

Tutorial on ABC Algorithms

Dr Chris Drovandi
Queensland University of Technology, Australia
c.drovandi@qut.edu.au

July 3, 2014



Notation

- Model parameter θ with prior $\pi(\theta)$
- Likelihood is $f(\mathbf{y}|\theta)$ with observed data \mathbf{y}
- Posterior $\pi(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta)\pi(\theta)$
- Simulated data \mathbf{x}
- ABC posterior $\pi_\epsilon(\theta, \mathbf{x}|\mathbf{y}) \propto K_\epsilon(\rho(\mathbf{y}, \mathbf{x}))f(\mathbf{x}|\theta)\pi(\theta)$

Ingredients for Implementing ABC Algorithm

- Code to simulate data from model
- Code to compute summary statistics
- Code to compute discrepancy function
- Combine within rejection, MCMC or SMC algorithm (not too hard)

Importance Sampling

- Define importance distribution $g(\boldsymbol{\theta})$
- Draw M iid samples from g , $\boldsymbol{\theta}_i \stackrel{iid}{\sim} g(\boldsymbol{\theta})$, $i = 1, \dots, M$
- Weight the samples

$$w_i = \frac{f(\mathbf{y}|\boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i)}{g(\boldsymbol{\theta}_i)}$$

- Normalise the weights to obtain W_i , $i = 1, \dots, M$
- $\{\boldsymbol{\theta}_i, W_i\}_{i=1}^M$ represents weighted sample from $\pi(\boldsymbol{\theta}|\mathbf{y})$
- Effective sample size:

$$ESS = \frac{1}{\sum_{i=1}^M (W_i)^2}$$

ABC Importance Sampling

- Require proposal on joint space of θ and \mathbf{x}
- Let $g(\theta, \mathbf{x}) = g(\theta)f(\mathbf{x}|\theta)$
- Generate sample $\{\theta_i, \mathbf{x}_i\}_{i=1}^M \stackrel{iid}{\sim} g(\theta, \mathbf{x})$
- Weight

$$w_i = \frac{K_\epsilon(\rho(\mathbf{y}, \mathbf{x}_i)) \cancel{f(\mathbf{x}_i|\theta_i)} \pi(\theta_i)}{g(\theta_i) \cancel{f(\mathbf{x}_i|\theta_i)}}$$

- If we choose $g(\theta) = \pi(\theta)$ and let $K_\epsilon(\rho(\mathbf{y}, \mathbf{x})) = 1(\rho(\mathbf{y}, \mathbf{x}) \leq \epsilon)$ then w_i is either 0 or 1

ABC Rejection Algorithm

ABC rejection (post-determination of ϵ)

- 1: Draw $\theta_i \sim \pi(\cdot)$ and simulate $\mathbf{x}_i \sim f(\cdot|\theta_i)$ for $i = 1, \dots, M$.
Generates collection $\{\theta_i, \mathbf{x}_i\}_{i=1}^M$.
- 2: Compute discrepancy $\rho_i = \rho(\mathbf{y}, \mathbf{x}_i)$. Obtain particles $\{\theta_i, \rho_i\}_{i=1}^M$.
- 3: Sort $\{\theta_i, \rho_i\}_{i=1}^M$ based on ρ
- 4: Keep $N = \alpha \times M$ of θ_i with lowest discrepancy (this defines the ϵ)

Choice of α trade off between accuracy and Monte Carlo error
(Fearnhead and Prangle (2012))

ABC Rejection Algorithm (Pritchard et al 1999)

ABC Rejection (pre-specification of ϵ)

- 1: Draw $\theta \sim \pi(\cdot)$
 - 2: Simulate $\mathbf{x} \sim f(\cdot|\theta)$
 - 3: If $\rho(\mathbf{y}, \mathbf{x}) \leq \epsilon$ then accept θ
 - 4: Repeat lines 1:, 2: and 3: until N samples are drawn
- Avoids storage requirements but choosing ϵ difficult

ABC Rejection Example

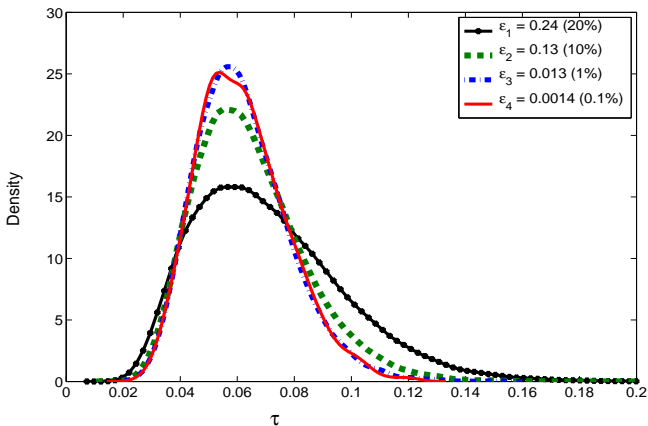


Figure: Plot courtesy of Brenda Nho Vo. ABC posterior for different α (leading to different ϵ).

ABC Rejection Algorithm

Advantages

- Simplicity
- Useful for testing different sets of summary statistics (Nunes and Balding 2010) or analysing multiple datasets from same model (see Drovandi and Pettitt 2013 for Bayesian experimental design example)

Disadvantages

- Storage requirements (depending on implementation chosen)
- Highly inefficient if posterior different to prior (too high ϵ required to obtain a reasonable size sample from ABC posterior)

Markov chain Monte Carlo

- Motivation: Keep proposals within non-negligible posterior regions
- Construct a Markov chain whose limiting distribution is $\pi(\boldsymbol{\theta}|\mathbf{y})$
- Assume current value of chain is $\boldsymbol{\theta}$. Propose next value of chain $\boldsymbol{\theta}^* \sim q(\cdot|\boldsymbol{\theta})$. Accept $\boldsymbol{\theta}^*$ as next value of chain with probability $\min(1, A)$ where

$$A = \frac{f(\mathbf{y}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})},$$

otherwise set $\boldsymbol{\theta}$ as the next value of chain.

MCMC ABC

- Require proposal on joint space of θ and x
- Let $q(\theta^*, x^* | \theta, x) = f(x^* | \theta^*)q(\theta^* | \theta)$
- Metropolis-Hastings Ratio

$$A = \frac{K_\epsilon(\rho(\mathbf{y}, \mathbf{x}^*)) \cancel{f(\mathbf{x}^* | \theta^*)} \pi(\theta^*) q(\theta | \theta^*) \cancel{f(\mathbf{x} | \theta)}}{K_\epsilon(\rho(\mathbf{y}, \mathbf{x})) \cancel{f(\mathbf{x} | \theta)} \pi(\theta) q(\theta^* | \theta) \cancel{f(\mathbf{x}^* | \theta^*)}}$$

- Choice of proposal leads to cancellation of intractable likelihoods!

MCMC ABC (Marjoram et al 2003, Sisson and Fan 2011)

- 1: Obtain θ^0, \mathbf{x}^0 using a burnin or from one sample of rejection ABC
- 2: Compute $\psi^0 = K_\epsilon(\rho(\mathbf{y}, \mathbf{x}^0))$
- 3: **for** $i = 1$ **to** N **do**
- 4: Draw $\theta^* \sim q(\cdot | \theta^{i-1})$
- 5: Simulate $\mathbf{x}^* \sim f(\cdot | \theta^*)$
- 6: Compute $\psi^* = K_\epsilon(\rho(\mathbf{y}, \mathbf{x}^*))$
- 7: Compute $r = \frac{\pi(\theta^*)\psi^*q(\theta^{i-1}|\theta^*)}{\pi(\theta^{i-1})\psi^{i-1}q(\theta^*|\theta^{i-1})}$
- 8: **if** $\text{uniform}(0, 1) < r$ **then**
- 9: $\theta^i = \theta^*, \psi^i = \psi^*$
- 10: **else**
- 11: $\theta^i = \theta^{i-1}, \psi^i = \psi^{i-1}$
- 12: **end if**
- 13: **end for**

MCMC ABC

Advantages

- Still pretty simple
- Tends to be more efficient than ABC rejection for single data analysis

Disadvantages

- Slightly more difficult to implement than ABC rejection
- Requires tuning of proposal distribution q
- Can get 'stuck' in low probability regions (thus long runs are often required)
- Must be re-run and tuned for each new dataset or summary statistic selection

MCMC ABC Example of “stickiness”

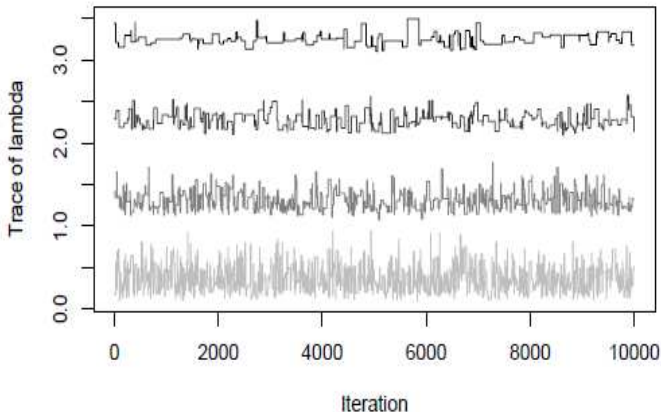


Figure: Plot taken from Sisson and Fan 2011. From bottom to top ϵ is 4.5, 4, 3.5, 3.

Variants on MCMC ABC

- Bortot et al 2007 - Include ϵ as a 'parameter' (every now and then propose larger value of ϵ)
- Baragatti et al 2013 - Population MCMC (parallel tempering, swapping chains with different ϵ)
- Picchini et al 2013 - Early rejection (when using uniform kernel, may reject proposal without simulating data)
- Aandahl et al 2014 - Multiple try MCMC ABC

Sequential Monte Carlo (SMC) ABC

- SMC for sampling from a sequence of target distribution
- For ABC natural to define sequence in terms of non-increasing set of tolerances $\epsilon_1 \geq \epsilon_2 \geq \epsilon_T$

$$\pi_t(\boldsymbol{\theta}, \mathbf{x} | \mathbf{y}, \epsilon) \propto f(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathbf{1}(\rho(\mathbf{x}, \mathbf{y}) \leq \epsilon_t), \text{ for } t = 1, \dots, T,$$

- Traverse set of N 'particles' through sequence of targets by iteratively applying re-weighting (importance sampling), re-sampling and mutation steps
- Ultimately obtain set of weighted samples from π_T
- Nice accessible reference to SMC is Chopin (2002) (see also Del Moral et al (2006))

SMC ABC Algorithm of a few papers

- 1: Initialise $\epsilon_1, \dots, \epsilon_T$ and specify the initial importance sampling distribution $\pi_0(\cdot)$
- 2: **for** $t = 1$ **to** T **do**
- 3: **for** $i = 1$ **to** N **do**
- 4: If $t = 1$ sample θ^{**} from $\pi_0(\cdot)$. If $t > 1$ sample θ^* from the previous population $\{\theta_i^{t-1}, W_i^{t-1}\}_{i=1}^N$ and perturb the particle $\theta^{**} \sim K_t(\cdot | \theta^*)$.
- 5: Generate a dataset $x^{**} \sim f(\cdot | \theta^{**})$.
- 6: If $\rho(y, x^{**}) > \epsilon_t$ then go back to step 4.
- 7: Set $\theta_t^i = \theta^{**}$ and re-weight

$$w_t^i = \begin{cases} \frac{\pi(\theta_t^i)}{\pi_0(\theta_t^i)} & \text{if } t = 1 \\ \frac{\pi(\theta_t^i)}{\sum_{j=1}^N W_{t-1}^j K_t(\theta_t^i | \theta_{t-1}^j)} & \text{if } t > 1 \end{cases} .$$

- 8: **end for**
- 9: Normalise the weights $W_t^i = w_t^i / \sum_{j=1}^N w_t^j$ for $i = 1, \dots, N$.
- 10: Update the tuning parameters of K_{t+1} using the set of particles $\{\theta_t^i, W_t^i\}_{i=1}^N$.
- 11: **end for**

SMC ABC

Advantages

- Advantages of SMC over MCMC (deals better with multi-modal posteriors, easy adaptation of proposal)
- Tends to be more efficient than ABC rejection and MCMC ABC (from adaptive proposal)

Disadvantages

- Not as simple to implement as ABC rejection and MCMC ABC
- Must be re-run and tuned for each new dataset or summary statistic selection
- Requires specification of sequence of tolerances

SMC ABC of Drovandi et al 2011

- 1: Set N_a as the integer part of αN
- 2: Perform the rejection sampling algorithm with ϵ_1 . This produces a set of particles $\{\theta^i, \rho^i\}_{i=1}^N$
- 3: Sort the particle set by ρ , so that $\rho^1 \leq \rho^2 \leq \dots \leq \rho^N$, and set $\epsilon_t = \rho^{N-N_a}$ and $\epsilon_{\max} = \rho^N$. If $\epsilon_{\max} \leq \epsilon_T$ then finish, otherwise go to 4.
- 4: Compute the tuning parameters of the MCMC kernel $q_t(\cdot|\cdot)$ using the particle set $\{\theta^i\}_{i=1}^{N-N_a}$
- 5: **for** $j = N - N_a + 1$ **to** N **do**
- 6: Resample θ^j from $\{\theta^i\}_{i=1}^{N-N_a}$
- 7: Perform R_t iterations of MCMC ABC using q_t and ϵ_t to update θ^j
- 8: **end for**
- 9: Compute R_t based on the overall MCMC acceptance rate of the previous iteration and go to 3.

See also Del Moral et al 2012 for similar algorithm

Movie

It's movie time!



SMC ABC of Drovandi et al 2011

Advantages

- Same advantages as previous but also do not require to specify sequence of tolerances, removes weight calculation and has stopping rule

Disadvantages

- Same disadvantages as previous but do not require sequence of tolerances but uses MCMC ABC so not all particles get moved (some duplication inevitable)

Other Variants on SMC ABC

- Silk et al (2013) - different way to determine adaptive sequence of tolerances
- Filippi et al (2013) - determine optimal proposal distribution for SMC ABC
- Probably a few other papers out there...

ABC Software Packages

- R Packages (ABC, EasyABC)
- DIYABC, PopABC (ABC for population genetics models)
- ABC-SysBio (Python package with GPU support(?))
- ABCtoolbox

See Wiki page

http://en.wikipedia.org/wiki/Approximate_Bayesian_computation

Closing Remarks on ABC Algorithms

- ABC Algorithms generally easy to implement
- MCMC and SMC versions tend to be more efficient but ABC rejection still has its charm
- Generally still quite computationally intensive
- Regression adjustment (see Beaumont et al 2002) can be applied to output of any ABC algorithm
- Many other Likelihood-free algorithms out there...

References

[Beaumont et al \(2002\)](#). Approximate Bayesian Computation in Population Genetics. Genetics.

[Marjoram et al \(2003\)](#). Markov chain Monte Carlo without likelihoods. PNAS.

[Sisson et al \(2007\)](#). Sequential Monte Carlo without likelihoods. PNAS.

[Drovandi and Pettitt \(2011\)](#). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics.

[Del Moral et al \(2012\)](#). An adaptive sequential Monte Carlo method for approximate Bayesian computation. Statistics and Computing.

[Drovandi and Pettitt \(2013\)](#). Bayesian Experimental Design for Models with Intractable Likelihoods. Biometrics.