ABC MCMC: a survey of theoretical results

Anthony Lee
University of Warwick & Alan Turing Institute

Validating and Expanding ABC Workshop
Banff International Research Station
February 20th, 2017

Outline

Outline

ABC pseudo-marginal Markov chains

Comparison and order

Intractable likelihood functions

- Let $y_{\mathrm{obs}} \sim f_{\mathrm{obs}}(\cdot \mid \theta_0)$ be the observed data.
  - Assume $f_{\mathrm{obs}}(y_{\mathrm{obs}} \mid \cdot)$ is intractable.
  - Assume can draw $x \sim f_{\mathrm{obs}}(\cdot \mid \theta)$ for any $\theta \in \Theta$.

- Approximation I: replace y_{obs} with $y := s(y_{\mathrm{obs}})$.
  - $f(y \mid \cdot)$ is typically also intractable.
  - We can draw $x \sim f(\cdot \mid \theta)$ for any $\theta \in \Theta$.

Intractable likelihood functions

- Approximation II: $\tilde{f}(y \mid \theta) := \int K(x,y) f(x \mid \theta) \mathrm{d}x$.
  - $\tilde{f}$ is in some sense "even less" tractable than f.
  - Standard choices include:

$$K(x,y) \propto \mathbb{I}\left\{d(x,y) \leq \epsilon\right\}, \quad K(x,y) = \mathcal{N}(y; x, \epsilon I).$$

- Alternatives exist, e.g.

$$\bar{f}(y \mid \theta) = \int f_A(y \mid \phi_N(x_{1:N}, \theta)) \prod_{i=1}^{N} f(x_i \mid \theta) \mathrm{d}x_{1:N}.$$

  - f_A is multivariate normal $\Rightarrow$ synthetic likelihood [Wood, 2010].

Why is it useful?

- Denote by p the prior density for θ.
- An auxiliary target can be defined:

$$\pi(\theta, w) \propto p(\theta)\tilde{f}(y \mid \theta)wQ_\theta(w),$$

where $W \sim Q_\theta$ is non-negative and $\mathbb{E}_{Q_\theta}[W] = 1$.

1. $\tilde{f}(y \mid \theta)W$ is a non-negative, r.v. with expectation $\tilde{f}(y \mid \theta)$
2. $\pi(\theta) = \int \pi(\theta, w)\mathrm{d}w \propto p(\theta)\tilde{f}(y \mid \theta)$.

- Rejection/importance sampling algorithms then follow.
- We can simulate a $\pi(\theta, w)$-invariant Markov chain.

ABC-MCMC pseudo-marginal kernels

- To sample from $P(\theta, w; \cdot)$:

1. Draw $\theta' \sim q(\theta, \cdot)$ and $w' \sim Q_{\theta'}$.
2. Output (θ', w') w.p.

$$1 \wedge \frac{p(\theta')\tilde{f}(y \mid \theta')w'q(\theta', \theta)}{p(\theta)\tilde{f}(y \mid \theta)wq(\theta, \theta')},$$

otherwise output (θ, w).

- Drawing $w' \sim Q_{\theta'}$ is equivalent to producing an unbiased estimate $\tilde{f}(y \mid \theta')w'$ of $\tilde{f}(y \mid \theta')$.

ABC examples of unbiased estimators

- Pseudo-marginal methods [Beaumont, 2003, Andrieu and Roberts, 2009] are generally applicable.

- Marjoram et al. [2003]:

$$w' = K(x', y)/\tilde{f}(y \mid \theta'), \qquad x' \sim f(\cdot \mid \theta').$$

- Becquet and Przeworski [2007]:

$$w' = \frac{1}{N}\sum_{i=1}^{N} K(x_i', y)/\tilde{f}(y \mid \theta'), \qquad x_i' \overset{iid}{\sim} f(\cdot \mid \theta').$$

- We denote the corresponding kernel by P_N.

- There are other possibilities, e.g. r-hit estimators [Lee, 2012]

Exact/marginal kernel $P_\star$

- We can compare this kind of chain with an "exact" variant.

- To sample from $P_\star(\theta; \cdot)$:

1. Draw $\theta' \sim q(\theta, \cdot)$
2. Output θ' w.p.

$$1 \wedge \frac{p(\theta')\tilde{f}(y \mid \theta')q(\theta', \theta)}{p(\theta)\tilde{f}(y \mid \theta)q(\theta, \theta')},$$

   otherwise output θ.

- Can think of this as P_∞, or the case where $w = 1$.

Outline

ABC pseudo-marginal Markov chains

Comparison and order

Performance measures

- To keep things simple, we will consider only

1. Asymptotic variance of ergodic averages:

$$\mathrm{var}(f, P) := \lim_{n\to\infty} n\,\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} f(\theta_i, w_i)\right),$$

   where $(\theta_0, w_0) \sim \pi$.

2. Geometric ergodicity (GE):

$$\|P^n(\theta_0, w_0; \cdot) - \pi(\cdot)\|_{TV} \leq C(x)\rho^n.$$

- Reversible P: P is geometrically ergodic $\Rightarrow$ finite asymptotic variance for all $f \in L^2(\pi)$.
- Almost necessary, for $\iff$ variance bounding instead of GE.

Comparisons with $P_\star$ 1/2

1. (B) [Andrieu and Vihola, 2015] For any $f \in L^2(\pi)$ with $f : \Theta \to \mathbb{R}$, $\mathrm{var}(f, P) \geq \mathrm{var}(f, P_\star)$.

2. (G) [Andrieu and Roberts, 2009, Andrieu and Vihola, 2015] If $W_\theta \sim Q_\theta$ is uniformly bounded in θ, then $P_\star$ GE $\Rightarrow$ P GE (at least for positive P).

3. (G) [Andrieu and Vihola, 2015] Under technical conditions on $f \in L^2(\pi)$ with $f : \Theta \to \mathbb{R}$,

$$\lim_{N \to \infty} \mathrm{var}(f, P_N) = \mathrm{var}(f, P_\star).$$

4. (B) [Andrieu and Roberts, 2009, Andrieu and Vihola, 2015] If $W_\theta \sim Q_\theta$ is unbounded for "enough" θ then P cannot be GE (not typically a problem in ABC).

Comparisons with $P_\star$ 2/2

4 (B) [Lee and Łatuszyński, 2014, Andrieu and Vihola, 2015] If $W_\theta \sim Q_\theta$ is bounded but not uniformly so, then P might not inherit GE from $P_\star$.

- For $K(x, y) = \mathbb{I}(d(x, y) \leq \epsilon)$, $\tilde{f}(y \mid \theta) > 0$ for all θ with $\tilde{f}(y \mid \theta) \to 0$ as $\|\theta\| \to \infty$ and q "local" then P_N cannot be GE for any N.

5 (G) [Deligiannidis and Lee, 2016] If $\sup_\theta \mathrm{var}(W_\theta) < \infty$ and P GE then $\mathrm{var}(f, P) < \infty$ for any $f \in L^2(\pi)$ with $f : \Theta \to \mathbb{R}$.

Ordering P's

- [Andrieu and Vihola, 2016] If $\{W_\theta; \theta \in \Theta\} \leq_{cx} \{W'_\theta; \theta \in \Theta\}$ then $\mathrm{var}(f, P) \leq \mathrm{var}(f, P')$.
  - Implies that $\mathrm{var}(f, P_N) \leq \mathrm{var}(f, P_{N+1})$ for $N \in \mathbb{N}$.
  - Also motivates stratification in ABC and dependent estimators.

- But how much better is P_{N+1} compared to P_N?
  - Improvement diminishes eventually as $\mathrm{var}(f, P_\star) \leq \mathrm{var}(f, P_N)$.

Computational considerations

- [Born et al., 2017, Sherlock et al., 2016] Let $M \leq N$. Then

$$M\left[\mathrm{var}(f, P_M) + \mathrm{var}_\pi(f)\right] \leq N\left[\mathrm{var}(f, P_N) + \mathrm{var}_\pi(f)\right],$$

which implies

$$\mathrm{var}(f, P_1) \leq N\left[\mathrm{var}(f, P_N) + \mathrm{var}_\pi(f)\right] - \mathrm{var}_\pi(f),$$

i.e. simple averaging cannot bring "too much" benefit.

- P_N positive implies $\mathrm{var}(f, P_1) \leq (2N-1)\mathrm{var}(f, P_N)$.
- Also shows that $\mathrm{var}(f, P_N) < \infty \iff \mathrm{var}(f, P_1) < \infty$.
- If comp. cost is proportional to N, often best to use $N = 1$.

Discussion 1/2

- There exist provably more robust Markov chains, e.g. 1-hit ABC [Lee et al., 2012, Lee and Łatuszyński, 2014], *r*-hit variants [Lee, 2012], correlated pseudo-marginal methods [Deligiannidis et al., 2015].
- Understanding is still incomplete.

- Other Monte Carlo methods, e.g. SMC samplers / PMC.
- Choice of summary statistics.
- How to exploit mappings $F^{-}(U) = X \sim f(\cdot \mid \theta)$ where $U \sim \mathcal{U}([0,1]^d)$.

Discussion 2/2

- There are potential benefits to alternative approximate likelihoods. E.g., in a *very* simple scenario [Price et al., 2017],

$$\bar{f}_N(y \mid \theta) = \int f_A(y \mid \phi_N(x_{1:N}, \theta)) \prod_{i=1}^{N} f(x_i \mid \theta) \mathrm{d}x_{1:N},$$

  is comp. more robust than $\tilde{f}(y \mid \theta) = \int K_\epsilon(x, y) f(x \mid \theta) \mathrm{d}x$.

- N acts like $1/\epsilon$, controls some approximation error.
- Natural estimator of $\bar{f}_N(y \mid \theta)$ converges in prob. as $N \to \infty$ with cost $\mathcal{O}(N)$, but for a given dimension d one needs $\mathcal{O}(N^{d/2})$ samples to stabilize the natural estimator of $\tilde{f}(y \mid \theta)$.
- Of course, in general $\bar{f}_N(y \mid \theta) \not\to f(y \mid \theta)$ as $N \to \infty$.

References I

C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009.

C. Andrieu and M. Vihola. Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.*, 25(2):1030–1077, 2015.

C. Andrieu and M. Vihola. Establishing some order amongst exact approximations of MCMCs. *Ann. Appl. Probab.*, 26(5):2661–2696, 2016.

M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164:1139–1160, 2003.

C. Becquet and M. Przeworski. A new approach to estimate parameters of speciation models with application to apes. *Genome research*, 17(10): 1505–1519, 2007.

L. Bornn, N. S. Pillai, A. Smith, and D. Woodard. The use of a single pseudo-sample in approximate Bayesian computation. *Statistics and Computing*, 2017. To appear.

G. Deligiannidis and A. Lee. Which ergodic averages have finite asymptotic variance? *arXiv preprint arXiv:1606.08373*, 2016.

G. Deligiannidis, A. Doucet, and M. K. Pitt. The correlated pseudo-marginal method. *arXiv preprint arXiv:1511.04992*, 2015.

References II

A. Lee. On the choice of MCMC kernels for approximate Bayesian computation with SMC samplers. In *Proceedings of the Winter Simulation Conference*, 2012.

A. Lee and K. Łatuszyński. Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika*, 101(3):655–671, 2014.

A. Lee, C. Andrieu, and A. Doucet. Discussion of "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation" by Fearnhead and Prangle. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 74(3):449–450, 2012.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100 (26):15324–15328, 2003.

L. F. Price, C. C. Drovandi, A. Lee, and D. J. Nott. Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 2017. To appear.

C. Sherlock, A. Thiery, and A. Lee. Pseudo-marginal Metropolis–Hastings using averages of unbiased estimators. *arXiv preprint arXiv:1610.09788*, 2016.

S. N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102–1104, 2010.