

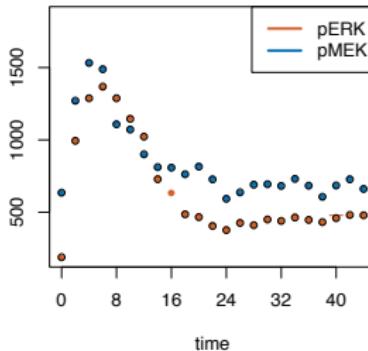
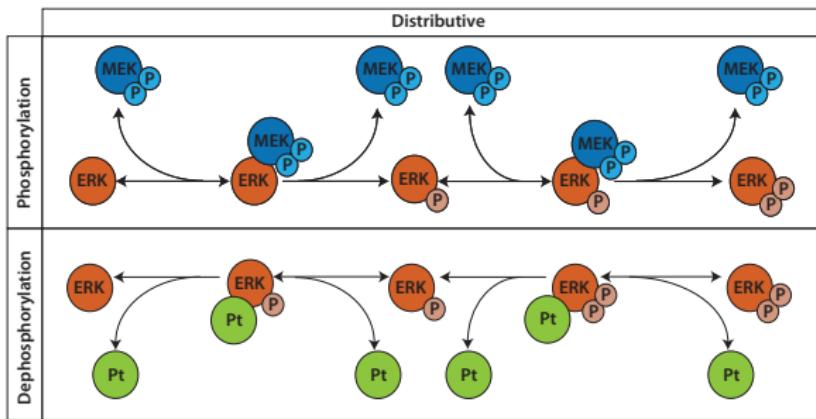
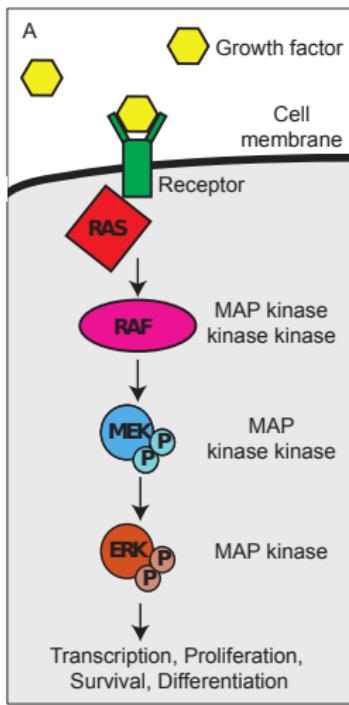
Approximate Bayesian Computation

Sarah Filippi

Department of Statistics
University of Oxford

09/02/2016

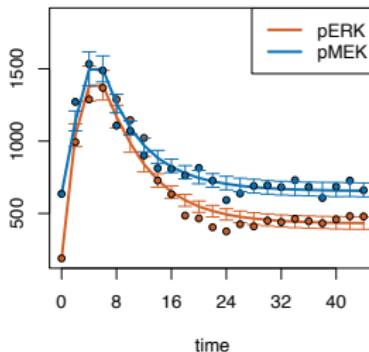
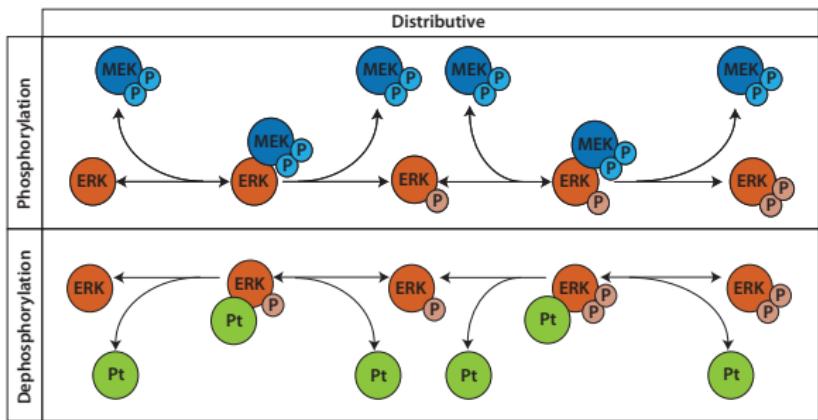
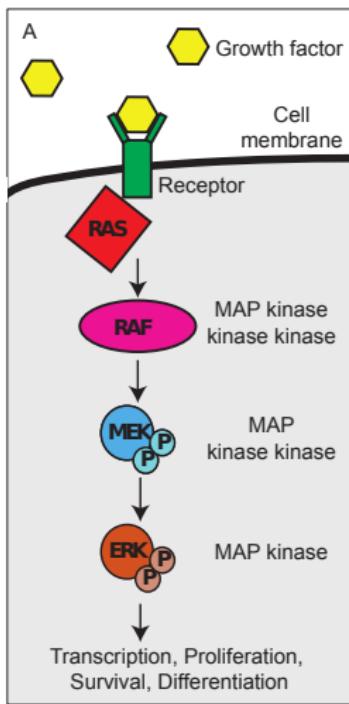
Parameter inference in a signalling pathway



Model via differential equations

Species	9
Parameters	16
Initial conditions	2

Parameter inference in a signalling pathway



Model via differential equations

Species	9
Parameters	16
Initial conditions	2

Notations

We have:

- an observed dataset, denoted by x^*
- a parametric model, either deterministic or stochastic.

Model = data-generating process:

given a value of parameter $\theta \in \mathbb{R}^d$, the output of the system is

$$X \sim f(\cdot | \theta)$$

The likelihood function: $\theta \mapsto f(x^* | \theta)$

"All models are wrong but some are useful." George E.P. Box

Aim:

- **parameter inference:** determine how likely it is for each value of θ to explain the data
- **model selection:** rank candidate models in terms of their ability to explain the data

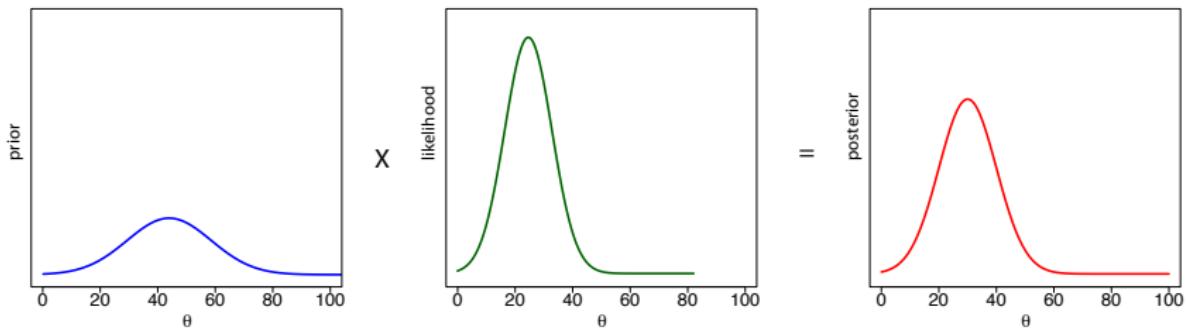
Bayesian methods

In the Bayesian framework, we combine

- the information brought by the data x^* , via the likelihood $f(x^*|\theta)$
- some a priori information, specified in a prior distribution $\pi(\theta)$

These informations are summarized in the posterior distribution, which is derived using the Bayes formula:

$$p(\theta|x^*) = \frac{f(x^*|\theta)\pi(\theta)}{\int f(x^*|\theta')\pi(\theta')d\theta'}$$



Sample from the posterior distribution

- In general, no closed-form solution
- Use computer simulation and Monte-Carlo techniques to sample from the posterior distribution
- Typical likelihood-based approaches:
 - Importance sampling
 - Markov Chain Monte-Carlo (MCMC)
 - Population Monte-Carlo
 - Sequential Monte-Carlo (SMC) sampler
 - ...

What if we can not compute the likelihood?

In many applications of interest: either impossible or computationally too expensive.

- stochastic differential equations
- population genetics
- ...

Use so called [likelihood-free](#) method such as [Approximate Bayesian Computation \(ABC\)](#).

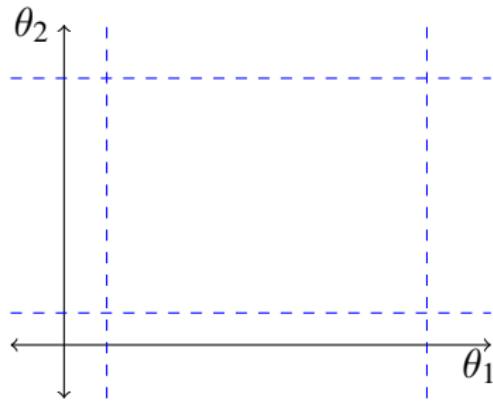
Approximate Bayesian Computation

- Principle:
 - sample parameters θ from the prior distribution
 - select the values of θ such that the simulated data are close to the observed data.
- More formally: given a small value of $\epsilon > 0$,

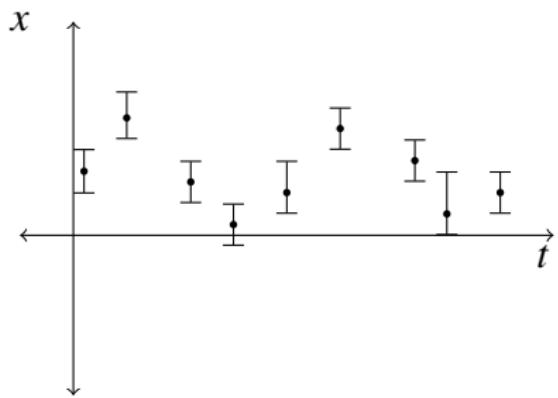
$$p(\theta|x^*) = \frac{f(x^*|\theta)\pi(\theta)}{p(x^*)} \approx p_\epsilon(\theta|x^*) = \frac{\int f(x|\theta)\pi(\theta)1\!\!1_{\Delta(x,x^*) \leq \epsilon} dx}{p(x^*)}$$

Approximate Bayesian Computation

Model



Data, x^*

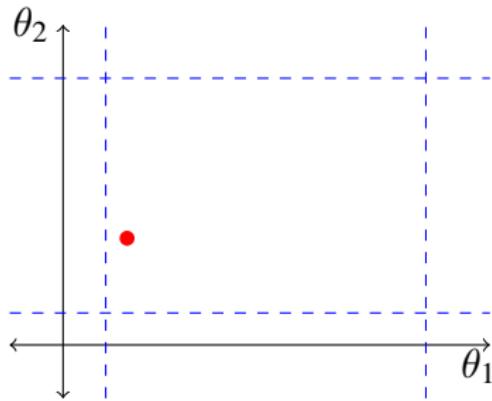


Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

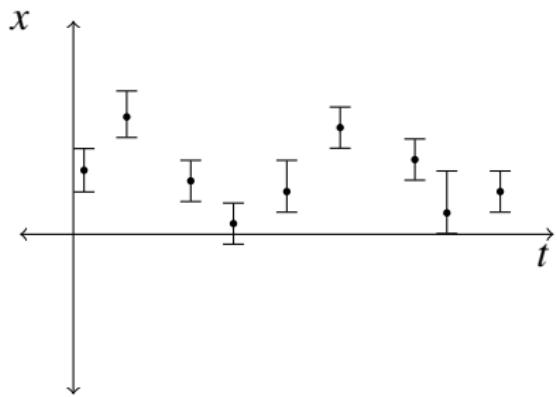
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*

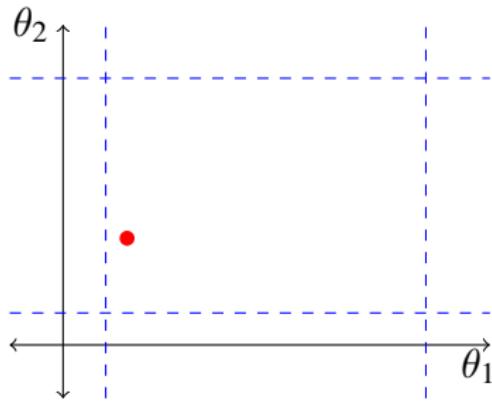


Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

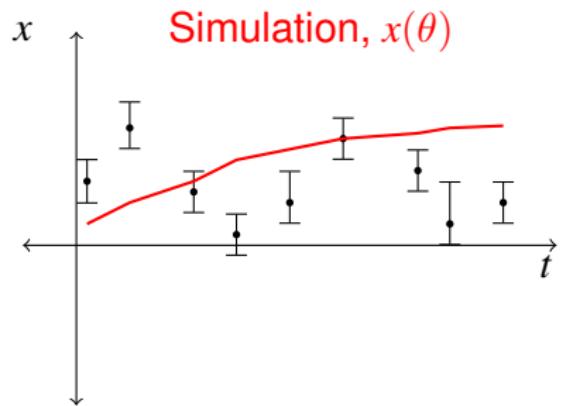
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*

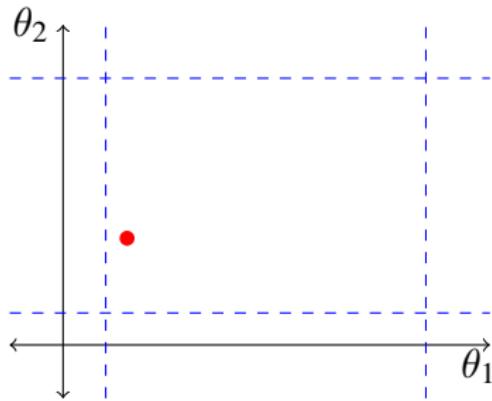


Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

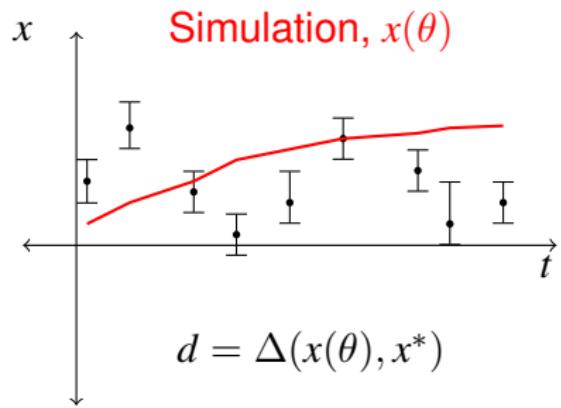
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*

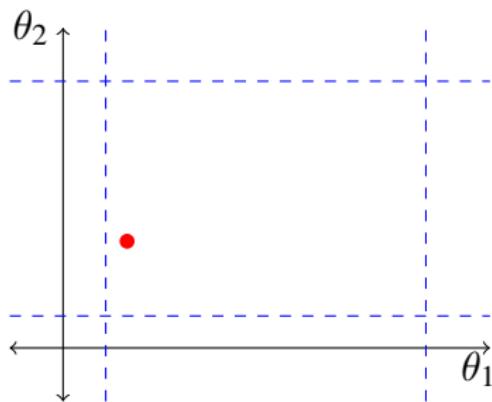


Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

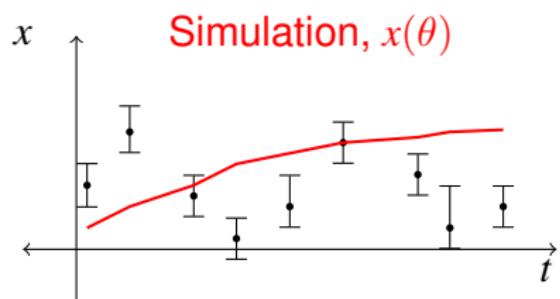
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*



$$d = \Delta(x(\theta), x^*)$$

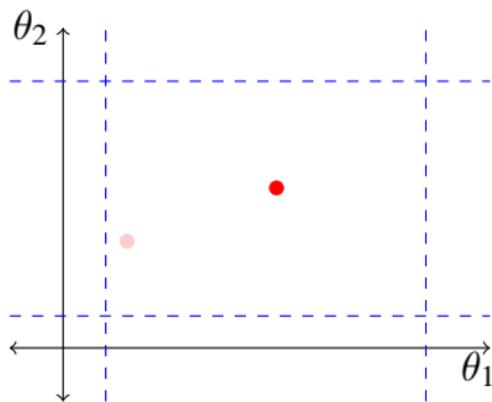
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

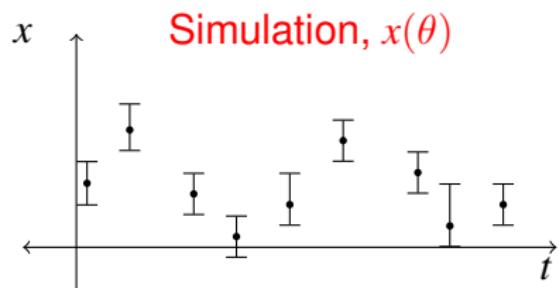
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*



Simulation, $x(\theta)$

$$d = \Delta(x(\theta), x^*)$$

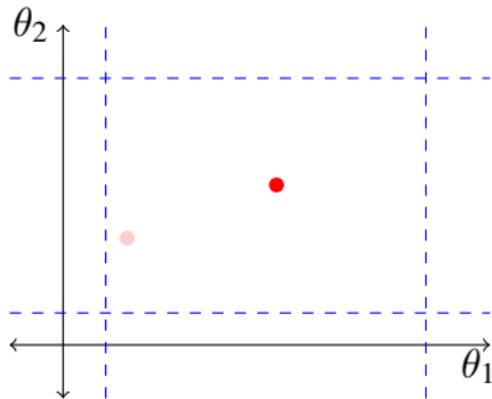
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

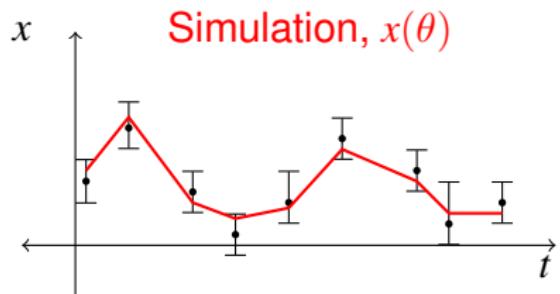
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*



$$d = \Delta(x(\theta), x^*)$$

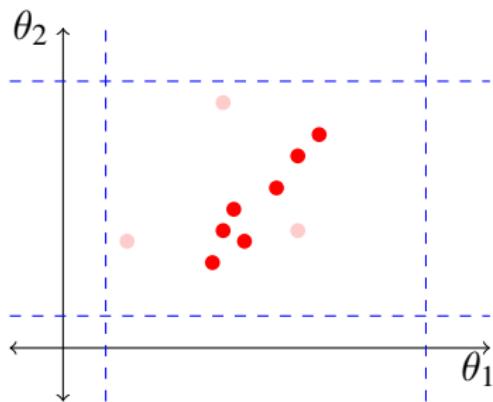
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

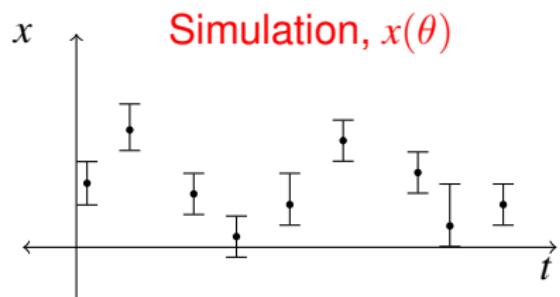
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*



$$d = \Delta(x(\theta), x^*)$$

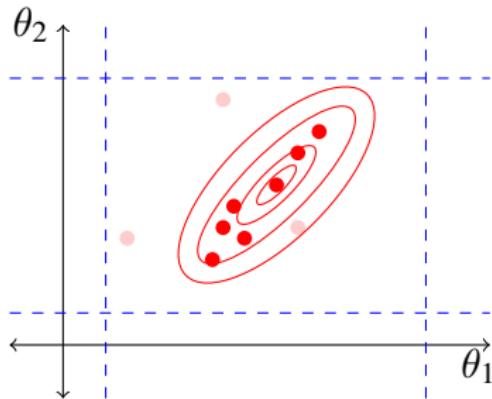
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

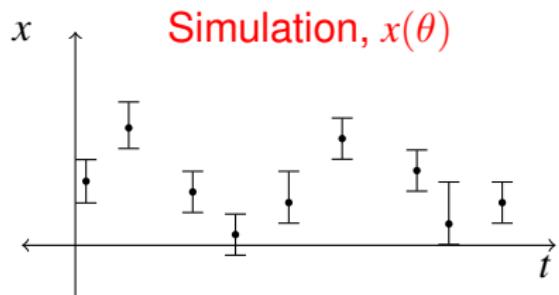
Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

Model



Data, x^*



Simulation, $x(\theta)$

$$d = \Delta(x(\theta), x^*)$$

Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.

Toni & Stumpf, Bioinformatics (2010)

Approximate Bayesian Computation

- Very sensitive to the choice of ϵ :
 - ϵ should be close to 0 to obtain samples from a distribution close to the posterior distribution

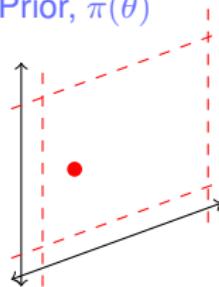
$$\lim_{\epsilon \rightarrow 0} p_\epsilon(\theta|x^*) = p(\theta|x^*)$$

- if ϵ too small, very low acceptance rate
- More efficient variants of ABC:
 - regression-adjusted ABC
Tallmon et al., 2004; Fagundes et al., 2007; Blum and François, 2010
 - Markov chain Monte Carlo ABC schemes
Marjoram and Molitor, 2003; Ratmann et al., 2007
 - ABC implementing some variant of sequential importance sampling (SIS) or sequential Monte Carlo (SMC)
Sisson et al., 2007; Beaumont et al., 2009; Toni et al., 2009; Del Moral et al., 2011

Sequential ABC – a schematic representation

Population 1

Prior, $\pi(\theta)$

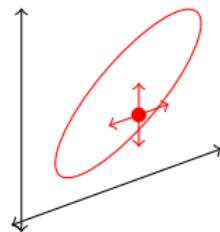


Define set of intermediate distributions, $\pi_t, t = 1, \dots, T$

$$\epsilon_1 > \epsilon_2 > \dots > \epsilon_T$$

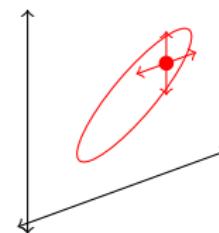
Population $t - 1$

$$\pi_{t-1}(\theta | \Delta(x^*, x(\theta)) < \epsilon_{t-1})$$



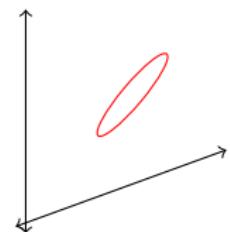
Population t

$$\pi_t(\theta | \Delta(x^*, x(\theta)) < \epsilon_t)$$



Population T

$$\pi_T(\theta | \Delta(x^*, x(\theta)) < \epsilon_T)$$



Acknowledgement to Prof. Michael Stumpf (Theoretical System Biology group, Imperial College London) for the slides.
Toni et al., J.Roy.Soc. Interface (2009); Toni & Stumpf, Bioinformatics (2010).

Sequential ABC – the approach (1)

In the rest of this lecture, we focus on the approach proposed by Beaumont et al., 2009; Toni et al., 2009.

The algorithms are related to the population Monte-Carlo algorithm (Cappé et al, 2004) and invoke importance sampling arguments:

- if a sample $\theta^{(t)} = (\theta^{(1,t)}, \dots, \theta^{(N,t)})$ is produced by simulating each $\theta^{(i,t)} \sim q_{it}$ independent of now another conditional on the past samples
- and if each point $\theta^{(i,t)}$ is associated to an important weight

$$\omega^{(i,t)} \propto \frac{\pi(\theta^{(i,t)})}{q_{it}}$$

then $\frac{1}{N} \sum_{i=1}^N \omega^{(i,t)} h(\theta^{(i,t)})$ is an unbiased estimator of $\int h(\theta) \pi(\theta) d\theta$.

Sequential ABC – the approach (2)

In the sequential ABC algorithm, at each population t , we consider a sample $\theta^{(t)} = (\theta^{(1,t)}, \dots, \theta^{(N,t)})$.

Each $\theta^{(i,t)}$ is sampled independently as follows:

- θ is sampled from the previous population $\{\theta^{(i,t-1)}, \omega^{(i,t-1)}\}_{1 \leq i \leq N}$,
- then it is perturbed using a perturbation kernel $\tilde{\theta} \sim K_t(\cdot | \theta)$

therefore the proposal distribution is

$$q_t(\cdot) = \sum_{j=1}^n \omega^{(j,t-1)} K_t(\cdot | \theta^{(j,t-1)})$$

and the important weight associated to $\tilde{\theta}$ is

$$\omega \propto \frac{\pi(\theta)}{\sum_{j=1}^n \omega^{(j,t-1)} K_t(\theta | \theta^{(j,t-1)})} .$$

Sequential ABC – the algorithm

```
1: for all  $t$  do
2:    $i \leftarrow 1$ 
3:   repeat
4:     if  $t=1$  then
5:       sample  $\tilde{\theta}$  from  $\pi(\theta)$ 
6:     else
7:       sample  $\theta$  from the previous population  $\{\theta^{(i,t-1)}, \omega^{(i,t-1)}\}_{1 \leq i \leq N}$ 
8:       perturb  $\tilde{\theta}$  from  $K_t(\cdot|\theta)$  so that  $\pi(\tilde{\theta}) > 0$ 
9:     end if
10:    sample  $x$  from  $f(\cdot|\tilde{\theta})$ 
11:    if  $\Delta(x^*, x) \leq \epsilon_t$  then
12:       $\theta^{(i,t)} \leftarrow \tilde{\theta}; \quad i \leftarrow i + 1$ 
13:    end if
14:  until  $i = N + 1$ 
15:  calculate the weights:  $\omega^{(i,t)} \propto \frac{\pi(\theta^{(i,t)})}{\sum_{j=1}^n \omega^{(j,t-1)} K_t(\theta^{(i,t)} | \theta^{(j,t-1)})}; \quad \omega^{(i,1)} = 1/N$ 
16: end for
```

Toni *et al.*, 2009; Beaumont *et al.*, 2009

Sequential ABC – the algorithm

```
1: for all  $t$  do
2:    $i \leftarrow 1$ 
3:   repeat
4:     if  $t=1$  then
5:       sample  $\tilde{\theta}$  from  $\pi(\theta)$ 
6:     else
7:       sample  $\theta$  from the previous population  $\{\theta^{(i,t-1)}, \omega^{(i,t-1)}\}_{1 \leq i \leq N}$ 
8:       perturb  $\tilde{\theta}$  from  $K_t(\cdot | \theta)$  so that  $\pi(\tilde{\theta}) > 0$ 
9:     end if
10:    sample  $x$  from  $f(\cdot | \tilde{\theta})$ 
11:    if  $\Delta(x^*, x) \leq \epsilon_t$  then
12:       $\theta^{(i,t)} \leftarrow \tilde{\theta}; \quad i \leftarrow i + 1$ 
13:    end if
14:  until  $i = N + 1$ 
15:  calculate the weights:  $\omega^{(i,t)} \propto \frac{\pi(\theta^{(i,t)})}{\sum_{j=1}^n \omega^{(j,t-1)} K_t(\theta^{(i,t)} | \theta^{(j,t-1)})}; \quad \omega^{(i,1)} = 1/N$ 
16: end for
```

Impact on computational efficiency:

- perturbation kernels $\{K_t\}_t$
- threshold schedule $\{\epsilon_t\}_t$

Toni *et al.*, 2009; Beaumont *et al.*, 2009

Use adaptive perturbation kernel

- Local perturbation kernel:
 - hardly moves particles
 - high acceptance rate if successive values of ϵ close enough
- Widely spread kernel:
 - exploration of the parameter space
 - low acceptance rate

Balance between exploring the parameter space and ensuring a high acceptance rate

Properties of optimal kernel

1. From sequential importance sampling theory:

similarity between two joint distributions of $(\theta^{(t-1)}, \theta^{(t)})$ where $\theta^{(t-1)} \sim p_{\epsilon_{t-1}}$ and

- $\theta^{(t)}$ constructed by perturbing $\theta^{(t-1)}$ and accepting according to threshold ϵ_t
- $\theta^{(t)} \sim p_{\epsilon_t}$

2. Computational efficiency: high acceptance rate

3. Theoretical requirements for convergence:

- kernel with larger support than the target distribution
⇒ guarantee asymptotic unbiasedness of the empirical mean
- vanish slowly enough in the tails of the target
⇒ guarantee finite variance of the estimator

Derivation of optimal kernel

Criteria 1: Resemblance between the two distributions

$$q_{\epsilon_{t-1}, \epsilon_t}(\theta^{(t-1)}, \theta^{(t)} | x) = \frac{p_{\epsilon_{t-1}}(\theta^{(t-1)} | x) K_t(\theta^{(t)} | \theta^{(t-1)}) \int f(x | \theta^{(t)}) \mathbb{1}(\Delta(x^*, x) \leq \epsilon_t) dx}{\alpha(K_t, \epsilon_{t-1}, \epsilon_t, x)}$$

and

$$q_{\epsilon_{t-1}, \epsilon_t}^*(\theta^{(t-1)}, \theta^{(t)} | x) = p_{\epsilon_{t-1}}(\theta^{(t-1)} | x) p_{\epsilon_t}(\theta^{(t)} | x)$$

in terms of the KL divergence ([Douc et al, 2007](#); [Cappé et al, 2008](#); [Beaumont et al, 2009](#))

$$KL(q_{\epsilon_{t-1}, \epsilon_t}; q_{\epsilon_{t-1}, \epsilon_t}^*) = -Q(K_t, \epsilon_{t-1}, \epsilon_t, x) + \log \alpha(K_t, \epsilon_{t-1}, \epsilon_t, x) + C(\epsilon_{t-1}, \epsilon_t, x)$$

where

$$Q(K_t, \epsilon_{t-1}, \epsilon_t, x) = \iint p_{\epsilon_{t-1}}(\theta^{(t-1)} | x) p_{\epsilon_t}(\theta^{(t)} | x) \log K_t(\theta^{(t)} | \theta^{(t-1)}) d\theta^{(t-1)} d\theta^{(t)}$$

Derivation of optimal kernel

Criteria 1: minimise the KL divergence:

$$KL(q_{\epsilon_{t-1}, \epsilon_t}; q_{\epsilon_{t-1}, \epsilon_t}^*) = -Q(K_t, \epsilon_{t-1}, \epsilon_t, x) + \log \alpha(K_t, \epsilon_{t-1}, \epsilon_t, x) + C(\epsilon_{t-1}, \epsilon_t, x)$$

Derivation of optimal kernel

Criteria 1: minimise the KL divergence:

$$KL(q_{\epsilon_{t-1}, \epsilon_t}; q_{\epsilon_{t-1}, \epsilon_t}^*) = -Q(K_t, \epsilon_{t-1}, \epsilon_t, x) + \log \alpha(K_t, \epsilon_{t-1}, \epsilon_t, x) + C(\epsilon_{t-1}, \epsilon_t, x)$$

Criteria 2: maximise the acceptance rate $\alpha(K_t, \epsilon_{t-1}, \epsilon_t, x)$

Method

Select the kernel K_t which maximises $Q(K_t, \epsilon_{t-1}, \epsilon_t, x)$ which is equivalent to maximise $-KL(q_{\epsilon_{t-1}, \epsilon_t}; q_{\epsilon_{t-1}, \epsilon_t}^*) + \log \alpha(K_t, \epsilon_{t-1}, \epsilon_t, x)$.

Remark:

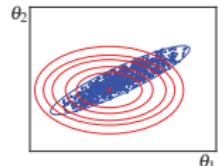
$$Q(K_t, \epsilon_{t-1}, \epsilon_t, x) = \iint p_{\epsilon_{t-1}}(\theta^{(t-1)}|x) p_{\epsilon_t}(\theta^{(t)}|x) \log K_t(\theta^{(t)}|\theta^{(t-1)}) d\theta^{(t-1)} d\theta^{(t)}$$

can be maximized easily for some families of kernels.

Gaussian random walk kernels

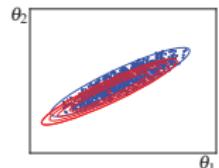
- **Component-wise kernel:**

diagonal covariance matrix $\Sigma^{(t)}$ [Beaumont et al, 2009](#)



- **Multi-variate normal kernel:**

$$\Sigma^{(t)} \approx \sum_{i=1}^N \sum_{k=1}^{N_0} \omega^{(i,t-1)} \tilde{\omega}^{(k)} (\tilde{\theta}^{(k)} - \theta^{(i,t-1)}) (\tilde{\theta}^{(k)} - \theta^{(i,t-1)})^T$$



where $\left\{ \tilde{\theta}^{(k)} \right\}_{1 \leq k \leq N_0} = \left\{ \theta^{(i,t-1)} \text{ s.t. } \Delta(x^*, x^{(i,t-1)}) \leq \epsilon_t, 1 \leq i \leq N \right\}$

- **Local multi-variate normal kernels:**

use a different kernel for each particle $\theta^{(t-1)}$.

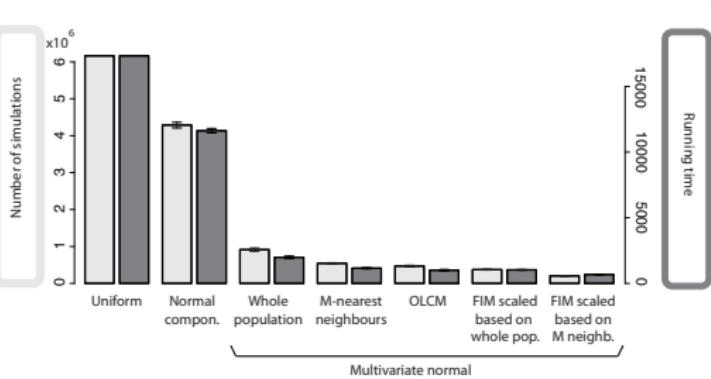
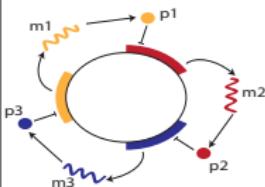
- Nearest neighbours: $\Sigma_{\theta^{(t-1)}, M}^{(t)}$ based on the M nearest neighbours
- Optimal local covariance matrix:

$$\Sigma_{\theta^{(t-1)}}^{(t)} \approx \sum_{k=1}^{N_0} \tilde{\omega}^{(k)} (\tilde{\theta}^{(k)} - \theta^{(t-1)}) (\tilde{\theta}^{(k)} - \theta^{(t-1)})^T$$

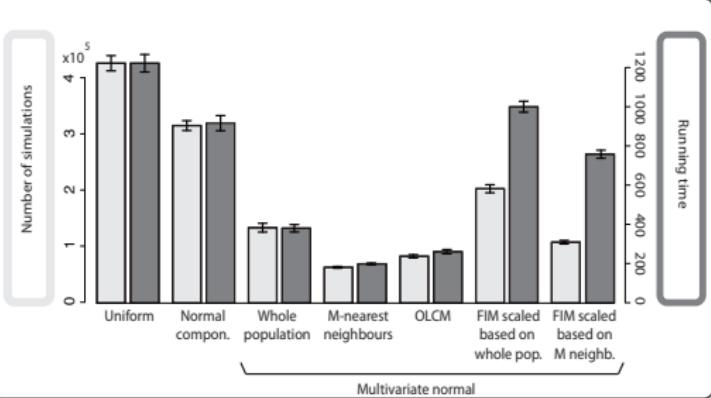
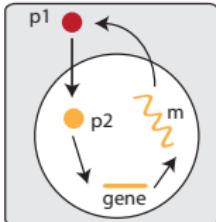
- Based on FIM for models defined by ODE or SDE

Computational efficiency of perturbation kernels

The Repressilator system



The Hes1 transcriptional regulation model



Computational cost of perturbation kernel

- simulating the data dominates
- computational cost of perturbation kernel implementation:

Component-wise normal	$O(dN^2)$
Multivariate normal based on the whole previous population	$O(d^2N^2)$
Multivariate normal based on the M nearest neighbours	$O((d + M)N^2 + d^2M^2N)$
Multivariate normal with OLCM	$O(d^2N^2)$
Multivariate normal based on the FIM (normalized with entire population)	$O(dCN + d^2N^2)$

N = population size; d = parameter dimension

Recommendation

Use of multivariate kernels with OLCM

- highest acceptance rate in our examples
- relatively easy to implement at acceptable computational cost

Choice of the threshold schedule

```
1: for all  $t$  do
2:    $i \leftarrow 1$ 
3:   repeat
4:     if  $t=1$  then
5:       sample  $\tilde{\theta}$  from  $\pi(\theta)$ 
6:     else
7:       sample  $\theta$  from the previous population  $\{\theta^{(i,t-1)}, \omega^{(i,t-1)}\}_{1 \leq i \leq N}$ 
8:       perturb  $\tilde{\theta}$  from  $K_t(\cdot|\theta)$  so that  $\pi(\tilde{\theta}) > 0$ 
9:     end if
10:    sample  $x$  from  $f(\cdot|\tilde{\theta})$ 
11:    if  $\Delta(x^*, x) \leq \epsilon_t$  then
12:       $\theta^{(i,t)} \leftarrow \tilde{\theta}; \quad i \leftarrow i + 1$ 
13:    end if
14:    until  $i = N + 1$ 
15:    calculate the weights:  $\omega^{(i,t)} \propto \frac{\pi(\theta^{(i,t)})}{\sum_{j=1}^n \omega^{(j,t-1)} K_t(\theta^{(i,t)} | \theta^{(j,t-1)})}; \quad \omega^{(i,1)} = 1/N$ 
16: end for
```

- final value of ϵ close to 0
- minimise number of simulations i.e.
 - minimise number of populations
 - maximise acceptance rate per pop.

Toni *et al.*, 2009; Beaumont *et al.*, 2009

Previous approaches

- Trial and error
- Adaptive method based on quantile

Beaumont *et al.*, 2009; Del Moral *et al.*, 2008; Drovandi *et al.*, 2011; Lenormand *et al.*, 2012

- ...

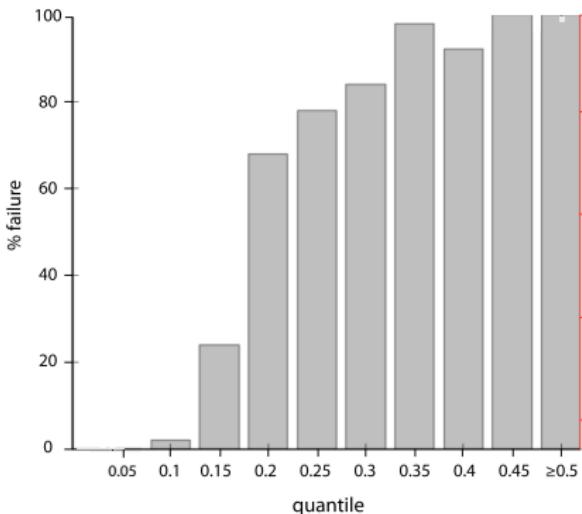
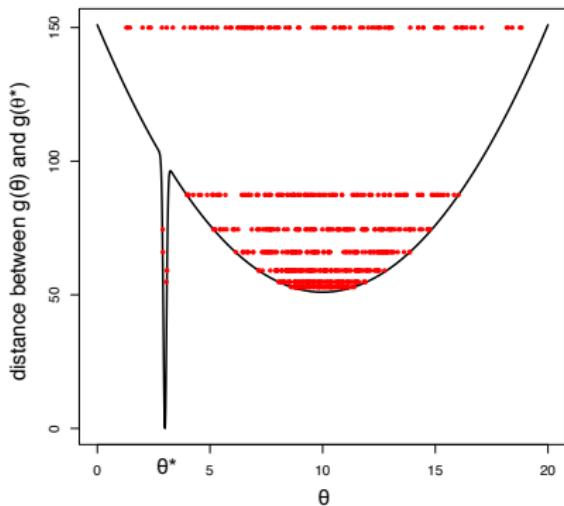
The quantile based approach:

$\epsilon_t = \alpha$ - quantile of the distances $\{\Delta(x^{(i,t-1)}, x^*)\}_{1 \leq i \leq N}$

Which value of α should we chose?

An adaptive choice of α : Sedki *et al.*, 2013.

Drawback of the quantile approach

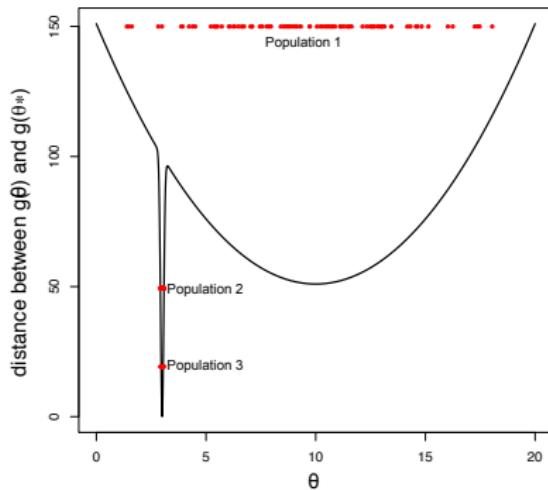
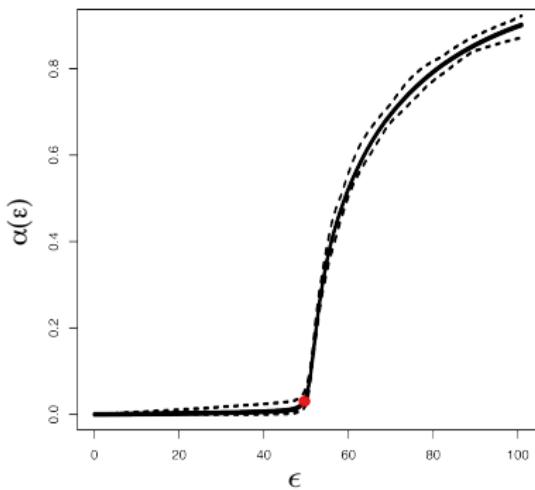


- For large α , the algorithm ‘fails’.
- The optimal (or safe) choice of α depends on the data, the model and the prior range.

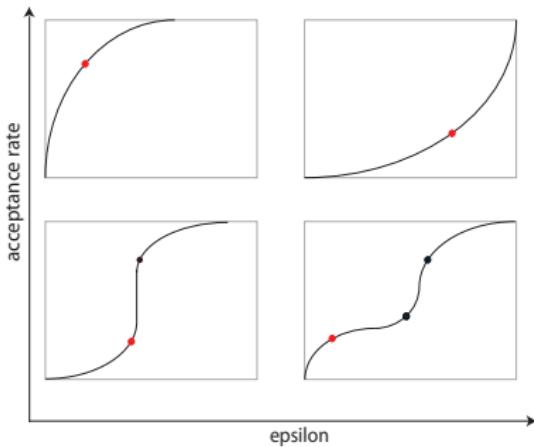
The threshold-acceptance rate curve

Idea:

- use the threshold-acceptance rate curve
- avoid area with excessively high acceptance rate



Exploit the threshold-acceptance rate curve



Proposed method:

- Set $\epsilon^* = \operatorname{argmax}_\epsilon \frac{\partial^2 \alpha_t}{\partial \epsilon^2}$
- if $\alpha(\epsilon^*)$ not too small, $\epsilon_t = \epsilon^*$
- otherwise
$$\epsilon_t = \operatorname{argmin}_\epsilon \Delta\left(\left(\frac{\epsilon}{\epsilon_{t-1}}, \frac{\alpha_t(\epsilon)}{\alpha_t(\epsilon_{t-1})}\right), (0, 1)\right)$$

Convex shapes: symptom of sampling from local optima.

Estimating the threshold-acceptance rate curve

- Acceptance rate

$$\alpha_t(\epsilon) = \int \int q_t(\theta) f(x|\theta) \mathbb{1} (\Delta(x, x^*) \leq \epsilon) dx d\theta$$

where $q_t(\theta) = \sum_{i=1}^N \omega^{(i,t-1)} K_t(\theta|\theta^{(i,t-1)})$

- No closed-form expression and difficult to approximate

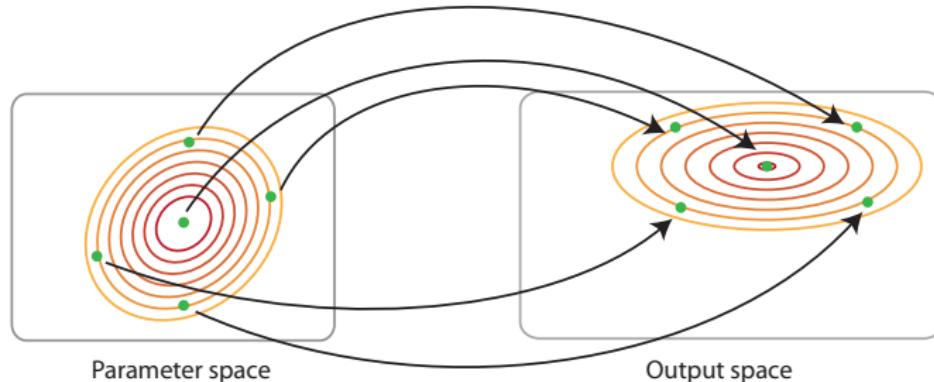
Estimating the threshold-acceptance rate curve

- Acceptance rate

$$\alpha_t(\epsilon) = \int \int q_t(\theta) f(x|\theta) \mathbb{1} (\Delta(x, x^*) \leq \epsilon) dx d\theta$$

where $q_t(\theta) = \sum_{i=1}^N \omega^{(i,t-1)} K_t(\theta|\theta^{(i,t-1)})$

- No closed-form expression and difficult to approximate
- For deterministic models, use Unscented Transform (UT)



Adaptive method for threshold choice

At each population t ,

1. generate a population of perturbed particles,
2. fit a Gaussian mixture model to the perturbed population
3. estimate

$$p_t(x) = \int q_t(\theta) f(x|\theta) d\theta$$

using the unscented transform independently for each component of the Gaussian mixture,

4. estimate

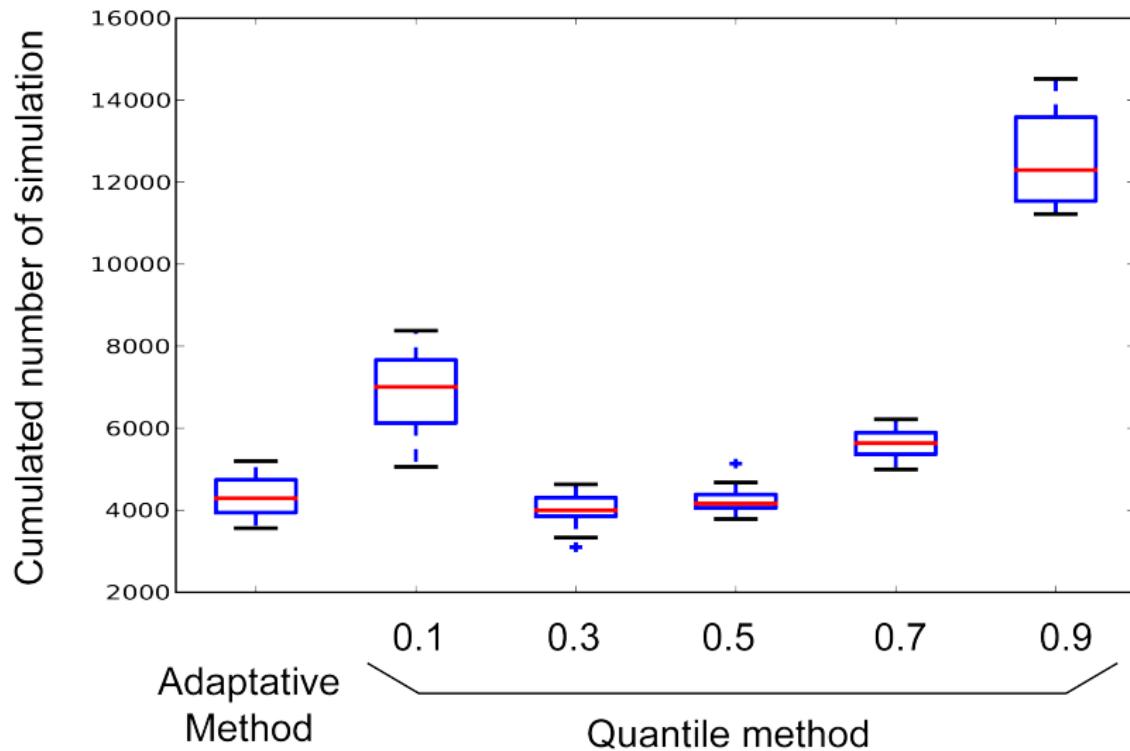
$$\alpha_t(\epsilon) = \int p_t(x) \mathbb{1} (\Delta(x, x^*) \leq \epsilon) dx$$

Remark

The threshold-acceptance rate curve depends on the perturbation kernel.

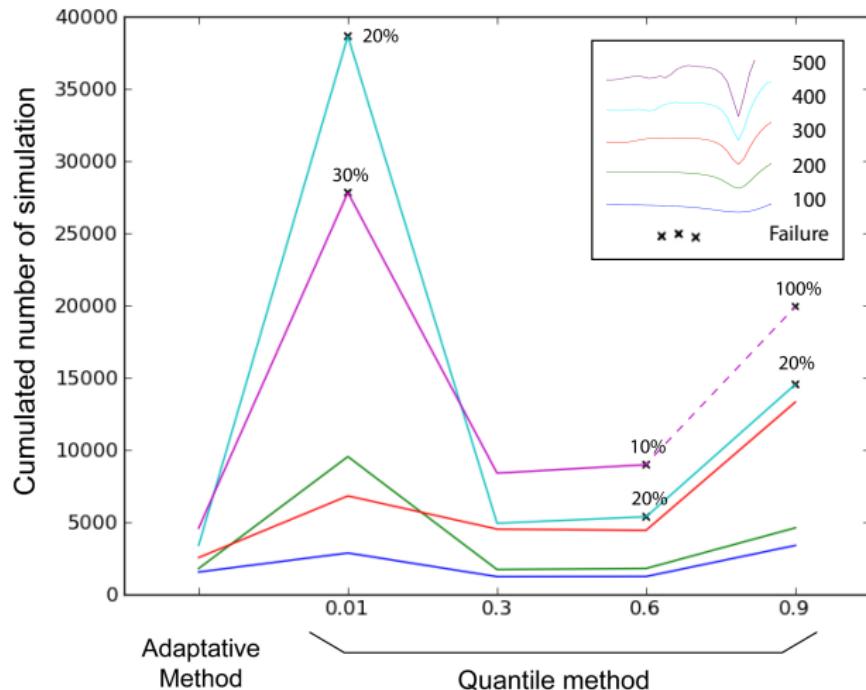
Computational efficiency of the adaptive method

The Repressilator system:



Computational efficiency of the adaptive method

Chemical reaction system illustrating the "local minimum problem"



Take-home messages

Approximate Bayesian Computation

To be used for parameter inference when the likelihood can not be computed but it is possible to simulate from the model.

Sequential approaches

Computationally more efficient.

Perturbation kernel

Gaussian random walk with the optimal (local) covariance matrix

- low computational cost
- high acceptance rate

Threshold schedule

- the quantile approach can lead to the wrong posterior
- exploit the threshold -acceptance rate curve

Filippi et al, 2013; Silk, Filippi, Stumpf (2013.)

References

- Beaumont, M., Cornuet, J.-M., Marin, J.-M., and Robert, C. (2009). *Adaptive approximate Bayesian computation*. Biometrika.
- Blum, M. and Francois, O. (2010). *Non-linear regression models for approximate Bayesian computation*. Statist. Comput.
- Cappe, O., Guillin, A., Marin, J.-M., and Robert, C. (2004). *Population Monte Carlo*. J. Comput. Graph. Statist.
- Del Moral, P., Doucet, A., and Jasra, A. (2006). *Sequential Monte Carlo samplers*. JRSS Series B
- Fagundes, N., et al. (2007). *Statistical evaluation of alternative models of human evolution*. PNAS.
- [Filippi, S., Barnes, C., Cornebise, J., and Stumpf, M. \(2013\)](#). *On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo*. SAGMB
- Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). *Markov chain Monte Carlo without likelihoods*. PNAS.

References

- Ratmann, O., et al. (2007). *Using likelihood-free inference to compare evolutionary dynamics of the protein networks of H. pylori and P. falciparum*. PLoS Comput Biol .
- Silk, D., Filippi, S., and Stumpf, M. (2013). *Optimizing threshold-schedules for sequential approximate Bayesian computation: applications to molecular systems*. SAGMB
- Sisson, S. A., Fan, Y., and Tanaka, M. (2007). *Sequential Monte Carlo without likelihoods*. PNAS.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. (2009). *Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems*. JRSS Interface.

Other references

- Blum, M. (2010). *Approximate Bayesian Computation: a non-parametric perspective*. J. American Statist. Assoc.
- Fearnhead, P. and Prangle, D. (2012). *Constructing summary statistics for Approximate Bayesian Computation: semi-automatic Approximate Bayesian Computation*. JRSS: Series B, (With discussion.).
- Marin, J., Pudlo, P., Robert, C., and Ryder, R. (2012). *Approximate Bayesian computational methods*. Statistics and Computing,.
- Robert, C., Cornuet, J.-M., Marin, J.-M., and Pillai, N. (2011). *Lack of confidence in ABC model choice*. PNAS.
- Rubin, D. (1984). *Bayesianly justifiable and relevant frequency calculations for the applied statistician*. Ann. Statist.
- Wilkinson, R. (2013). *Approximate Bayesian computation (ABC) gives exact results under the assumption of model error*. SAGMB.