{ij}, \quad i = 1:59, \quad j = 1:4,\\
\beta &\sim \mathcal{N}(0, 100^2), \quad \forall \beta,\\
\epsilon_i &\sim \mathcal{N}(0, 1/\tau_\epsilon),\\
\nu_{ij} &\sim \mathcal{N}(0, 1/\tau_\nu),\\
\tau_\epsilon &\sim \Gamma(0.001, 0.001),\\
\tau_\nu &\sim \Gamma(0.001, 0.001).
\end{aligned}
$$

Epil example III

- aghq package interfaces really easily with TMB!
- This is the code I used to fit the model with TMB

```
obj <- MakeADFun(
  data = dat,
  parameters = param,
  # These are the ones integrated out with a Laplace approximation
  random = c("epsilon", "nu"),
  DLL = "epil"
)
```

Epil example IV

- Then to fit it with `aghq` it's only a very small modification

```
fit <- aghq::marginal_laplace_tmb(
  obj,
  k = 3,
  startingvalue = c(param$beta, param$l_tau_epsilon, param$l_tau_nu)
)
```

	Stan	INLA_G	INLA_SL	INLA_L	TMB	glmmTMB	tmbstan	aghq
beta_0	1.572	1.626	1.573	1.573	1.579	1.579	1.571	1.573
sd(beta_0)	0.076	0.077	0.078	0.078	0.073	0.073	0.081	0.076
beta_1	-0.968	-0.927	-0.954	-0.956	-0.949	-0.949	-0.961	-0.955
sd(beta_1)	0.420	0.419	0.419	0.419	0.396	0.396	0.426	0.411
beta_2	0.879	0.859	0.880	0.881	0.880	0.880	0.875	0.881
sd(beta_2)	0.136	0.138	0.138	0.138	0.129	0.129	0.135	0.135
beta_3	-0.103	-0.101	-0.103	-0.104	-0.103	-0.103	-0.105	-0.103
sd(beta_3)	0.087	0.086	0.086	0.086	0.086	0.086	0.086	0.086
beta_4	0.488	0.471	0.484	0.485	0.490	0.490	0.489	0.486
sd(beta_4)	0.352	0.364	0.364	0.364	0.342	0.342	0.375	0.357
beta_5	0.356	0.340	0.350	0.351	0.349	0.349	0.360	0.351
sd(beta_5)	0.212	0.213	0.213	0.213	0.200	0.200	0.215	0.209

Plan

- Test `aghq` for toy Naomi example
  - Do as above with the Epil example, testing versus a long MCMC run
  - Sometimes you have to look pretty hard for a node (element of the latent field) where there are differences. In the INLA paper they do this by computing a SKLD and ordering by maximum difference. Probably good to do here as well
- Extend `aghq` to replicate INLA functionality by adding the more complex versions of $\tilde{p}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$ then test that wih Naomi
  - Håvard Rue philosophy: "do one thing and do it well"
  - `R-INLA` implementation of INLA based on sparsity of $\boldsymbol{Q}(\boldsymbol{\theta})$ that doesn't hold up for extended LGMs
  - Wood (2020) on how to still do it
- Try the INLA without `R-INLA` on other almost LGMs and see how far it can be pushed

References I

Amoah, Benjamin, Peter J Diggle, and Emanuele Giorgi. 2020. "A Geostatistical Framework for Combining Spatially Referenced Disease Prevalence Data from Multiple Diagnostics." *Biometrics* 76 (1): 158–70.

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*.

References II

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.

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.

Wood, Simon N. 2020. "Simplified Integrated Nested Laplace Approximation." *Biometrika* 107 (1): 223–30.