Fast approximate Bayesian inference for the Naomi model

Machine Learning and Global Health Network

Adam Howes

Imperial College London

June 2023

Doing precision public health requires granular data

1. The right interventions
2. in the right place
3. to the right populations
4. at the right time

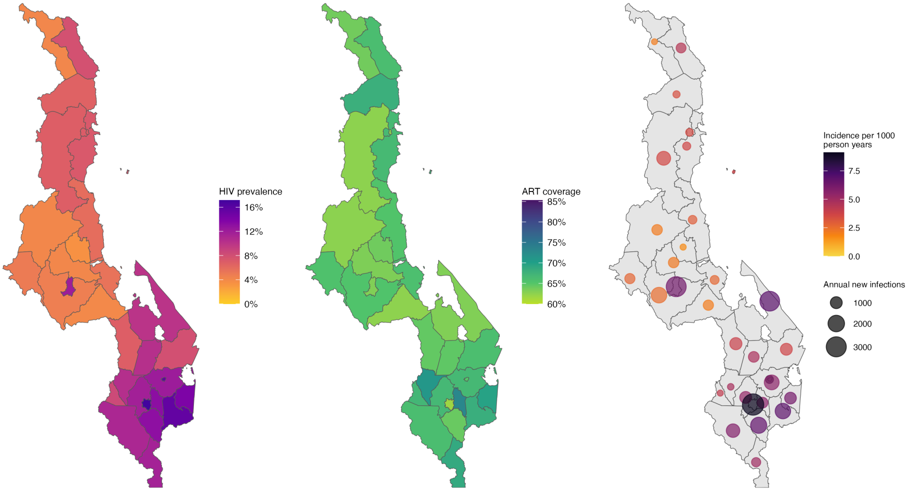

Figure 1: Naomi is a age-sex-district model of HIV indicators, like prevalence, treatment coverage, and incidence, for countries in sub-Saharan Africa.

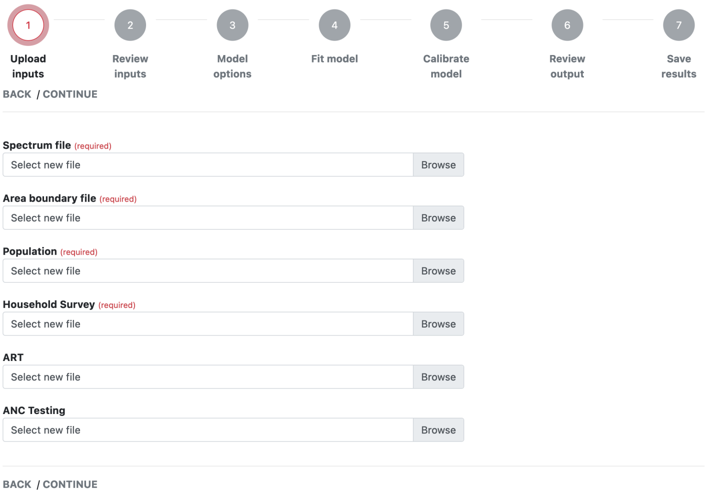

Figure 2: Estimates are generated yearly via a web interface. This promotes data ownership, data use, and data quality. From https://naomi.unaids.org/.

Integration of information from many sources

1. Household surveys infreqent, but gold-standard
2. Antenatal care clinic data frequent, only for pregnant women
3. Treatment service provision data frequent, but hard to interpret

A challenging Bayesian inference problem

We want our inference procedure to be

1. Fast enough for interactive review of estimates
2. Accurate enough for precision public health
3. Flexible enough for compatibility with a complex model

The model has a big, structured, Gaussian latent field **x**

- Fixed effects Gaussian
- Age structure AR1
- Spatial structure IID, ICAR, BYM2

Concatenate together as **x**, length 467

Smaller, non-Gaussian, hyperparameters $\boldsymbol{\theta}$

- Standard deviations Half-Gaussian
- BYM2 proportion parameters Beta
- AR1 autocorrelation parameters Uniform

Concatenate together as $\boldsymbol{\theta}$, length 24

Approximate the marginal posterior of $\mathbf{x}$ by a Gaussian

Given hyperparameters $\boldsymbol{\theta}$ we compute this as

$$\tilde{p}_{\mathrm{G}}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(\hat{\mathbf{x}}(\boldsymbol{\theta}), \hat{\mathbf{H}}(\boldsymbol{\theta}))$$

If you input 24 length $\boldsymbol{\theta}$ then it'll return a 467 length mean vector $\hat{\mathbf{x}}(\boldsymbol{\theta})$ and 467×467 length covariance matrix $\hat{\mathbf{H}}(\boldsymbol{\theta})$. Mean calculated using gradient based optimisation. Gradients, and Hessian, obtained using automatic differentiation.

Optimise the resulting Laplace approximation

$$\hat{\boldsymbol{\theta}}_{\mathrm{LA}} = \arg\max_{\boldsymbol{\theta}} \tilde{p}_{\mathrm{LA}}(\boldsymbol{\theta}, \mathbf{y}) = \arg\max_{\boldsymbol{\theta}} \frac{p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})}{\tilde{p}_{\mathrm{G}}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x} = \hat{\mathbf{x}}(\boldsymbol{\theta})}$$

Use adaptive Gauss-Hermite quadrature to integrate over $\boldsymbol{\theta}$

Quadrature method based on the theory of polynomial interpolation which:

1. Works well when the integrand looks like a polynomial times a Gaussian
2. Adapts to the particular integrand based on the mode and Hessian
3. Is implemented by the `aghq` package (Stringer 2021)

$$p(\mathbf{y}) \approx \int_{\boldsymbol{\theta}} \tilde{p}_{\text{LA}}(\boldsymbol{\theta}, \mathbf{y}) \text{d}\boldsymbol{\theta} \approx \sum_{\mathbf{z} \in \mathcal{Q}} p_{\text{LA}}(\mathbf{z}, \mathbf{y}) \omega(\mathbf{z})$$

where $\mathbf{z} \in \mathcal{Q}$ are a set of nodes and $\omega : \mathcal{Q} \to \mathbb{R}$ is a weighting function.

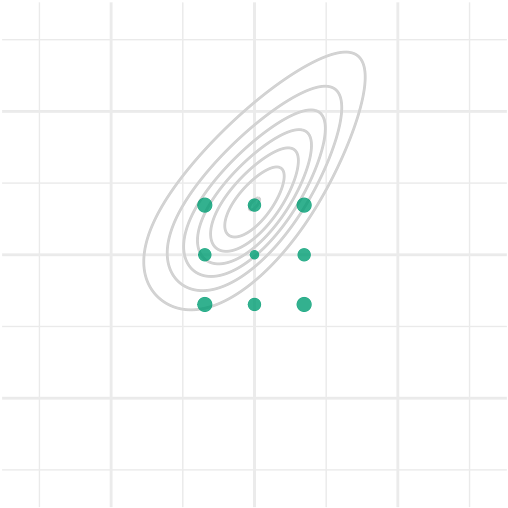

Figure 3: Unadapted Gauss-Hermite nodes in two dimensions with $k = 3$.

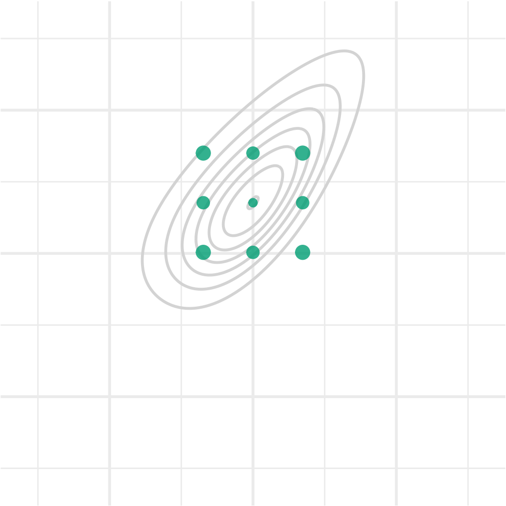

Figure 4: Add the mode $\mathbf{z} + \hat{\boldsymbol{\theta}}$.

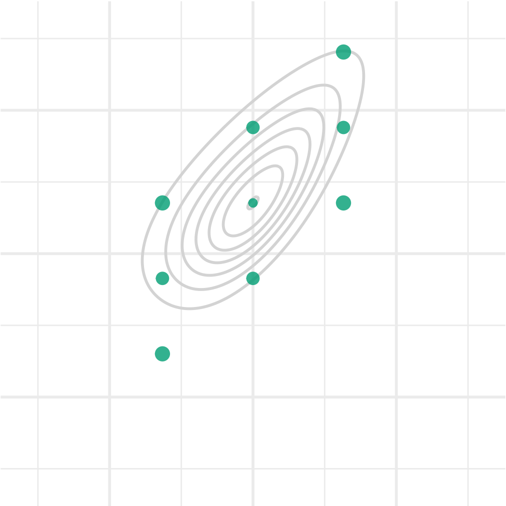

Figure 5: You could rotate by the lower Cholesky $\mathbf{Lz} + \hat{\boldsymbol{\theta}}$.

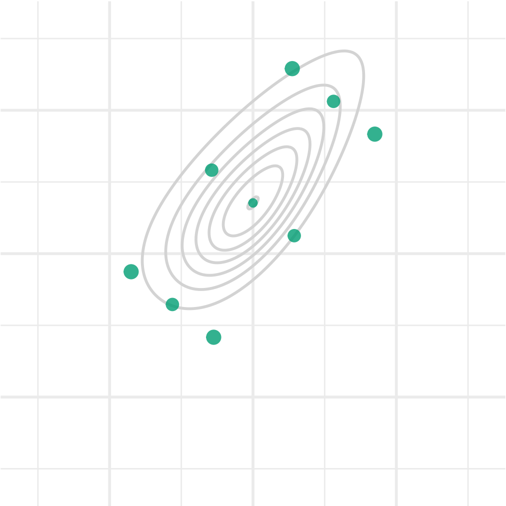

Figure 6: Or you could rotate using the eigendecomposition $\mathbf{E}\boldsymbol{\Lambda}^{1/2}\mathbf{z} + \hat{\boldsymbol{\theta}}$.

24 hyperparameters is too many for a dense grid

$k = 3$ points in 24 dimensions is not feasible

$$3 \times 3 \times \cdots \times 3 = 3^{24}$$

So we need to find something smaller (that still does a good job!)

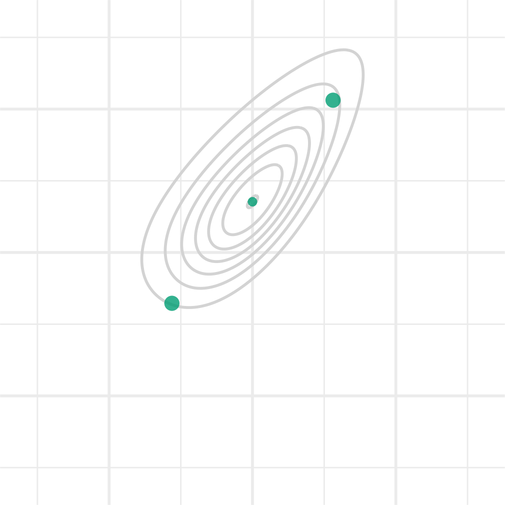

Figure 7: An obvious thing to try is only keeping points from the largest eigenvectors: we call this PCA-AGHQ. Corresponds to variable choice of k by dimension.

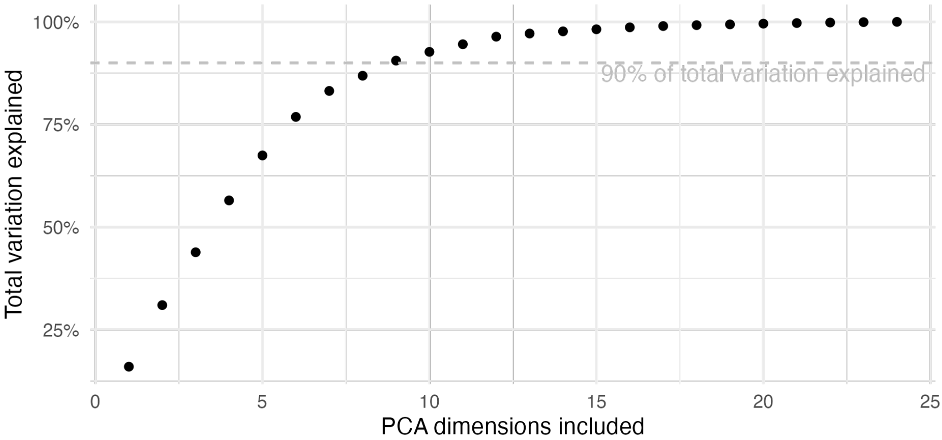

Figure 8: Including 8 dimensions you can explain close to 90% of the total variation.

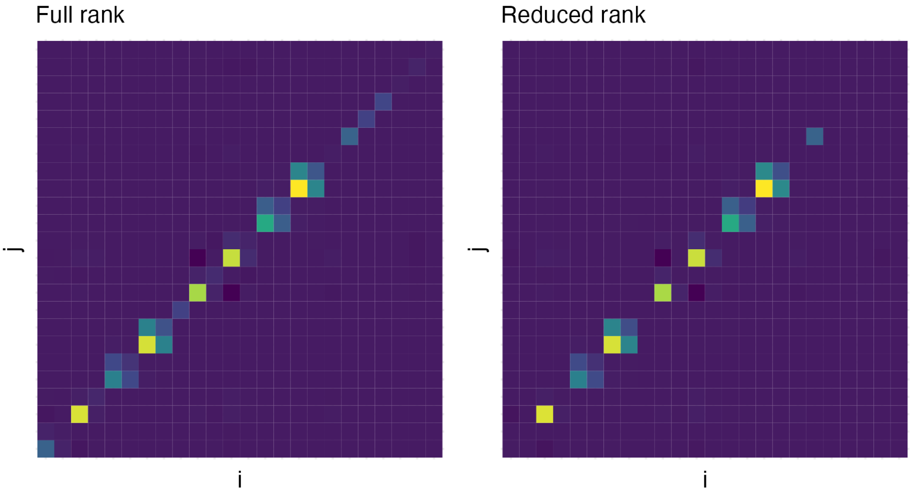

Figure 9: If you reconstruct the Hessian with just those 8 dimensions, it looks pretty similar.

Yes, but did it work?

Run NUTS[1] as gold-standard, then compare to TMB to PCA-AGHQ using:

1. Marginal distributions point estimates, ECDF e.g. KS or AD
2. Joint distributions PSIS, MMD
3. Policy relevant outcomes second 90, high incidence

[1] For 3 days!

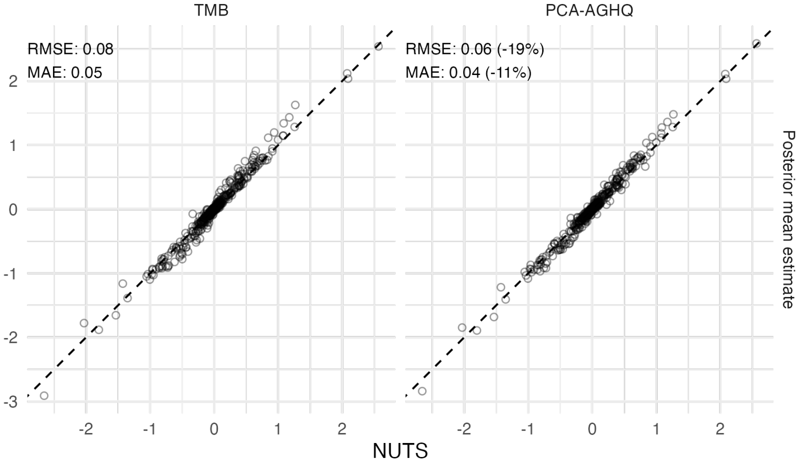

Figure 10: PCA-AGHQ modestly improves estimation of the posterior mean.

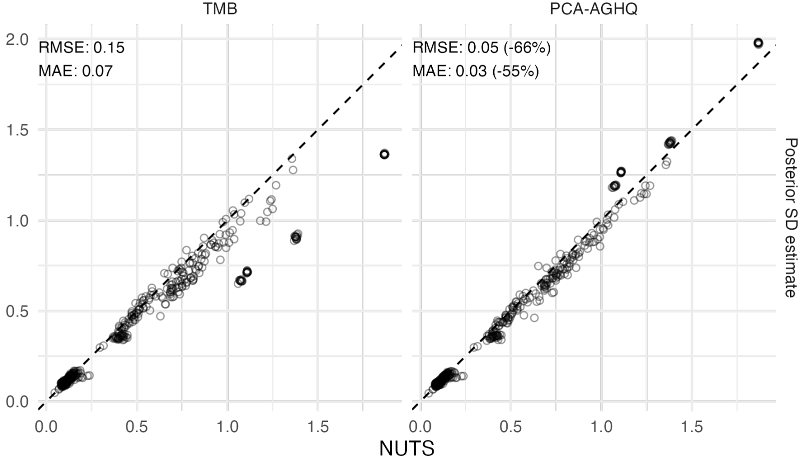

Figure 11: PCA-AGHQ substantially improves estimation of the posterior standard deviation. TMB systematically underestimates, which you'd expect.

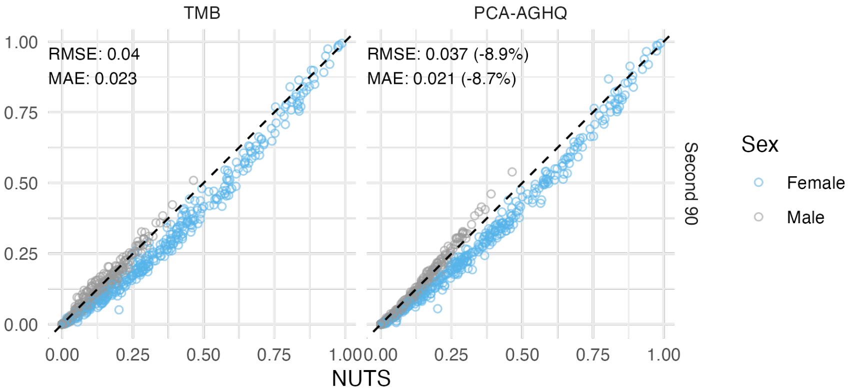

Figure 12: Strata probabilities of having greater than 81% ART coverage, and as such meeting the second 90 target. Both approximate methods are inaccurate for females.

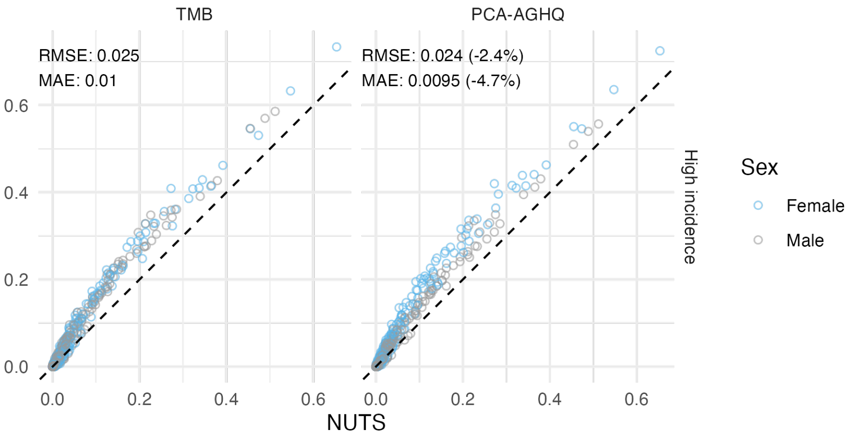

Figure 13: Strata probabilities of having greater than 1% HIV incidence, and as such being classified high incidence. Again, both approximate methods are inaccurate.

Further improvements look possible

1. Fix issues with scaling
   Logit scaled not uniformly more important than log scaled
2. Take into account importance for outputs of interest
   Variance of inputs isn't really what we care about
3. Take into account marginal skewness
   The more skewed, the more we should be placing lots of points

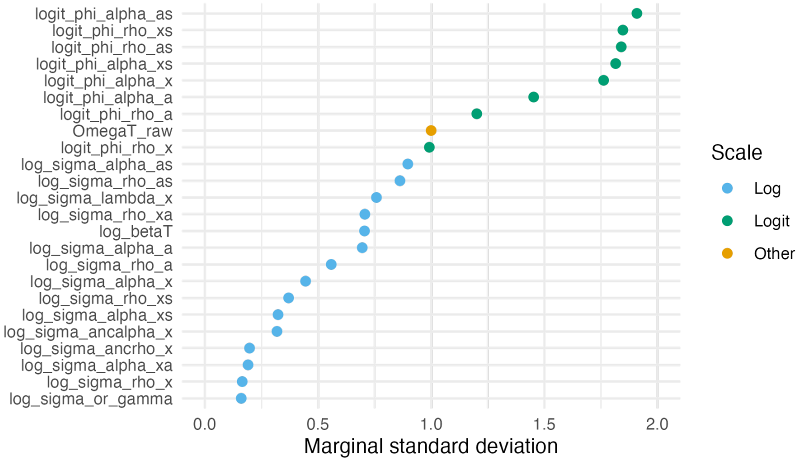

Figure 14: On the real scale, $[0, 1]$ hyperparameters appear to have more marginal variance than $[0, \infty)$ hyperparameters. This doesn't really make them more important though.

Get in touch to chat about

1. You have suggestions to improve the work done so far!
2. Further directions for this research
   e.g. suggestions for short masters or PhD projects
3. Impactful academic or industry jobs using Bayesian statistics
   to begin around the end of this year when I graduate (hopefully!)

Method	Details
Zulip	Adam Howes
Email	`ath19@ic.ac.uk`
Calendly	`adamthowes`

For more information

- Code and notebooks: `github.com/athowes/elgm-inf`
- Working paper on the way[2], Any early readers greatly appreciated!
  - Fast approximate Bayesian inference for small-area estimation of HIV indicators using the Naomi model Adam Howes, Alex Stringer, Seth Flaxman, Jeff Eaton

[2] For sufficiently vague definition of "on the way"

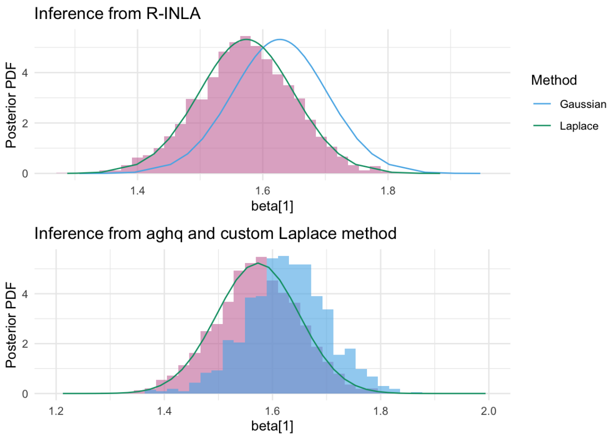

Figure 15: Custom version of Laplace marginals implemented using `TMB`. Speed-up possible using Simon Wood's method? For the Epilepsy example from Rue (2009), see `athowes/elgm-inf/src/epil`.

Figure 16: Much of this work done in Waterloo, Canada visiting Alex Stringer last fall! Would definitely recommend the SAS department.

References I

Stringer, Alex. 2021. "Implementing Approximate Bayesian Inference using Adaptive Quadrature: the aghq Package." https://arxiv.org/abs/2101.04468.