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]

8 January, 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. 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 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

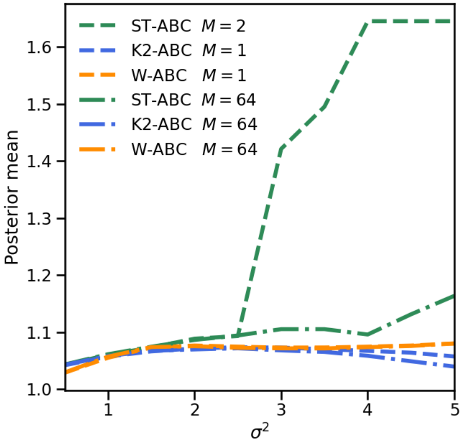

Figure: The posterior mean as a function of the degree of misspecification of the model. A comparison of ABC algorithms with a soft-threshold loss (ST-ABC), a maximum mean discrepancy (MMD) loss (K2-ABC), and a Wasserstein loss (W-ABC).

References

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.

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.

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!