Improving approximate Bayesian computation through randomized quasi Monte Carlo

Rencontre des jeunes statisticiens 2017 – SFdS

Alexander Buchholz, joint work with Nicolas Chopin

April 6, 2017

CREST – ENSAE Paris Tech

Table of contents

Approximate Bayesian computation and Sequential Monte Carlo

Concept of ABC

- Suppose we have a model that is amenable to simulation but without a closed form for the likelihood
- How can we recover the most likely set of parameters for the model that led to the observed data?
- **Examples**: biological models (mutation of bacteria), epidemiology (propagation of diseases)
- **Aim**: Making Bayesian inference in a model with intractable likelihood $L(y|\theta)$:

$$p(\theta|y) \propto L(y|\theta)p(\theta), \tag{1}$$

  but simulation of $y|\theta$ is possible.
- **Solution**: Introduce distance $\delta(\cdot, \cdot)$ between simulation y and observed y^*, keep y that are close enough for a given θ
- See [Marin et al., 2012] for a review

Concept of ABC

Algorithm [Beaumont et al., 2002]:
For $i \leq N$:

1. Simulate $\theta_i \sim p(\theta)$ from the prior
2. Simulate $y_i \sim L(y|\theta_i)$ from the model
3. Accept θ_i, y_i with probability $\mathcal{K}_\epsilon(\delta(y_i, y^*))$

Result: $\theta_i, y_i | y^*, \epsilon \sim L(y|\theta)p(\theta)\mathcal{K}_\epsilon(\delta(y, y^*))$,
correct sampling if $\epsilon \to 0$,

Use simulations to approximate $\mathbb{E}_{p(\theta, y|y^*, \epsilon)}[\Phi(\theta, y)] \approx \frac{1}{N}\sum_i \Phi(\theta_i, y_i)$

Sequential Monte Carlo (SMC) and ABC

Sequential setup: Use previous steps to guide the simulation towards regions of high posterior probability, reduce ϵ_t at every iteration, see [Sisson et al., 2007, Beaumont et al., 2009]

Importance sampling: we want to approximate the integral $\mathbb{E}_{p(\theta,y|y^*,\epsilon_t)}[\Phi(\theta,y)]$

- **Target**: $L(y|\theta)p_0(\theta)\mathcal{K}_{\epsilon_t}(\delta(y,y^*))$, **proposal**: $L(y|\theta)p_t(\theta)$
- **Weight**: $\hat{\mathcal{K}}_{\epsilon_t}(\delta(y,y^*)) = \frac{p_0(\theta)\mathcal{K}_{\epsilon_t}(\delta(y,y^*))}{p_t(\theta)}$
- Auto-normalized estimator:

$$\frac{\frac{1}{N}\sum_{i=1}^{N}\Phi(\theta_i,y_i)\hat{\mathcal{K}}_{\epsilon_t}(\delta(y_i,y^*))}{\frac{1}{N}\sum_{i=1}^{N}\hat{\mathcal{K}}_{\epsilon_t}(\delta(y_i,y^*))} \to_N \mathbb{E}_{p(\theta,y|y^*,\epsilon_t)}[\Phi(\theta,y)] \qquad (2)$$

- Use a sequence of $\epsilon_0 > \cdots > \epsilon_t$ and an adaptable proposal $p_t(\theta)$

Sequential Monte Carlo (SMC) and ABC

Algorithm [Sisson et al., 2007, Beaumont et al., 2009]:

For time $t \leq T$:

For particle $i \leq N$:

1. Simulate $\theta_i^t \sim p_t(\theta)$ from the proposal
2. Simulate $y_i^t \sim L(y|\theta_i^t)$ from the model
3. Accept θ_i^t, y_i^t with probability $\hat{\mathcal{K}}_{\epsilon_t}(\delta(y_i^t, y^*))$
4. Construct $p_{t+1}(\theta)$ based on accepted θ_i^t

Return $\frac{\frac{1}{N}\sum_{i=1}^N \Phi(\theta_i^T, y_i^T)\hat{\mathcal{K}}_{\epsilon_T}(\delta(y_i^T, y^*))}{\frac{1}{N}\sum_{i=1}^N \hat{\mathcal{K}}_{\epsilon_T}(\delta(y_i^T, y^*))}$

Quasi Monte Carlo

Monte Carlo and Quasi Monte Carlo point sets

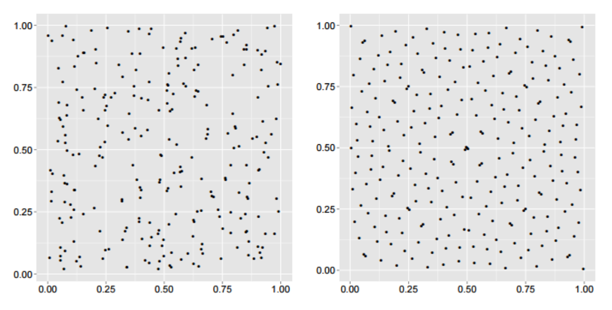

Figure 1: 256 uniform Monte Carlo and quasi Monte Carlo points on $[0,1)^2$

- **Aim of MC and QMC**: solve $I = \int_{[0,1)^s} \Phi(x)dx$
- Therefore, sample $\{x_i\}_{i=1:N} \in [0,1)^s$
- Evaluate $\hat{I}_N = \frac{1}{N}\sum_i^N \Phi(x_i)$
- What about the error of the approximation?

Monte Carlo and Quasi Monte Carlo point sets

Error rates in MC, QMC and RQMC, see [Leobacher and Pillichshammer, 2014, Owen, 1997]

- Convergence of $|\hat{I}_N - I|^2 = \mathcal{E}(\Phi, N)$
- For Monte Carlo: $\mathbb{E}[\mathcal{E}(\Phi, N^{MC})] = \mathcal{O}(N^{-1})$
- For quasi Monte Carlo: $\mathcal{E}(\Phi, N^{QMC})] = \mathcal{O}(N^{-2+\tau})$
- For randomized quasi Monte Carlo: $\mathbb{E}[\mathcal{E}(\Phi, N^{RQMC})] = \mathcal{O}(N^{-3+\tau})$
- *Note:* $\tau = \tau(s)$ grows with dimension s.

Monte Carlo and Quasi Monte Carlo point sets

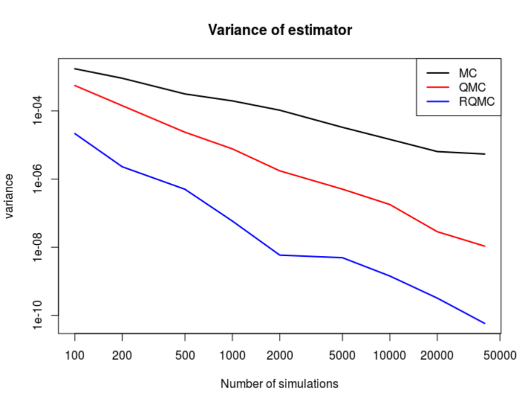

Figure 2: Convergence rate of the error of Monte Carlo, quasi Monte Carlo and randomized quasi Monte Carlo

- **Idea**: Use RQMC for the proposal distribution in ABC

Theoretical results

Asymptotic and non-asymptotic behavior of mixed sequences

Convergence results for mixed sequences

- We investigate the behavior of ABC estimators with mixed sequences with fixed ϵ
- $\theta \sim p(\theta)$, where we will use a (R)QMC point set
- $y \sim p(y|\theta)$, where the simulator generates a MC sample
- $\theta, y \sim p(y|\theta)p(\theta)$ is a mixed sequence
- Study $\text{Var}(\hat{Z}_N)$ and $\text{Var}(\hat{\Phi}_N)$

$$\hat{Z}_N = \frac{1}{N}\sum_{i=1}^{N} \mathcal{K}_\epsilon(\delta(y_i, y^*)) \tag{3}$$

$$\hat{\Phi}_N = \frac{\frac{1}{N}\sum_{i=1}^{N} \Phi(\theta_i, y_i)\mathcal{K}_\epsilon(\delta(y_i, y^*))}{\frac{1}{N}\sum_{i=1}^{N} \mathcal{K}_\epsilon(\delta(y_i, y^*))} \tag{4}$$

Non-Asymptotic behavior of mixed sequences

- Let ϵ be fixed. Consider the normalization constant of the posterior distribution $Z = \int \mathcal{K}_\epsilon(\delta(y, y^*))p(y|\theta)p(\theta)dyd\theta = p_\epsilon(y^*)$
- We approximate the constant by the mixed random sample $(\theta_i, y_i)_{i\in\{1,\ldots,N\}}$ and calculation of $\hat{Z}_N = \frac{1}{N}\sum_{i=1}^N \mathcal{K}_\epsilon(\delta(y_i, y^*))$

Theorem

$\text{Var}(\hat{Z}_N) \leq \mathcal{O}(N^{-3+\tau}) + \mathcal{O}(N^{-1})$

Proof.

Use decomposition of variance to identify the corresponding terms and use [Owen, 1997] (Hoelder continuity of conditional expectation).

$$\text{Var}(\hat{Z}_N) = \underbrace{\text{Var}_{\theta^{1:N}}[\mathbb{E}[\hat{Z}_N|\theta^{1:N}]]}_{\sim\mathcal{O}(N^{-3+\tau})} + \underbrace{\mathbb{E}_{\theta^{1:N}}[\text{Var}[\hat{Z}_N|\theta^{1:N}]]}_{\sim\mathcal{O}(N^{-1})} \tag{5}$$

□

Non-Asymptotic behavior of mixed sequences

- Consider general importance sampling estimators of the following form

$$\hat{\Phi}_N = \frac{\frac{1}{N}\sum_{i=1}^{N} \Phi(\theta_i, y_i)\mathcal{K}_\epsilon(\delta(y_i, y^*))}{\frac{1}{N}\sum_{i=1}^{N} \mathcal{K}_\epsilon(\delta(y_i, y^*))} \tag{6}$$

Theorem

$\mathrm{Var}(\hat{\Phi}_N) \leq \mathcal{O}(N^{-3+\tau}) + \mathcal{O}(N^{-1})$

Proof.

Upper bound the variance of the importance sampling estimator by the variance of the denominator and the numerator as in [Agapiou et al., 2015]. Apply the same reasoning as before. □

Asymptotic behavior of mixed sequences

- Can we get an asymptotic variance reduction?
- Yes! Multivariate extension of CLT by [Ökten et al., 2006]

Theorem

Let $u^k = (q^k_{1:d}, X^k_{d+1:s})$ *be a mixed sequence of dimension* s *where* $q^k_{1:d}$ *is the QMC part and* $X^k_{d+1:s}$ *is the random MC part. Let* $I = \int_{[0,1)^s} f(x)dx$ *be the target,* $\hat{I}_N = \frac{1}{N}\sum_k^N f(u^k)$ *the estimator. If* $\lim_{N\to\infty} \frac{\mathrm{Var}[\sum_k^N f(u^k)]}{N} = C^2_{QMC}$ *is finite, invertible and not dominated by any summand, then*

$$\sqrt{N}\left(\hat{I}_N - I\right) \to^{\mathcal{L}} \mathcal{N}(0, C_{QMC}) \tag{7}$$

and $C_{QMC} \preceq C_{MC}$

Conclusion: Using mixed sequences we can achieve an asymptotic variance reduction. Results hold when using RQMC instead of QMC.

Application

Toy example: recovering the mean of a Gaussian mixture

$y^* = 0,\ y \sim \frac{1}{2}\mathcal{N}(\theta, 1) + \frac{1}{2}\mathcal{N}(\theta, 0.01)$ and $\theta \sim \pi(\cdot)$

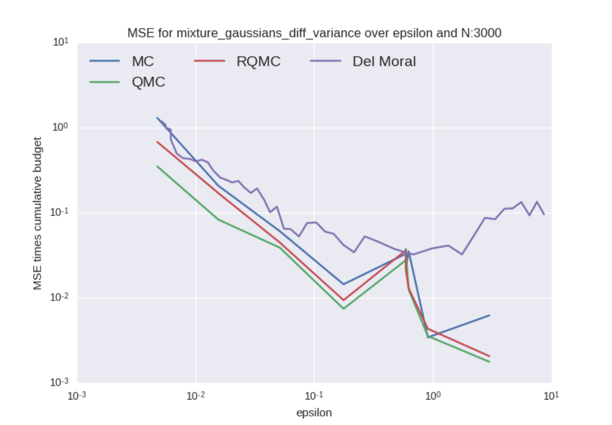

Figure 3: Adaptive importance sampling for the ABC target. Comparison of cost×variance to the SMC-ABC sampler of [Del Moral et al., 2012].

Inference in a birth-death-mutation process in epidemiology

Results for variance reduction

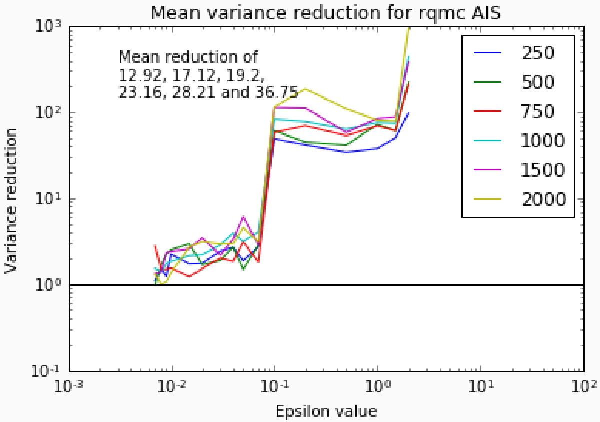

Figure 4: Adaptive importance sampling for the ABC target. Ratio of variances of MC-ABC vs. RQMC-ABC for a contracting epsilon schedule.

Conclusion

Summary

- Replacing the ABC prior by a (R)QMC point set yields direct variance reduction
- No change in the algorithm necessary $\Rightarrow$ **free lunch!**
- Theoretical results consistent with application
- **But:** The rate of convergence stays the same
- **But:** The variance reduction vanishes if ϵ goes to zero
- **Ongoing work:** Apply to more complex models, study properties in sequential setup

Thank you for your attention!

Questions?

References I

Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., and Stuart, A. (2015).
Importance sampling: computational complexity and intrinsic dimension.
arXiv preprint arXiv:1511.06196.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009).
Adaptive approximate bayesian computation.
Biometrika, page asp052.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002).
Approximate bayesian computation in population genetics.
Genetics, 162(4):2025–2035.

References II

Blum, M. G. (2012).
Approximate bayesian computation: a nonparametric perspective.
Journal of the American Statistical Association.

Chopin, N. (2004).
Central limit theorem for sequential monte carlo methods and its application to bayesian inference.
Annals of statistics, pages 2385–2411.

Del Moral, P., Doucet, A., and Jasra, A. (2012).
An adaptive sequential monte carlo method for approximate bayesian computation.
Statistics and Computing, 22(5):1009–1020.

References III

Gerber, M. and Chopin, N. (2015).
Sequential quasi monte carlo.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(3):509–579.

Leobacher, G. and Pillichshammer, F. (2014).
Introduction to quasi-Monte Carlo integration and applications.
Springer.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012).
Approximate bayesian computational methods.
Statistics and Computing, 22(6):1167–1180.

References IV

Ökten, G., Tuffin, B., and Burago, V. (2006).
A central limit theorem and improved error bounds for a hybrid-monte carlo sequence with applications in computational finance.
Journal of Complexity, 22(4):435–458.

Owen, A. B. (1997).
Scrambled net variance for integrals of smooth functions.
The Annals of Statistics, 25(4):1541–1562.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007).
Sequential monte carlo without likelihoods.
Proceedings of the National Academy of Sciences, 104(6):1760–1765.

References V

Tanaka, M. M., Francis, A. R., Luciani, F., and Sisson, S. (2006).
Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data.
Genetics, 173(3):1511–1520.