Generalized Posteriors in Approximate Bayesian Computation

Sebastian Schmon[1], **Patrick Cannon**[1], Jeremias Knoblauch[2]

Improbable[1], University of Warwick[2], The Alan Turing Institute[2]

June 2021

Overview

- Cast approximate Bayesian computation (ABC) as exact Bayesian inference in the presence of an implicit error model (see Wilkinson, 2013).
- Draw connections to generalized Bayesian inference (GBI).
- Use the connection to develop our understanding of kernel choice in ABC, particularly with respect to misspecification (see e.g. David T. Frazier, Robert, and Rousseau, 2020).

Problem setting

- Simulator $f(\cdot \mid \theta)$ parameterized by $\theta \in \Theta$, intractable.
- Observations $y \sim p^*(\cdot)$.
- Misspecified so that $p^* \notin \{f(\cdot \mid \theta); \theta \in \Theta\}$, and we instead assume the true DGP satisfies $p^* \in \{p(\cdot \mid \theta); \theta \in \Theta\}$ with

$$p(y \mid \theta) = \int_{\mathcal{X}} g(y \mid x) f(x \mid \theta)\,\mathrm{d}x \quad \text{and} \quad p(\theta \mid y) \propto p(y \mid \theta) p(\theta)$$

ABC

Algorithm 1: ABC

Input: Simulator f, prior p, data y, summary statistics η, probability kernel K_h

for $i = 1 : N$ **do**

  Sample from the prior $\theta_i \sim p(\cdot)$

  Sample from the simulator $x_i \sim f(\cdot \mid \theta_i)$

Calculate weights $w_i = K_h(\|\eta(x_i) - \eta(y)\|)$

Output: Weighted approximate posterior samples $\{\theta_i, w_i\}$

ABC

This induces an augmented distribution

$$p(\theta, x \mid y) := K_h(\|\eta(x) - \eta(y)\|)f(x|\theta)p(\theta). \tag{1}$$

The marginal

$$p_{\mathrm{ABC}}(\theta \mid y) := \int p(\theta, x \mid y)\,\mathrm{d}x \tag{2}$$

is taken as an approximation to the Bayesian posterior $p(\theta \mid y)$.

The general view

In GBI, 'generalized' Bayesian posteriors of the following form are considered

$$p_\ell(\theta \mid y) \propto \int \exp\{-w \cdot \ell(y; x)\} f(x \mid \theta) p(\theta) \,\mathrm{d}x. \tag{3}$$

The choice $w = 1$ and

$$\ell(y; x) = -\log K_h(\|\eta(x) - \eta(y)\|) \tag{4}$$

recovers the ABC posterior.

The general view

- A typical ABC choice is the 'hard-threshold' kernel

$$K_\varepsilon(x, y) = \mathbb{1}\{\|\eta(x) - \eta(y)\| < \varepsilon\}$$

  which corresponds to noise model $g(y \mid x) = K_\varepsilon(x, y)$.

- GBI framework liberates us to pick any loss function we find meaningful.

Simulation study

Toy example as in David T. Frazier, Robert, and Rousseau, 2020.

We consider three choices of loss function, namely

ST-ABC $\tilde{\ell}(x, y) = \|\eta(x) - \eta(y)\|^2$, equivalent to standard ABC Gaussian kernel,

K2-ABC $\tilde{\ell}(x, y) = \widehat{\mathrm{MMD}}^2(x, y)$ as used in Park, Jitkrittum, and Sejdinovic, 2016,

W-ABC $\tilde{\ell}(x_i, y) = \mathcal{W}(x, y)$, related to Bernton et al., 2019.

Simulation study

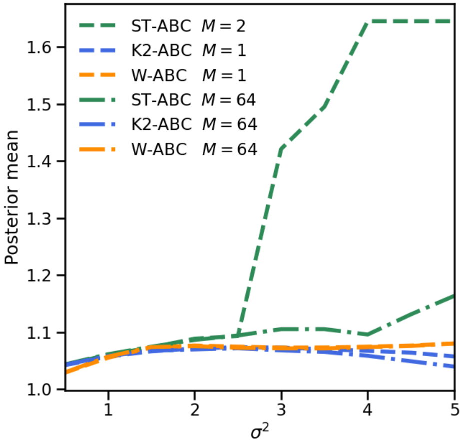

Figure: The posterior mean as a function of the degree of misspecification of the model. A comparison of ABC algorithms with a soft-threshold loss (ST-ABC), a maximum mean discrepancy (MMD) loss (K2-ABC), and a Wasserstein loss (W-ABC).

Extension: signature kernel ABC

work with **Joel Dyer** (Mathematical Institute, Oxford)

Introduction

Recall that the generalized posterior is given by

$$p_\ell(\theta \mid y) \propto \int \exp\{-w \cdot \ell(y; x)\} f(x \mid \theta) p(\theta) \, \mathrm{d}x. \tag{5}$$

What choices of loss function ℓ are appropriate for time-series?

We investigate the *path signature*.

Path signatures

The *signature* of a path $X = (X^1, X^2, \dots, X^d) : [0, T] \to \mathbb{R}^d$ is an infinite collection of statistics that characterizes the path up to a negligible equivalence class. It is defined by the infinite collection of statistics

$$\mathrm{Sig}(X) = (1, S(X)^1_{0,T}, S(X)^2_{0,T}, \dots, S(X)^d_{0,T}, S(X)^{1,1}_{0,T}, S(X)^{1,2}_{0,T}, \dots) \tag{6}$$

consisting of the k-fold iterated integral of X with multi-index $i_1, \dots, i_k$ defined as

$$S(X)^{i_1,\dots,i_k}_{0,T} = \int \cdots \int_{0 \le t_1 < \cdots < t_k \le T} \mathrm{d}X^{i_1}_{t_1} \dots \mathrm{d}X^{i_k}_{t_k}. \tag{7}$$

When the underlying path X is of bounded variation, the integral (7) can be understood as the Riemann-Stieltjes integral with respect to X.

Path signatures

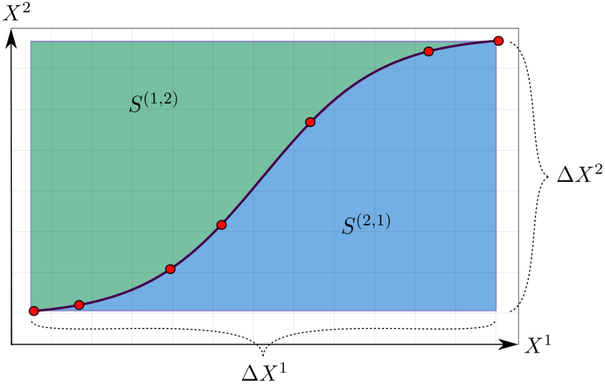

Figure: A schematic demonstrating the geometric interpretation of signature terms for an example two-dimensional path. Red circles indicate the (possibly irregular) observations, and the black curve shows the underlying continuous path. Depth-1 terms correspond to the increments ΔX^1 and ΔX^2, while the depth-2 terms $S^{(1,2)}$ and $S^{(2,1)}$ correspond to the green and blue areas, respectively.

Path signature kernel trick

- We might think of using path signature directly as a summary statistic.
- Unfortunately, it's infinite dimensional and so we can't calculate e.g. $\|\mathrm{Sig}(x) - \mathrm{Sig}(y)\|$.
- (Cass et al., 2020) demonstrate a 'kernel trick' for path signatures. Briefly, they define a *signature kernel*

$$k_{x,y}(s,t) = \langle \mathrm{Sig}(x)_s, \mathrm{Sig}(y)_t \rangle$$

  and show it to be a solution of a second order, hyperbolic PDE known as a Goursat PDE. This can be solved numerically.

Signature as summary

It follows that it is possible to use the signature without calculating it explicitly, through the loss function

$$\ell(x, y) = \|\mathrm{Sig}(x) - \mathrm{Sig}(y)\|_2^2 = k(x, x) + k(y, y) - 2\,k(x, y), \tag{8}$$

where k is the signature kernel.

Taking y as the observation, the PDE for $k(y, y)$ need only be solved once for the whole ABC algorithm, and the PDEs for $k(x, x)$ and $k(x, y)$ must be solved at each iteration.

Signature kernel ABC

- Appealing to make use of the signature in combination with the semi-automatic approach.
- It is infinite, but the signature ‘kernel trick’ is still relevant here. Simply replace standard regression techniques with kernel ridge regression.
- Nakagome, Fukumizu, and Mano, 2013 use a kernel ridge regression with a Gaussian RBF kernel to perform semi-automatic ABC.

Experiments

- We compare Wasserstein ABC (Wass), K2-ABC (MMD), semi-automatic ABC (SA), signature ABC (S) and signature kernel ABC (SK).
- SA-ABC and SK-ABC use 300 training samples from the model.
- For SA-ABC, we use the first, second, third, and fourth powers of the time series as the regressors.
- We use a simple rejection ABC scheme to ensure fairness and simplicity in comparing different loss functions.
- ABC posteriors are assessed by calculating an estimate of the MMD between ABC posteriors samples and 'ground truth' samples.
- We use 10^5 iterations of the ABC algorithm, retaining the 10^3 with lowest loss.

Results - MA(2)

Observation $x = (x_t)_{t=0,\dots,T}$ is generated as

$$x_0 = 0, \tag{9}$$

$$x_1 = \epsilon_1 + \theta_1 \epsilon_0, \tag{10}$$

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}, \quad t = 2, \dots, T, \tag{11}$$

with $\epsilon_t \sim \mathcal{N}(0, 1)$ and unknown parameters $\theta = (\theta_1, \theta_2)$.

- Take as the prior a uniform distribution on the triangle defined as $\theta_1 \in [-2, 2]$, $\theta_1 + \theta_2 > -1$, $\theta_1 - \theta_2 < 1$ and $\theta_2 < 1$.
- Generate an observation $y \sim p(\cdot|\theta^*)$, with $\theta^* = (0.6, 0.2)$

Results - MA(2)

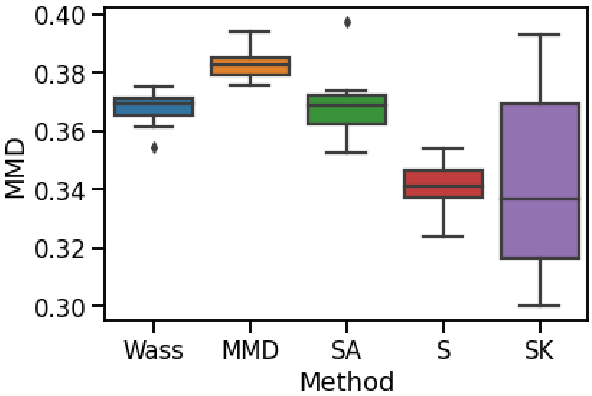

Figure: (Moving average model of order 2) Maximum mean discrepancies between the posteriors recovered from each loss function and 'ground truth' samples.

Results - GBM

Dynamics of a stock price x_t evolving with time t according to

$$\mathrm{d}x_t = \mu x_t \mathrm{d}t + \sigma x_t \mathrm{d}W_t, \tag{12}$$

where μ is the percentage drift, σ is the volatility, and W_t is a Brownian motion. This model permits an exact discretisation with $i = 1, 2, \ldots, T-1$ as

$$\log x_{i\Delta t} = \log x_{(i-1)\Delta t} + \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}\,\epsilon_i, \tag{13}$$

which we use for simulation. For all simulations, we fix $x_0 = 10$, $T = 100$, and $\Delta t = 1/(T-1)$.

We consider the task of recovering the posterior for parameters $\theta = (\mu, \sigma)$ on the basis of an observation $y = (y_0, y_{\Delta t}, y_{2\Delta t}, \ldots, y_{(T-1)\Delta t}) \sim p(x \mid \theta^*)$ with $\theta^* = (0.2, 0.5)$. We assume independent, uniform priors on the parameters as follows:

$$\mu \sim \mathcal{U}(-1, 1), \quad \sigma \sim \mathcal{U}(0.2, 2). \tag{14}$$

Results - GBM

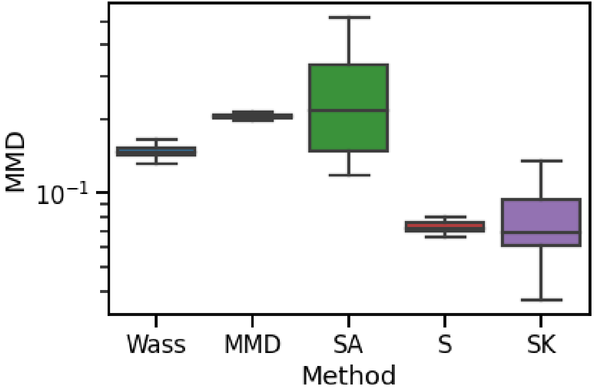

Figure: (Geometric Brownian motion) Maximum mean discrepancies between the posteriors recovered from each loss function and 'ground truth' samples.

Espen Bernton, Pierre E. Jacob, Mathieu Gerber, and Christian P. Robert. "Approximate Bayesian computation with the Wasserstein distance". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81.2 (2019), pp. 235–269.

Patric Bonnier *et al.* "Deep Signature Transforms". In: *Neural Information Processing Systems* (2019).

Thomas Cass, Terry Lyons, Cristopher Salvi, and Weixin Yang. "Computing the full signature kernel as the solution of a Goursat problem". In: *arXiv* (2020), pp. 1–39. ISSN: 23318422. arXiv: 2006.14794.

Christopher Drovandi and David T Frazier. *A Comparison of Likelihood-Free Methods With and Without Summary Statistics*. 2021. arXiv: 2103.02407 [stat.ME].

Paul Fearnhead and Dennis Prangle. "Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation". In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 74.3 (2012), pp. 419–474. ISSN: 13697412. DOI: 10.1111/j.1467-9868.2011.01010.x. arXiv: arXiv:1004.1112v2.

David T. Frazier, Christian P. Robert, and Judith Rousseau. "Model misspecification in approximate Bayesian computation: consequences and diagnostics". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.2 (2020), pp. 421–444.

James Morrill, Adeline Fermanian, Patrick Kidger, and Terry Lyons. "A Generalised Signature Method for Time Series". In: *arXiv preprint* (2020).

Shigeki Nakagome, Kenji Fukumizu, and Shuhei Mano. "Kernel approximate Bayesian computation in population genetic inferences". In: *Statistical Applications in Genetics and Molecular Biology* 12.6 (2013), pp. 667–678.

Mijung Park, Wittawat Jitkrittum, and Dino Sejdinovic. "K2-ABC: Approximate Bayesian computation with kernel embeddings". In: vol. 51. Proceedings of Machine Learning Research. PMLR, 2016, pp. 398–407.

Richard David Wilkinson. "Approximate Bayesian computation (ABC) gives exact results under the assumption of model error". In: *Statistical applications in genetics and molecular biology* 12.2 (2013), pp. 129–141.

Thanks

IMPROBABLE

WARWICK
THE UNIVERSITY OF WARWICK