Stochastic Convergence Rates and Applications of Adaptive Quadrature in Bayesian Inference

Blair Bilodeau[1,2] (with Alex Stringer[1,3] and Yanbo Tang[1,2])

June 29, 2021

World Meeting of the International Society for Bayesian Analysis

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

Approximate Bayesian Inference

Approximate Bayesian Inference

Setup

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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 $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}^{(n)})$

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$
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-dimensional models.

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

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

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

- Can be *very fast* and *stable*, but the quadrature rule must adapt to the data.
- Theoretical guarantees for the stochastic setting are limited.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

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

- Can be *very fast* and *stable*, but the quadrature rule must adapt to the data.
- Theoretical guarantees for the stochastic setting are limited.

We provide...

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

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

- Can be *very fast* and *stable*, but the quadrature rule must adapt to the data.
- Theoretical guarantees for the stochastic setting are limited.

We provide...

Stochastic convergence rates of multi-point quadrature for Bayesian inference.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

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

- Can be *very fast* and *stable*, but the quadrature rule must adapt to the data.
- Theoretical guarantees for the stochastic setting are limited.

We provide...

Stochastic convergence rates of multi-point quadrature for Bayesian inference.

A fast procedure to use numerical quadrature for high-dimensional models.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

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

- Can be *very fast* and *stable*, but the quadrature rule must adapt to the data.
- Theoretical guarantees for the stochastic setting are limited.

We provide...

Stochastic convergence rates of multi-point quadrature for Bayesian inference.

A fast procedure to use numerical quadrature for high-dimensional models.

An efficient and robust implementation of these methods in the `aghq` package.

Approximate Bayesian Inference

Setup

Model: $\boldsymbol{Y}^{(n)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \sim \pi(\boldsymbol{Y}^{(n)} \mid \boldsymbol{\theta}^*)$, $\boldsymbol{\theta}^* \in \Theta \subseteq \mathbb{R}^p$

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-dimensional models.

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

- Can be *very fast* and *stable*, but the quadrature rule must adapt to the data.
- Theoretical guarantees for the stochastic setting are limited.

We provide...

Stochastic convergence rates of multi-point quadrature for Bayesian inference.

A fast procedure to use numerical quadrature for high-dimensional models.

An efficient and robust implementation of these methods in the `aghq` package.

Adaptive Numerical Quadrature

Adaptive Numerical Quadrature

Numerical Quadrature

Approximate an integral using

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

Adaptive Numerical Quadrature

Numerical Quadrature

Approximate an integral using

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

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 $\boldsymbol{Y}^{(n)}$.

Adaptive Numerical Quadrature

Numerical Quadrature

Approximate an integral using

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

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 $-\partial^2_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}, \boldsymbol{Y}^{(n)})\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n}$

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

Adaptive Numerical Quadrature

Numerical Quadrature

Approximate an integral using

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

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 $-\partial^2_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}, \boldsymbol{Y}^{(n)})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n}$

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

Convergence Rates

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

Convergence Rates

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

Theorem (BST21)

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,

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

Convergence Rates

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

Theorem (BST21)

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,

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

Gauss-Hermite Rule: in 1-dimension, defines $(z_i)_{i \in [k]}$ and $\omega : \mathbb{R} \to \mathbb{R}_+$ that exactly integrates polynomials of degree $\leq 2k-1$.

Convergence Rates

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

Theorem (BST21)

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,

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

Gauss-Hermite Rule: in 1-dimension, defines $(z_i)_{i \in [k]}$ and $\omega : \mathbb{R} \to \mathbb{R}_+$ that exactly integrates polynomials of degree $\leq 2k-1$.

The adaptive version of the Gauss-Hermite rule (AGHQ) is implemented in `aghq`.

Convergence Rates

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

Theorem (BST21)

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,

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

Gauss-Hermite Rule: in 1-dimension, defines $(z_i)_{i \in [k]}$ and $\omega : \mathbb{R} \to \mathbb{R}_+$ that exactly integrates polynomials of degree $\leq 2k - 1$.

The adaptive version of the Gauss-Hermite rule (AGHQ) is implemented in `aghq`.

In p dimensions...

Product Rule: Exact integration using exponential in p points (grid).

Convergence Rates

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

Theorem (BST21)

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,

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

Gauss-Hermite Rule: in 1-dimension, defines $(z_i)_{i \in [k]}$ and $\omega : \mathbb{R} \to \mathbb{R}_+$ that exactly integrates polynomials of degree $\leq 2k-1$.

The adaptive version of the Gauss-Hermite rule (AGHQ) is implemented in `aghq`.

In p dimensions...

Product Rule: Exact integration using exponential in p points (grid).

Smolyak Rule: Exact integration using polynomial in p points (sparse grid).

(Heiss and Winschel, 2008)

Convergence Rates

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

Theorem (BST21)

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,

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

Gauss-Hermite Rule: in 1-dimension, defines $(z_i)_{i \in [k]}$ and $\omega : \mathbb{R} \to \mathbb{R}_+$ that exactly integrates polynomials of degree $\leq 2k - 1$.

The adaptive version of the Gauss-Hermite rule (AGHQ) is implemented in `aghq`.

In p dimensions...

Product Rule: Exact integration using exponential in p points (grid).

Smolyak Rule: Exact integration using polynomial in p points (sparse grid).

(Heiss and Winschel, 2008)

Low-Dimensional Example

Low-Dimensional Example

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

Low-Dimensional Example

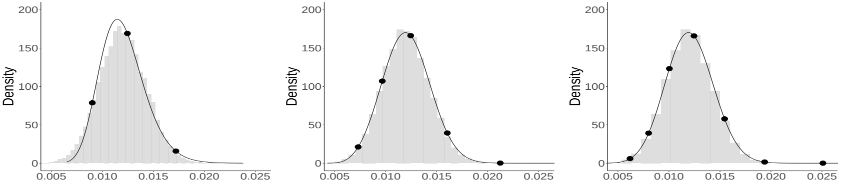
- Spread of disease among a grid of $n = 520$ tomato plants.
- `EpiILMCT`: continuous-time, individual-levels models for disease spread.

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

Low-Dimensional Example

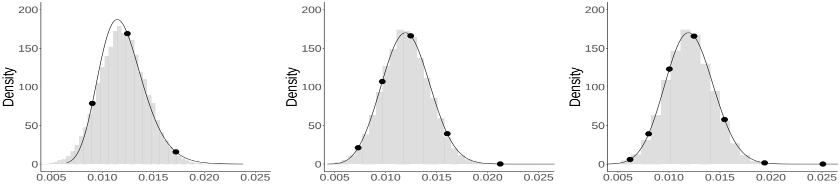
- Spread of disease among a grid of $n = 520$ tomato plants.
- `EpiILMCT`: continuous-time, individual-levels models for disease spread.
- The `aghq` package provides:
  - Optimization
  - Quadrature
  - Summary statistics (expectations, quantiles)
  - Marginal densities
- AGHQ is practically indistinguishable from MCMC and runs in milliseconds.



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

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.

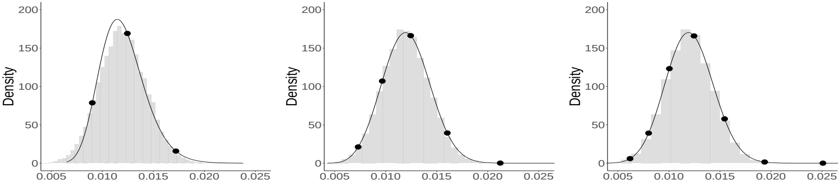


AGHQ (•, —) with $k \in \{3, 5, 7\}$ v.s. MCMC (■) for infectivity scale parameter.

Our theoretical convergence rate applies here!

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!

High-Dimensional Example

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (AGHQ with $k = 1$) to $\int \pi(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{Y}^{(n)}) \mathrm{d}\boldsymbol{W}$.

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (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

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (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 $\pi(\boldsymbol{W} \mid \boldsymbol{\theta}, \boldsymbol{Y}^{(n)})$.

High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (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)

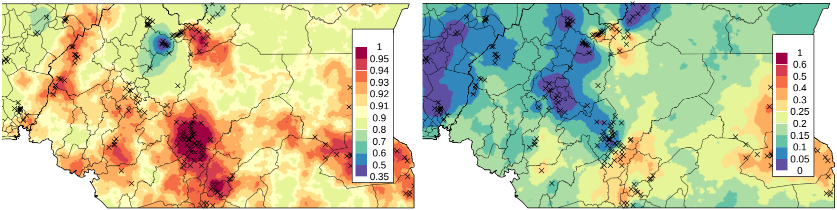
High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (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)

Example: Zero-inflated binomial regression for loaloa spread (Giorgi et al., 2018).



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

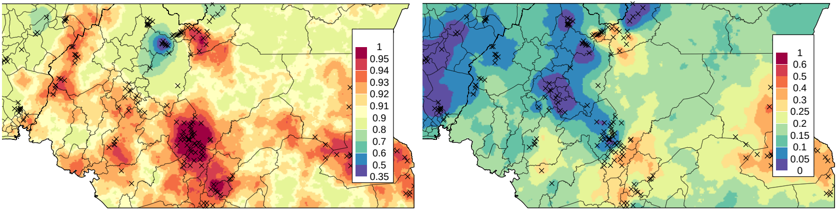
High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, aghq implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (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)

Example: Zero-inflated binomial regression for loaloa spread (Giorgi et al., 2018).



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

Open problem to develop theoretical guarantees for this procedure!

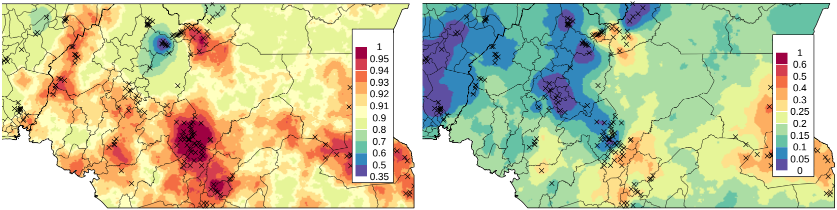
High-Dimensional Example

Even sparse rules are infeasible for large dimension p.

Instead, `aghq` implements the following procedure:

- Split the parameter into a low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{W}$.
- Apply a Laplace approximation (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)

Example: Zero-inflated binomial regression for loaloa spread (Giorgi et al., 2018).



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

Open problem to develop theoretical guarantees for this procedure!

Theoretical Details

Theoretical Details

High-Dimensional Convergence

Theoretical Details

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result; see Kleijn and van der Vaart (2012) for BVM analogue.

Theoretical Details

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result; see Kleijn and van der Vaart (2012) for BVM analogue.
- In high dimensions ($p \approx n$), the posterior may not concentrate and the Gaussian adaptation may be unsuitable. More structure required.

Theoretical Details

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result; see Kleijn and van der Vaart (2012) for BVM analogue.
- In high dimensions ($p \approx n$), the posterior may not concentrate and the Gaussian adaptation may be unsuitable. More structure required.

Previous Work

Theoretical Details

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result; see Kleijn and van der Vaart (2012) for BVM analogue.
- In high dimensions ($p \approx n$), the posterior may not concentrate and the Gaussian adaptation may be unsuitable. More structure required.

Previous Work

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

We match their rate, and provide the first explicit proof for $p > 1$ dimensions.

Theoretical Details

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result; see Kleijn and van der Vaart (2012) for BVM analogue.
- In high dimensions ($p \approx n$), the posterior may not concentrate and the Gaussian adaptation may be unsuitable. More structure required.

Previous Work

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

We match their rate, and provide the first explicit proof for $p > 1$ dimensions.

Jin and Andersson (2020) study deterministic 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.

Theoretical Details

High-Dimensional Convergence

- Convergence of our AGHQ renormalization requires a *misspecified* variant of our result; see Kleijn and van der Vaart (2012) for BVM analogue.
- In high dimensions ($p \approx n$), the posterior may not concentrate and the Gaussian adaptation may be unsuitable. More structure required.

Previous Work

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

We match their rate, and provide the first explicit proof for $p > 1$ dimensions.

Jin and Andersson (2020) study deterministic 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.

Summary of Contributions

Summary of Contributions

- The first stochastic convergence rates for approximate Bayesian inference via multi-point adaptive quadrature.
  - Our rate recovers the previous result for Laplace approximations.
  - Our rate applies to usual AGHQ as well as sparse versions.
  - Rate appears tight for some models, remains open to prove lower bounds.

Summary of Contributions

- The first stochastic convergence rates for approximate Bayesian inference via multi-point adaptive quadrature.
  - Our rate recovers the previous result for Laplace approximations.
  - Our rate applies to usual AGHQ as well as sparse versions.
  - Rate appears tight for some models, remains open to prove lower bounds.
- Using AGHQ within approximation methods for high-dimensional models.
  - Applies to a broader class of models than `R-INLA`.
  - We highlight remaining challenges to prove theory for high-dimensional AGHQ.

Summary of Contributions

- The first stochastic convergence rates for approximate Bayesian inference via multi-point adaptive quadrature.
  - Our rate recovers the previous result for Laplace approximations.
  - Our rate applies to usual AGHQ as well as sparse versions.
  - Rate appears tight for some models, remains open to prove lower bounds.
- Using AGHQ within approximation methods for high-dimensional models.
  - Applies to a broader class of models than `R-INLA`.
  - We highlight remaining challenges to prove theory for high-dimensional AGHQ.
- A robust implementation of AGHQ in the `aghq` package, available on CRAN.
  - Package: https://github.com/awstringer1/aghq
  - Paper: https://arxiv.org/abs/2102.06801.
  - Code: https://github.com/awstringer1/aghq-paper-code.
  - Vignette: https://arxiv.org/abs/2101.04468.

Summary of Contributions

- The first stochastic convergence rates for approximate Bayesian inference via multi-point adaptive quadrature.
  - Our rate recovers the previous result for Laplace approximations.
  - Our rate applies to usual AGHQ as well as sparse versions.
  - Rate appears tight for some models, remains open to prove lower bounds.
- Using AGHQ within approximation methods for high-dimensional models.
  - Applies to a broader class of models than `R-INLA`.
  - We highlight remaining challenges to prove theory for high-dimensional AGHQ.
- A robust implementation of AGHQ in the `aghq` package, available on CRAN.
  - Package: https://github.com/awstringer1/aghq
  - Paper: https://arxiv.org/abs/2102.06801.
  - Code: https://github.com/awstringer1/aghq-paper-code.
  - Vignette: https://arxiv.org/abs/2101.04468.