Fast approximate Bayesian inference of HIV indicators using PCA adaptive Gauss-Hermite quadrature

Adam Howes[1, 2], Alex Stringer[3], Seth R. Flaxman[4], Jeffrey W. Eaton[5, 2]

[1] Department of Mathematics, Imperial College London
[2] MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London
[3] Department of Statistics and Actuarial Science, University of Waterloo
[4] Department of Computer Science, University of Oxford
[5] Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health

Summary

- We developed an approximate Bayesian inference method using Laplace approximation, adaptive Gauss-Hermite quadrature and principal component analysis
- Motivation: an evidence synthesis model for small-area estimation of HIV indicators in >35 countries in sub-Saharan Africa
- Implemented using the `aghq` package (Stringer 2021), and compatible with any model with a Template Model Builder `TMB` (Kristensen et al. 2016) C++ user template

1. The Naomi HIV model

- District-level model of HIV indicators which synthesises data from 1) household surveys, 2) antenatal care (ANC) clinics, and 3) routine service provision of antiretroviral therapy (ART) (Eaton et al. 2021)
  - Combining evidence from multiple data sources helps overcome the limitations of any one
  - Small-area estimation methods (structured random effects) to overcome limited district-level sample sizes
- Yearly estimation process: model run interactively by by country teams using `naomi.unaids.org` web-app
  - Figure 1 illustrates the seven stages of using the app
- Inference conducted in minutes using empirical Bayes and a Gaussian approximation
- Days to get accurate answers with MCMC via `tmbstan` (Monnahan and Kristensen 2018): not practical!
- Naomi has a large latent field x controlled by a smaller number of hyperparameters θ
- Extended latent Gaussian model (Stringer, Brown, and Stafford 2022): more complex dependency structures than a latent Gaussian model
- Looking for a fast, approximate approach, that properly takes uncertainty in hyperparameters into account

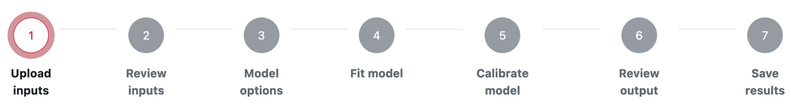

Figure 1: Model fitting occurs interactively in stages

2. Inference procedure

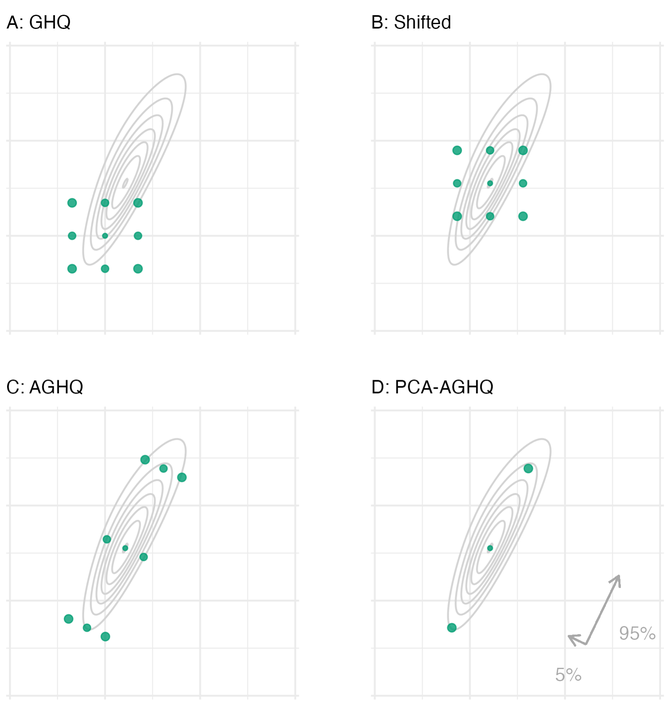

Figure 2: Demonstration of Gauss-Hermite quadrature, adaption, and our principal components approach.

- **Laplace approximation** Integrate out latent field using a Gaussian approximation to the denominator

$$p(\theta, y) \approx \tilde{p}_{\texttt{LA}}(\theta, y) = \frac{p(y, x, \theta)}{\tilde{p}_{\texttt{G}}(x \mid \theta, y)}\Big|_{x=\hat{x}(\theta)},$$

where $\tilde{p}_{\texttt{G}}(x \mid \theta, y) = \mathcal{N}(x \mid \hat{x}(\theta), \hat{H}(\theta)^{-1})$
  - Use automatic differentiation via `CppAD` in `TMB`
- **Adaptive Gauss-Hermite Quadrature** (AGHQ) perform quadrature over the hyperparameters

$$\int_\Theta p_{\texttt{LA}}(\theta, y) \mathrm{d}\theta \approx |L| \sum_{z \in \mathcal{Q}(m,k)} p_{\texttt{LA}}(\hat{\theta} + Lz, y)\omega(z),$$

where $m = \dim(\theta)$ and the Gauss-Hermite quadrature rule $z \in \mathcal{Q}(m, k)$ with weights $\omega : \mathcal{Q} \mapsto \mathbb{R}$ and k points per dimension is adapted (Figure 2) based upon
  - The mode $\hat{\theta} = \text{argmax}_{\theta \in \Theta} p_{\texttt{LA}}(\theta, y)$
  - A matrix decomposition $LL^\top = -\partial^2_\theta \log p_{\texttt{LA}}(\theta, y)|_{\theta=\hat{\theta}}$
- **PCA-AGHQ** To integrate over large spaces, use the spectral decomposition $L = E\Lambda^{1/2}$ and keep only the first $s < m$ principal components

3. Application to Malawi

- Malawi is a relatively small country but still has latent field $\dim(x) = 467$ and hyperparameters $\dim(\theta) = 24$

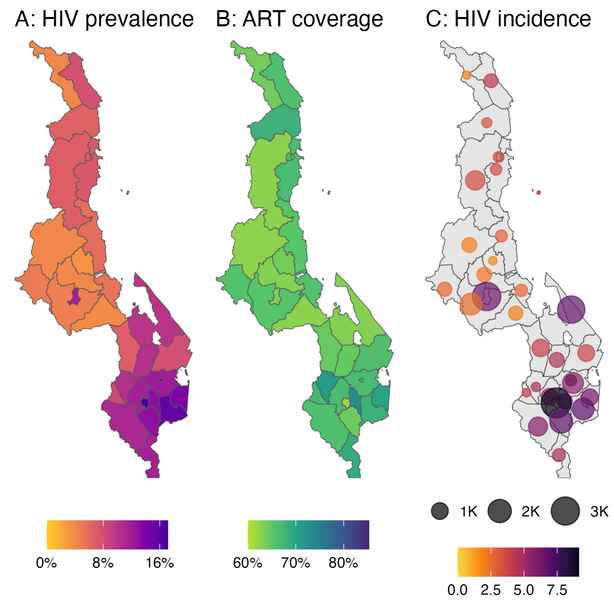

Figure 3: District-level model outputs for adults 15-49.

- Run three methods: 1. TMB (baseline, 54 secs), 2. PCA-AGHQ (new, 1 hour), 3. NUTS (gold-standard, 3.3 days)
  - For PCA-AGHQ $k = 3$ and $s = 8$ chosen using Scree plot to explain ~90% of variance
  - For NUTS 4 chains of 100,000 thinned by 40 were required for good diagnostics e.g. all $\hat{R} < 1.025$

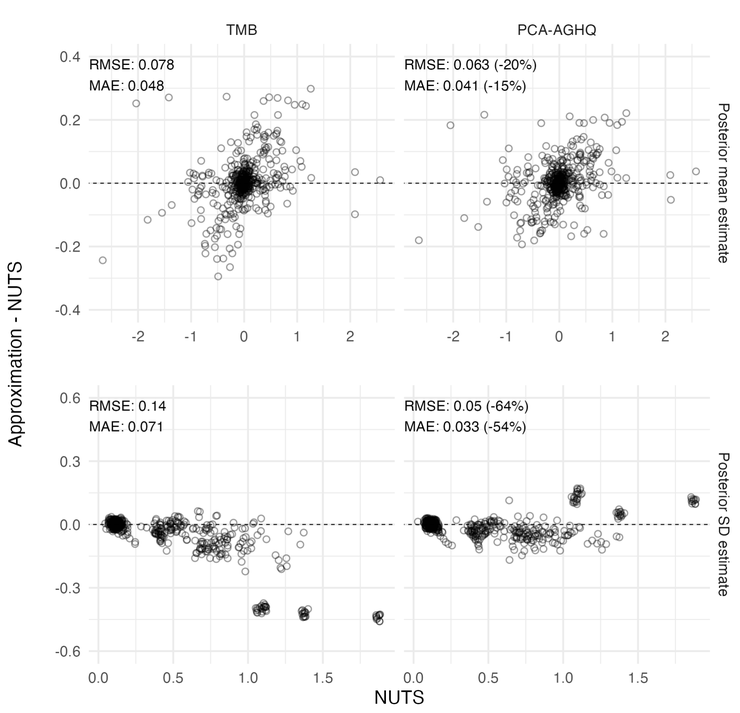

Figure 4: PCA-AGHQ moderately improves the posterior mean and substantially improves the posterior standard deviation

- PCA-AGHQ improves latent field point estimates (Figure 4) and distributional quantities like the Kolmogorov-Smirnov (KS, Figure 5) test statistic (-9% on average)

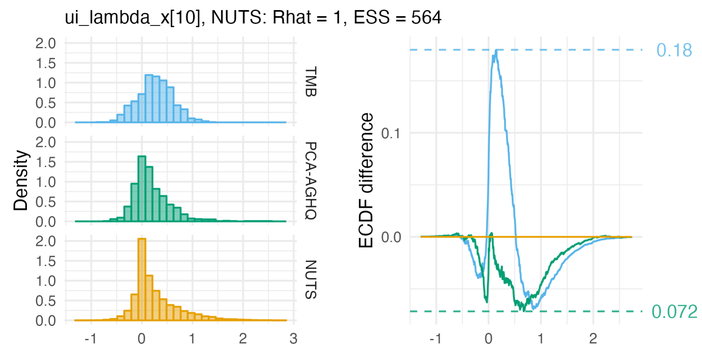

Figure 5: KS test demonstration.

- Naomi can be used to assess probabilities targets have been met e.g. 90% of those who know their HIV status are on ART ("second 90")
  - Though PCA-AGHQ is better (-16% RMSE), both TMB and PCA-AGHQ are biased (Figure 6)

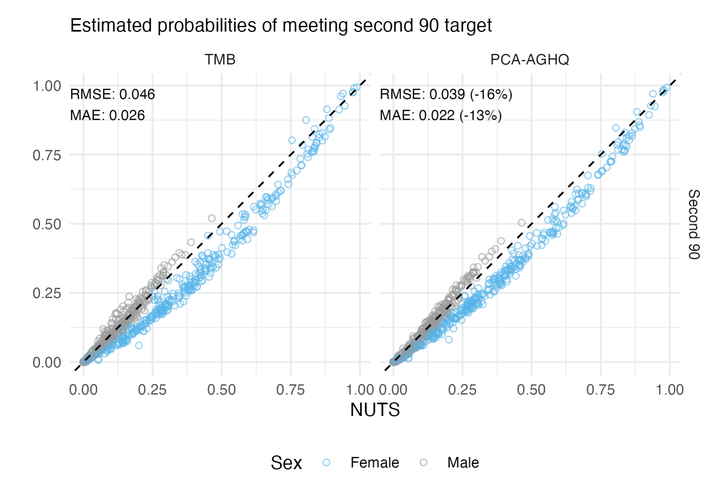

Figure 6: PCA-AGHQ moderately improves the second 90 bias.

4. Future directions

- Can we do better than moderate improvements?
  - Especially for the quantities of policy interest
- Laplace marginals with matrix algebra approximations (Wood 2020) to speed up calculations
- Further methods for allocation of effort to "important" dimensions of hyperparameter grid

Contact

- `github.com/athowes/naomi-aghq`
- `athowes.github.io`
- `ath19@ic.ac.uk`

Funding AH was supported by the EPSRC and Bill & Melinda Gates Foundation. This research was supported by the MRC Centre for Global Infectious Disease Analysis.

References

Eaton, Jeffrey W., Laura Dwyer-Lindgren, Steve Gutreuter, Megan O'Driscoll, Oliver Stevens, Sumali Bajaj, Rob Ashton, et al. 2021. "Naomi: A New Modelling Tool for Estimating HIV Epidemic Indicators at the District Level in Sub-Saharan Africa." *Journal of the International AIDS Society* 24 (S5): e25788.

Kristensen, Kasper, Anders Nielsen, Casper W Berg, Hans Skaug, Bradley M Bell, et al. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70 (i05).

Monnahan, Cole C, and Kasper Kristensen. 2018. "No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages." *PloS One* 13 (5): e0197954.

Stringer, Alex. 2021. "Implementing Approximate Bayesian Inference using Adaptive Quadrature: the aghq Package." *arXiv Preprint arXiv:2101.04468*.

Stringer, Alex, Patrick Brown, and Jamie Stafford. 2022. "Fast, Scalable Approximations to Posterior Distributions in Extended Latent Gaussian Models." *Journal of Computational and Graphical Statistics*, 1–15.

Wood, Simon N. 2020. "Simplified integrated nested Laplace approximation." *Biometrika* 107 (1): 223–30.