Tutorial on ABC rejection and ABC SMC for parameter estimation and model selection

Tina Toni, Michael P. H. Stumpf

ttoni@imperial.ac.uk, m.stumpf@imperial.ac.uk

In this tutorial we schematically illustrate four algorithms:

1. ABC rejection for parameter estimation [1, 2],

2. ABC SMC for parameter estimation [3, 4, 5],

3. ABC rejection for model selection on the joint space [6],

4. ABC SMC for model selection on the joint space [7].

We suggest to read this tutorial from the beginning. We start with a detailed explanation of the ABC rejection algorithm, which later helps to understand ABC SMC as it is based on the same concepts. Also, both model selection algorithms are closely related to parameter estimation algorithms and it is therefore helpful to understand those first.

This tutorial forms a part of the supplementary material of the paper "Simulation-based model selection for dynamical systems in systems and population biology, Bioinformatics, 26 (1), 104-110, 2010" (T. Toni, M. P. H. Stumpf).

ABC rejection

(a) We define a prior distribution $P(\theta)$ and we would like to approximate the posterior distribution $P(\theta|D_0)$. We start by sampling a parameter θ^* from the prior distribution. We call this sampled parameter *a particle*.

(b) We simulate a data set D^* according to some simulation framework $f(D|\theta^*)$. In our examples we use different simulation frameworks. If we simulate a deterministic dynamical model, we add some noise at the time points of interest. If we simulate a stochastic dynamical model, we do not add any additional noise to the trajectories. We compare the simulated data set D^* (circles) to the experimental data D_0 (crosses) using a distance function, d, and tolerance ϵ; if $d(D_0, D^*) \leq \epsilon$, we accept θ^*. The tolerance $\epsilon \geq 0$ is the desired level of agreement between D_0 and D^*.

(c) The particle θ^* is accepted because D^* and D_0 are sufficiently close.

(d) We sample another parameter θ^* from the prior distribution and simulate a corresponding dataset D^*. In this case D^* and D_0 are very different and we reject the particle (we "throw it away").

(e) We repeat the whole procedure until N particles have been accepted. They represent a sample from $P(\theta|d(D_0, D^*) \leq \epsilon)$, which approximates the posterior distribution. If ϵ is sufficiently small then the distribution $P(\theta|d(D_0, D^*) \leq \epsilon)$ will be a good approximation for the "true" posterior distribution, $P(\theta|D_0)$.

(f) Many particles were rejected in the procedure, for which we have spent a lot of computational effort for simulation. ABC rejection is therefore computationally inefficient. We can use ABC SMC to reduce the computational cost.
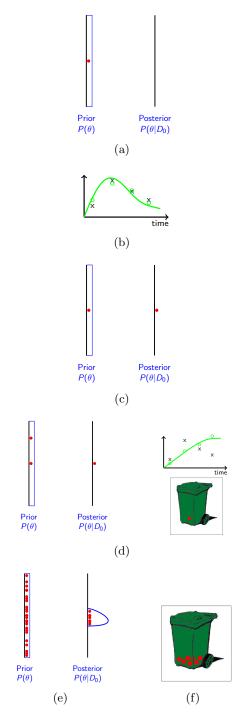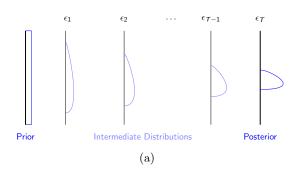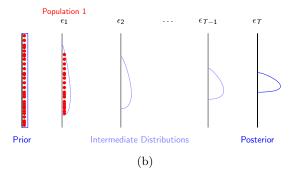

Figure 1: Schematic representation of ABC rejection.
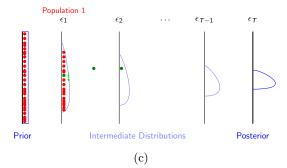
ABC SMC (Toni *et al.*, 2009)

(a) As in ABC rejection, we define a prior distribution $P(\theta)$ and we would like to approximate a posterior distribution $P(\theta|D_0)$. In ABC SMC we do this sequentially by constructing intermediate distributions, which converge to the posterior distribution. We define a tolerance schedule $\epsilon_1 > \epsilon_2 > \ldots \epsilon_T \geq 0$.

(a)

(b) We sample particles from a prior distribution until N particles have been accepted (have reached the distance smaller than ϵ_1). For all accepted particles we calculate weights (see [4] for formulas and derivation). We call the sample of all accepted particles "Population 1".

(b)

(c) We then sample a particle θ^* from population 1 and perturb it to obtain a perturbed particle $\theta^{**} \sim K(\theta|\theta^*)$, where K is a perturbation kernel (for example a Gaussian random walk). We then simulate a dataset $D^* \sim f(D|\theta^{**})$ and accept the particle θ^{**} if $d(D_0, D^{**}) \leq \epsilon_2$. We repeat this until we have accepted N particles in population 2. We calculate weights for all accepted particles.

(c)

(d) We repeat the same procedure for the following populations, until we have accepted N particles of the last population T and calculated their weights. Population T is a sample of particles that approximates the posterior distribution.

ABC SMC is computationally much more efficient than ABC rejection (see [4] for comparison).
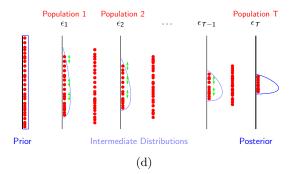
(d)

Figure 2: Schematic representation of ABC SMC.

3

ABC rejection for model selection

Prior
$P(m, \theta)$

Posterior
$P(m, \theta|D_0)$

(a)

A sample from the prior distribution (white) and accepted particles (blue)

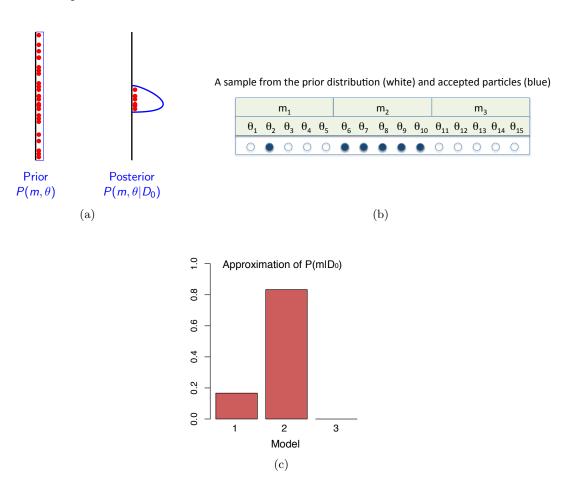| m_1					m_2					m_3				
θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}

(b)

Approximation of $P(m|D_0)$

Model

(c)

Figure 3: (a) Prior and posterior distributions, $P(m, \theta)$ and $P(m, \theta|D_0)$, are now defined on a joint model and parameter space. (b) Particles (m, θ) are sampled from the prior distribution and accepted/rejected according to the distance between the simulated and experimental datasets. The accepted particles are shown in dark blue. (c) Six particles have been accepted: one from model 1 and five from model 2. The approximated marginal posterior probability of the model can be calculated as $P(m = 1|D_0) = \frac{1}{6}$, $P(m = 2|D_0) = \frac{5}{6}$, $P(m = 3|D_0) = \frac{0}{6}$. For illustrative purposes we have chosen a small number of particles. In principle this algorithm will yield consistent marginal posterior model distributions for $N \to \infty$.
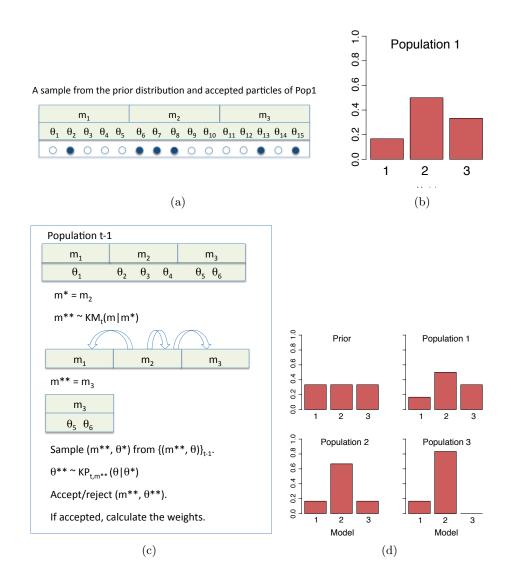
4

ABC SMC for model selection

A sample from the prior distribution and accepted particles of Pop1

m_1					m_2					m_3				
θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}
○	●	○	○	○	●	●	●	●	○	○	○	○	●	○	●

(a)

(b)

Population t-1

m_1	m_2	m_3
θ_1	θ_2 θ_3 θ_4	θ_5 θ_6

$m^* = m_2$

$m^{**} \sim KM_t(m|m^*)$

m_1	m_2	m_3

$m^{**} = m_3$

m_3
θ_5 θ_6

Sample (m^{**}, θ^*) from $\{(m^{**}, \theta)\}_{t-1}$.

$\theta^{**} \sim KP_{t,m^{**}}(\theta|\theta^*)$

Accept/reject (m^{**}, θ^{**}).

If accepted, calculate the weights.

(c)

(d)

5

Figure 4:

(a) Particles are sampled from the prior $P(m, \theta)$ until N particles have been accepted (in this example $N = 6$ for illustration purposes, in practice N should be larger), that is the distance is smaller than ϵ_1. The weights $w_{t=1}$ are calculated for all accepted particles and normalized.

(b) To obtain the marginal probability of m, we sum over all the weights corresponding to the model of interest: $P_{t=1}(m') = \sum_{m=m'} w_{t=1}(m, \theta)$. The histogram of the marginal population 1 is presented.

(c) This figure shows how we propose and accept a particle of population t, $t = 2, \ldots, T$. We sample a model m^* from population $t - 1$ with probability $P_{t-1}(m^*)$. For example, we might have sampled $m^* = m_2$. We then perturb the model using a model perturbation kernel KM_t to obtain $m^{**} \sim KM_t(m|m^*)$, for example $m^{**} = m_3$. After we have obtained the model m^{**}, we sample a parameter θ^* belonging to model m^{**} from population $t - 1$ and perturb it to obtain $\theta^{**} \sim KP_{t,m^{**}}(\theta|\theta^*)$. We simulate a dataset D^* for a particle (m^{**}, θ^{**}) and accept $(d(D_0, D^*) \leq \epsilon_t)$ or reject the particle. If a particle is accepted, we calculate its weight $w_t(m^{**}, \theta^{**})$.

(d) When N particles of population t have been accepted, we normalize the weights and marginalize them in order to obtain the marginal intermediate population of the model $P_t(m)$. We continue until population T, which is the approximation of the joint posterior distribution $P(m, \theta|D_0)$. The quantities of interest are the intermediate and the last marginal population of the model. In this example there are three populations, $T = 3$.

References

[1] Beaumont MA, Zhang W and Balding DJ. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

[2] Marjoram P, Molitor J, Plagnol V and Tavare S. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA*, 100(26):15324–8, 2003.

[3] Sisson SA, Fan Y and Tanaka MM. Sequential Monte Carlo without likelihoods. *Proc Natl Acad Sci USA*, 104(6):1760–5, 2007.

[4] Toni T, Welch D, Strelkowa N, Ipsen A and Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface*, 6:187–202, 2009.

[5] Beaumont MA, Cornuet JM, Marin JM and Robert CP. Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990, 2009.

[6] Grelaud A, Robert CP, Marin JM, Rodolphe F and Taly JF. ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*, 4(2):317–336, 2009.

[7] Toni T and Stumpf MPH. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1):104–10, 2010.