Stochastic Convergence Rates and Applications of Adaptive Quadrature in Bayesian Inference

Blair Bilodeau [1,2] Alex Stringer [1,3] Yanbo Tang [1,2]

[1] University of Toronto [2] Vector Institute [3] Center for Global Health Research, St. Michael's Hospital

VECTOR INSTITUTE

cghr

Contribution Summary

- **First stochastic convergence rates** of multi-point adaptive quadrature for Bayesian inference.
- **Fast procedure for high-dimensional models** combining multi- and single-point adaptive quadrature techniques.
- **Efficient and robust implementation** of these methods in the `aghq` package, demonstrated on multiple challenging examples.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (Y_1, \dots, Y_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

Inference: $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}^{(n)}) = \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})/\pi(\boldsymbol{Y}^{(n)})$

We want moments, quantiles, marginals, predictive distributions, etc.

MCMC: Draw samples from an *implicit* approximation of $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}^{(n)})$

- Easy to obtain summaries once you have samples.
- Theoretical guarantees exist.
- Can be (computationally) challenging, even for low-dimensions.

Quadrature: Numerically obtains an *explicit* approximation of $\pi(\boldsymbol{Y}^{(n)})$

- Software can make this *very fast* and *stable*.
- Theoretical guarantees for the stochastic setting are limited.

Adaptive Numerical Quadrature

Numerical Quadrature

Approximate an integral using

$$\int_\Theta f(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \approx \sum_{\boldsymbol{z} \in \mathcal{Q}} f(\boldsymbol{z})\omega(\boldsymbol{z}).$$

Adaptive Quadrature

Recall $\pi(\boldsymbol{Y}^{(n)}) = \int \pi(\boldsymbol{\theta}, \boldsymbol{Y}^{(n)})\mathrm{d}\boldsymbol{\theta}$.

The function $f_n(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}, \boldsymbol{Y}^{(n)})$ changes with n through $\boldsymbol{Y}^{(n)}$.

Have to adapt the numerical quadrature rule to $\boldsymbol{Y}^{(n)}$.

Naylor and Smith (1982) propose...

Mode: $\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \Theta} \pi(\boldsymbol{Y}^{(n)}, \boldsymbol{\theta})$

Curvature: $\widehat{\boldsymbol{L}}_n =$ lower Cholesky of $-\partial^2_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}, \boldsymbol{Y}^{(n)})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n}$

$$\widetilde{\pi}(\boldsymbol{Y}^{(n)}) = |\widehat{\boldsymbol{L}}_n| \sum_{\boldsymbol{z} \in \mathcal{Q}} \pi(\widehat{\boldsymbol{L}}_n \boldsymbol{z} + \widehat{\boldsymbol{\theta}}_n, \boldsymbol{Y}^{(n)})\,\omega(\boldsymbol{z}).$$

Low-Dimensional Example

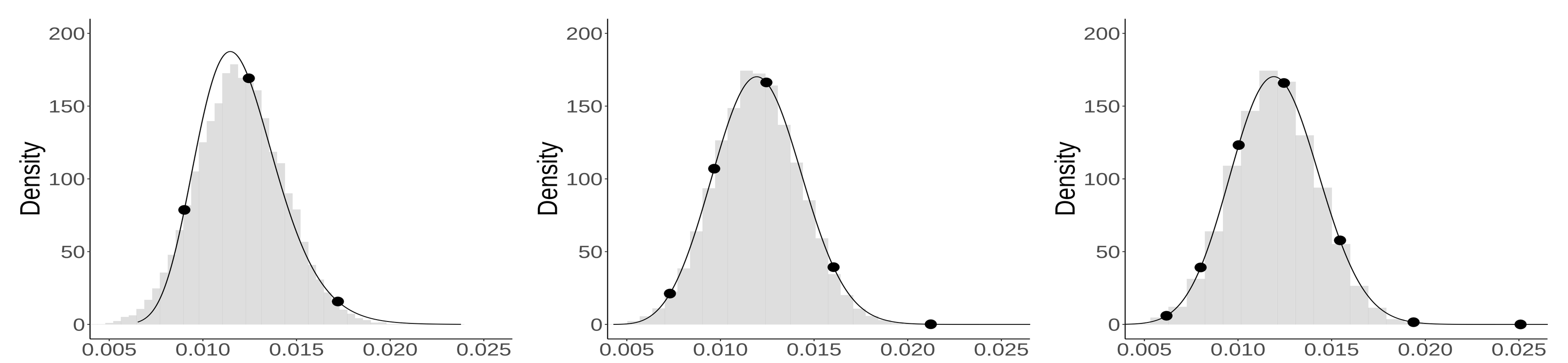
Spread of disease among a grid of $n = 520$ tomato plants.

`EpiILMCT`: continuous-time, individual-levels models for disease spread.

The `aghq` package provides:

- Optimization + Quadrature
- Summary statistics (expectations, quantiles)
- Marginal densities

AGHQ is practically indistinguishable from MCMC and runs in milliseconds.



AGHQ (•, —) with $k \in \{3, 5, 7\}$ v.s. MCMC (■) for infectivity scale parameter.

Our theoretical convergence rate applies here!

Stochastic Convergence Rate

Suppose...

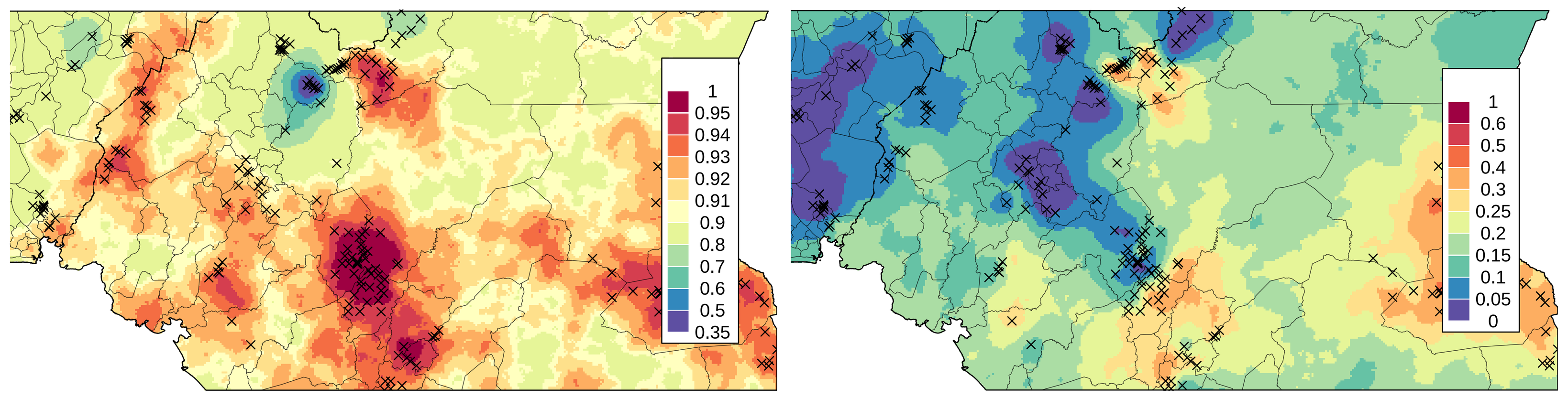
- the base quadrature rule exactly integrates polynomials of degree $\leq 2k - 1$
- the model $\pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta})$ satisfies standard regularity conditions.

Then,

$$\widetilde{\pi}(\boldsymbol{Y}^{(n)}) = \pi(\boldsymbol{Y}^{(n)})\left[1 + \mathcal{O}_P\left(n^{-\lfloor \frac{k+2}{3} \rfloor}\right)\right].$$

High-Dimensional Example

Zero-inflated binomial regression for spread of a tropical disease (loaloa).



AGHQ posterior mean (a) suitability probabilities and (b) incidence rates.

A Bayesian version of this model has not previously been fit!

High-Dimensional Procedure

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dim $\boldsymbol{\theta}$ and high-dim $\boldsymbol{W}$.
- Apply AGHQ with $k = 1$ to $\int \pi(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{Y}^{(n)})\mathrm{d}\boldsymbol{W}$.
  - Initially proposed by Tierney and Kadane (1986) to obtain $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}^{(n)})$
  - They ignore renormalization; we use another application of AGHQ
- Apply a Gaussian approximation to $\pi(\boldsymbol{W} \mid \boldsymbol{\theta}, \boldsymbol{Y}^{(n)})$.
  - Proposed by Stringer et al. (2021); builds on Rue et al. (2009)

Gauss Quadrature Rules

Gauss-Hermite (e.g., Davis and Rabinowitz, 1975)

- Nodes are the zeroes of the kth *Hermite polynomial*
- Weights defined to exactly integrate desired polynomials

In p dimensions, our theorem covers (for example)...

Product Rule: Gauss-Hermite using exponential in p points.

Smolyak Rule: Gauss-Hermite using polynomial in p points.

Any rule that satisfies the necessary exact integration applies.

Theoretical Considerations

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result.
- In high dimensions ($p \approx n$), the posterior may not concentrate and the Gaussian adaptation may be unsuitable.

Previous Work

Kass et al. (1990) prove a convergence rate for $k = 1$.

We match their rate, and extend the proof to $p > 1$ dimensions.

Jin and Andersson (2020) study AGHQ in a restricted setting.

- Their integrand cannot vary with n, and consequently the result does not apply to likelihoods (or Bayesian inference).
- They require properties to hold almost surely for all n, while we use properties that hold in the probabilistic limit.

NSERC CRSNG