What to do when exact Bayes is impossible? Some tools for approximate Bayesian inference

Umberto Picchini
Centre for Mathematical Sciences,
Lund University
www.maths.lth.se/matstat/staff/umberto/
twitter: @uPicchini

Bayes@Lund 12 April 2018

What to do when exact Bayes is impossible?
Some tools for approximate Bayesian inference

Umberto Picchini
from May I will be at Department of Mathematical Sciences,
Chalmers and Gothenburg University
www.here/therewillbe/something/else/
twitter: @uPicchini

Bayes@Lund 12 April 2018

- I will briefly introduce a methodology that has literally revolutionised statistical inference for complex models in the last 10-15 years.
- For the last 30 years advancements in computer hardware have enabled modellers to become more and more ambitious.
- Complex models are needed to make sense of advanced experiments and multivariate (large) datasets.

However the advancements of statistical algorithms didn't proceed at the same (fast) pace as hardware and modelling advancements.

We wanted to consider realistic model for our data, but often **we could not** because of the lack of flexible statistical methods.

- I will briefly introduce a methodology that has literally revolutionised statistical inference for complex models in the last 10-15 years.
- For the last 30 years advancements in computer hardware have enabled modellers to become more and more ambitious.
- Complex models are needed to make sense of advanced experiments and multivariate (large) datasets.

However the advancements of statistical algorithms didn't proceed at the same (fast) pace as hardware and modelling advancements.

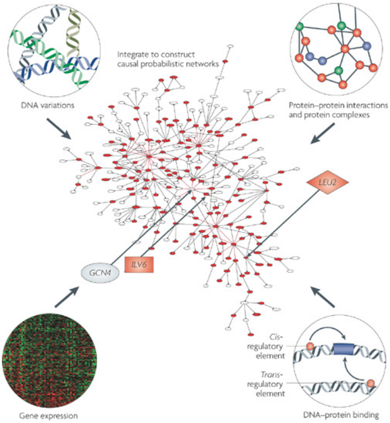
We wanted to consider realistic model for our data, but often **we could not** because of the lack of flexible statistical methods.

- I will briefly introduce a methodology that has literally revolutionised statistical inference for complex models in the last 10-15 years.
- For the last 30 years advancements in computer hardware have enabled modellers to become more and more ambitious.
- Complex models are needed to make sense of advanced experiments and multivariate (large) datasets.

However the advancements of statistical algorithms didn't proceed at the same (fast) pace as hardware and modelling advancements.

We wanted to consider realistic model for our data, but often **we could not** because of the lack of flexible statistical methods.

Most real-life modelling is way more complex than examples from courses textbooks. The likelihood of the object below might be totally out of reach.



Nature Reviews | Drug Discovery

[Pic from Schadt et al. (2009) doi:10.1038/nrd2826]

What we typically want is the **likelihood function** for model parameters θ:

- We have some data: $\mathbf{y}^o$.
- the likelihood function: $p(\mathbf{y}^o|\theta)$
- We consider data as the outcome of some probabilistic model, and write $\mathbf{y}^o \sim p(\mathbf{y}|\theta = \theta_0)$
- θ_0 is the **unknown** *ground-truth* value of θ.

Main issue

For realistically complex models, the likelihood function is **unavailable** in closed form.

Hence exact likelihood based inference is often **not possible**.

What we typically want is the **likelihood function** for model parameters θ:

- We have some data: $\mathbf{y}^o$.
- the likelihood function: $p(\mathbf{y}^o|\theta)$
- We consider data as the outcome of some probabilistic model, and write $\mathbf{y}^o \sim p(\mathbf{y}|\theta = \theta_0)$
- θ_0 is the **unknown** *ground-truth* value of θ.

Main issue

For realistically complex models, the likelihood function is **unavailable** in closed form.

Hence exact likelihood based inference is often **not possible**.

A paradigm shift is the concept of **generative model**.

You code a mathematical model $\mathcal{M}(\theta)$ as an idealized representation of the phenomenon under study.

$$\theta^* \rightarrow \mathcal{M}(\theta^*) \rightarrow \mathbf{y}^*$$

As long as we are able to run an instance of the model, we simulate/generate artificial data $\mathbf{y}^*$with $\mathbf{y}^* \sim p(\mathbf{y}^*|\theta = \theta^*)$.

So we have obtained a random *realization* $\mathbf{y}^*$ of the generative model $\mathcal{M}(\theta)$

Therefore the simulator $\mathcal{M}(\theta)$ defines the model pdf $p(\mathbf{y}|\theta)$ **implicitly**!

A paradigm shift is the concept of **generative model**.

You code a mathematical model $\mathcal{M}(\theta)$ as an idealized representation of the phenomenon under study.

$$\theta^* \to \mathcal{M}(\theta^*) \to \mathbf{y}^*$$

As long as we are able to run an instance of the model, we simulate/generate artificial data $\mathbf{y}^*$ with $\mathbf{y}^* \sim p(\mathbf{y}^*|\theta = \theta^*)$.

So we have obtained a random *realization* $\mathbf{y}^*$ of the generative model $\mathcal{M}(\theta)$

Therefore the simulator $\mathcal{M}(\theta)$ defines the model pdf $p(\mathbf{y}|\theta)$ **implicitly**!

A paradigm shift is the concept of **generative model**.

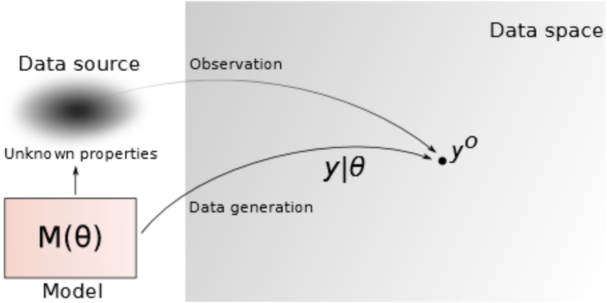
You code a mathematical model $\mathcal{M}(\theta)$ as an idealized representation of the phenomenon under study.

$$\theta^* \to \mathcal{M}(\theta^*) \to \mathbf{y}^*$$

As long as we are able to run an instance of the model, we simulate/generate artificial data $\mathbf{y}^*$with $\mathbf{y}^* \sim p(\mathbf{y}^*|\theta = \theta^*)$.

So we have obtained a random *realization* $\mathbf{y}^*$ of the generative model $\mathcal{M}(\theta)$

Therefore the simulator $\mathcal{M}(\theta)$ defines the model pdf $p(\mathbf{y}|\theta)$ **implicitly**!



We can use simulations from the generative model to produce inference about θ, **without explicit knowledge of the likelihood $p(\mathbf{y}|\theta)$.**

This is at the basis of **likelihood-free methods**

ABC, **a**pproximate **B**ayesian **c**omputation

ABC is probably the most important likelihood-free methodology.

We start by imposing a prior $\pi(\theta)$.

The first and simplest ABC algorithm is called **ABC rejection sampling**.

1. simulate from the prior $\theta^* \sim \pi(\theta)$
2. plug $\theta^* \to \mathcal{M}(\theta^*) \to \mathbf{y}^*$
3. if $\| \mathbf{y}^* - \mathbf{y}^o \| < \epsilon$ accept θ^* otherwise discard. Go to step 1 and repeat many times.

Each accepted pair $(\theta^*, \mathbf{y}^*)$ is from the augmented-posterior $\pi_\epsilon(\theta, \mathbf{y}^* | \mathbf{y}^o)$.

But we do not really care for $\mathbf{y}^*$, so if we retain only accepted θ^* then

$$\theta^* \sim \pi_\epsilon(\theta | \mathbf{y})$$

ABC, **a**pproximate **B**ayesian **c**omputation

ABC is probably the most important likelihood-free methodology.

We start by imposing a prior $\pi(\theta)$.

The first and simplest ABC algorithm is called **ABC rejection sampling**.

1. simulate from the prior $\theta^* \sim \pi(\theta)$
2. plug $\theta^* \to \mathcal{M}(\theta^*) \to \mathbf{y}^*$
3. if $\parallel \mathbf{y}^* - \mathbf{y}^o \parallel < \epsilon$ accept θ^* otherwise discard. Go to step 1 and repeat many times.

Each accepted pair $(\theta^*, \mathbf{y}^*)$ is from the augmented-posterior $\pi_\epsilon(\theta, \mathbf{y}^* | \mathbf{y}^o)$.

But we do not really care for $\mathbf{y}^*$, so if we retain only accepted θ^* then

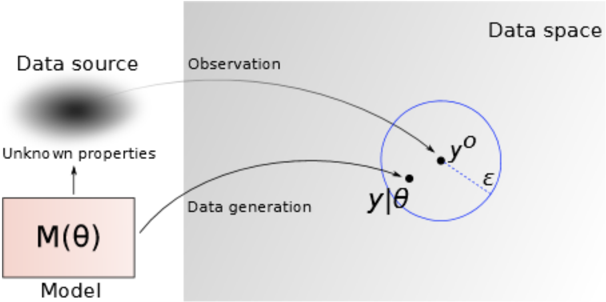
$$\theta^* \sim \pi_\epsilon(\theta | \mathbf{y})$$

Say that θ^* has been accepted by the ABC rejection sampling. Then:

- if $\epsilon = 0$ then $\theta^* \sim \pi(\theta|\mathbf{y})$, the **exact posterior**
- if $\epsilon = \infty$ then $\theta^* \sim \pi(\theta)$, the **prior**

Say that θ^* has been accepted by the ABC rejection sampling. Then:

- if $\epsilon = 0$ then $\theta^* \sim \pi(\theta|\mathbf{y})$, the **exact posterior**
- if $\epsilon = \infty$ then $\theta^* \sim \pi(\theta)$, the **prior**



Simulated data $\mathbf{y}^*$ inside the blue circle correspond to accepted parameters θ^*.

Bonus slide for the maths enthusiast

ABC rejection sampling produces draws from the joint "augmented posterior" $\pi_\epsilon(\theta, \mathbf{y}^*|\mathbf{y}^o)$ where

$$\pi_\epsilon(\theta, \mathbf{y}^*|\mathbf{y}^o) \propto \mathbb{I}_\epsilon(\mathbf{y}^*, \mathbf{y}^o) p(\mathbf{y}^*|\theta^*)\pi(\theta^*)$$

where $\mathbb{I}_\epsilon(\mathbf{y}^*, \mathbf{y}^o)$ equals 1 if $\| \mathbf{y}^* - \mathbf{y}^o \| < \epsilon$ and 0 otherwise.

However, in reality we do not need to store the $\mathbf{y}^*$ (we can just discard those immediately after we have evaluated $\| \mathbf{y}^* - \mathbf{y}^o \| < \epsilon$), and then $\theta^* \sim \pi_\epsilon(\theta|\mathbf{y}^o)$ where

$$\pi_\epsilon(\theta|\mathbf{y}^o) \propto \pi(\theta^*) \int_{\mathcal{Y}} \mathbb{I}_\epsilon(\mathbf{y}^*, \mathbf{y}^o) p(\mathbf{y}^*|\theta^*) d\mathbf{y}^*$$

Bonus slide for the maths enthusiast

ABC rejection sampling produces draws from the joint “augmented posterior” $\pi_\epsilon(\theta, \mathbf{y}^*|\mathbf{y}^o)$ where

$$\pi_\epsilon(\theta, \mathbf{y}^*|\mathbf{y}^o) \propto \mathbb{I}_\epsilon(\mathbf{y}^*, \mathbf{y}^o) p(\mathbf{y}^*|\theta^*) \pi(\theta^*)$$

where $\mathbb{I}_\epsilon(\mathbf{y}^*, \mathbf{y}^o)$ equals 1 if $\| \mathbf{y}^* - \mathbf{y}^o \| < \epsilon$ and 0 otherwise.

However, in reality we do not need to store the $\mathbf{y}^*$ (we can just discard those immediately after we have evaluated $\| \mathbf{y}^* - \mathbf{y}^o \| < \epsilon$), and then $\theta^* \sim \pi_\epsilon(\theta|\mathbf{y}^o)$ where

$$\pi_\epsilon(\theta|\mathbf{y}^o) \propto \pi(\theta^*) \int_{\mathcal{Y}} \mathbb{I}_\epsilon(\mathbf{y}^*, \mathbf{y}^o) p(\mathbf{y}^*|\theta^*) d\mathbf{y}^*$$

Toy model

Let's try something really trivial. We show how ABC rejection can become easily inefficient.

Suppose we have $n = 5$ i.i.d. observations $y_i \sim \mathit{Weibull}(2, 5)$.

Want to estimate parameters of the Weibull, so $\theta_0 = (2, 5) = (a, b)$ are the true values.

- take $\| \mathbf{y}^o - \mathbf{y}^* \| = \sqrt{\sum_{i=1}^n (y_i^o - y_i^*)^2}$ (you can try a different distance, this is not really crucial).
- We'll use different thresholds ϵ.
- Run 50,000 iterations of ABC rejection.

Toy model

Let's try something really trivial. We show how ABC rejection can become easily inefficient.

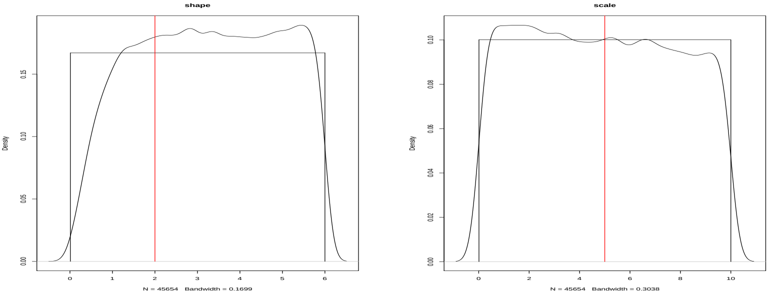
Suppose we have $n = 5$ i.i.d. observations $y_i \sim Weibull(2,5)$.

Want to estimate parameters of the Weibull, so $\theta_0 = (2,5) = (a,b)$ are the true values.

- take $\parallel \mathbf{y}^o - \mathbf{y}^* \parallel = \sqrt{\sum_{i=1}^n (y_i^o - y_i^*)^2}$ (you can try a different distance, this is not really crucial).
- We'll use different thresholds ϵ.
- Run 50,000 iterations of ABC rejection.

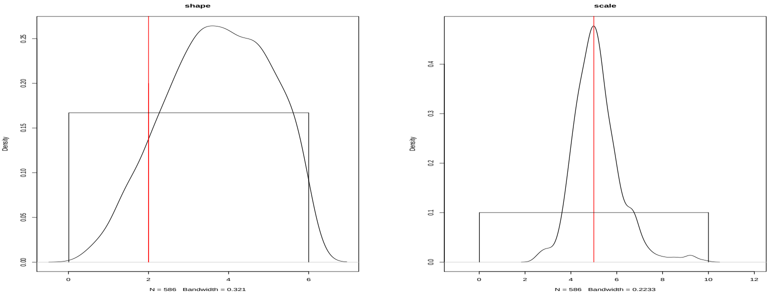
Wide priors for the “shape” parameter $a \sim U(0.01, 6)$ and “scale” $b \sim U(0.01, 10)$.

Try $\epsilon = 20$. True parameter values in red.



We are evidently sampling from the prior. Must reduce ϵ. About 92% draws were accepted. Way too large percentage!

Reduce ϵ from $\epsilon = 20$ to $\epsilon = 3$



About 1% of the produced simulations has been accepted.

Of course $n = 5$ is a very small sample size, so inference quality is necessarily limited, but you got the idea of the method.

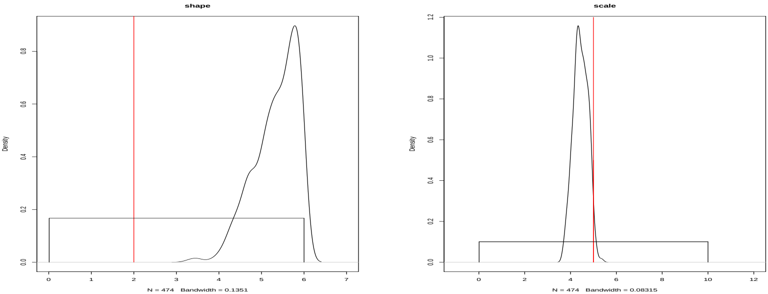
Curse of dimensionality

- results will degrade for a larger sample size *n* because of a "necessarily too large" ϵ;
- even for a moderately long dataset $\mathbf{y}^o$, how likely is that we simulate a $\mathbf{y}^*$ such that $\sum_{i=1}^n (y_i - y_i^*)^2 < \epsilon$ for **small** ϵ? Very unlikely.
- inevitably, we'll be forced to enlarge ϵ thus degrading the quality of the inference.
- Serious trade-off between computational efficience and statistical precision.

Curse of dimensionality

- results will degrade for a larger sample size n because of a "necessarily too large" ϵ;
- even for a moderately long dataset $\mathbf{y}^o$, how likely is that we simulate a $\mathbf{y}^*$ such that $\sum_{i=1}^n (y_i - y_i^*)^2 < \epsilon$ for **small** ϵ? Very unlikely.
- inevitably, we'll be forced to enlarge ϵ thus degrading the quality of the inference.
- Serious trade-off between computational efficience and statistical precision.

Here we take $n = 200$. To compare with our "best" previous result, we use $\epsilon = 31$ (to obtain again a 1% acceptance rate on 50,000 iterations).



Notice `shape` is completely off!

The approach is just **not going to be of any practical use with large datasets**.

Break the curse of dimensionality

Compress data information using some **summary statistics** $S(\mathbf{y})$.

Example: $S(\mathbf{y})$ may contain sample mean, standard deviation, autocorrelations, quantiles etc.

Idea: instead of comparing $\mathbf{y}^o$ with $\mathbf{y}^*$, compare $S(\mathbf{y}^o)$ with $S(\mathbf{y}^*)$.

Requirements:

- $\mathbf{S}(\cdot)$ should be "informative" regarding θ, as we give up on using the full data $\mathbf{y}$.
- $\mathbf{S}(\cdot)$ should not be too large. Ideally $\dim(\mathbf{S}) \equiv \dim(\theta)$ [Fearnhed& Prangle '12].[1]

[1] Fearnhead & Prangle (2012). JRSS-B, 74(3), 419-474.

Break the curse of dimensionality

Compress data information using some **summary statistics** $S(\mathbf{y})$.

Example: $S(\mathbf{y})$ may contain sample mean, standard deviation, autocorrelations, quantiles etc.

Idea: instead of comparing $\mathbf{y}^o$ with $\mathbf{y}^*$, compare $S(\mathbf{y}^o)$ with $S(\mathbf{y}^*)$.

Requirements:

- $\mathbf{S}(\cdot)$ should be "informative" regarding θ, as we give up on using the full data $\mathbf{y}$.
- $\mathbf{S}(\cdot)$ should not be too large. Ideally $\dim(\mathbf{S}) \equiv \dim(\theta)$ [Fearnhed& Prangle '12].[1]

[1] Fearnhead & Prangle (2012). JRSS-B, 74(3), 419-474.

Break the curse of dimensionality

Compress data information using some **summary statistics** $S(\mathbf{y})$.

Example: $S(\mathbf{y})$ may contain sample mean, standard deviation, autocorrelations, quantiles etc.

Idea: instead of comparing $\mathbf{y}^o$ with $\mathbf{y}^*$, compare $S(\mathbf{y}^o)$ with $S(\mathbf{y}^*)$.

Requirements:

- $\mathbf{S}(\cdot)$ should be "informative" regarding θ, as we give up on using the full data $\mathbf{y}$.
- $\mathbf{S}(\cdot)$ should not be too large. Ideally $\dim(\mathbf{S}) \equiv \dim(\theta)$ [Fearnhed& Prangle '12].[1]

[1] Fearnhead & Prangle (2012). JRSS-B, 74(3), 419-474.

ABC rejection with summaries (Pritchard et al.[2])

1. simulate from the prior $\theta^* \sim \pi(\theta)$
2. simulate $\mathcal{M}(\theta^*) \to \mathbf{y}^*$, compute $S(\mathbf{y}^*)$
3. if $\| S(\mathbf{y}^*) - S(\mathbf{y}^o) \| < \epsilon$ store θ^*. Go to step 1 and repeat.

Samples are from $\pi_\epsilon(\theta|S(\mathbf{y}^o))$

with

$$\pi_\epsilon(\theta|S(\mathbf{y}^o)) \propto \pi(\theta^*)\int_{\mathcal{Y}} \mathbb{I}_{A_{\epsilon,\mathbf{y}^o}}(\mathbf{y}^*)p(\mathbf{y}^*|\theta^*)d\mathbf{y}^*$$

$A_{\epsilon,\mathbf{y}^o}(\mathbf{y}^*) = \{\mathbf{y}^* \in \mathcal{Y}; \| S(\mathbf{y}^*) - S(\mathbf{y}^o) \| < \epsilon\}$.

[2] Pritchard et al. 1999, Molecular Biology and Evolution, 16:1791-1798.

ABC rejection with summaries (Pritchard et al.[2])

1. simulate from the prior $\theta^* \sim \pi(\theta)$
2. simulate $\mathcal{M}(\theta^*) \rightarrow \mathbf{y}^*$, compute $S(\mathbf{y}^*)$
3. if $\| S(\mathbf{y}^*) - S(\mathbf{y}^o) \| < \epsilon$ store θ^*. Go to step 1 and repeat.

Samples are from $\pi_\epsilon(\theta|S(\mathbf{y}^o))$

with

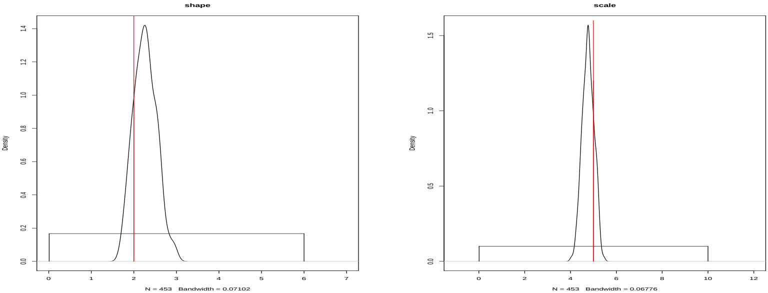
$$\pi_\epsilon(\theta|S(\mathbf{y}^o)) \propto \pi(\theta^*) \int_{\mathcal{Y}} \mathbb{I}_{A_{\epsilon,\mathbf{y}^o}}(\mathbf{y}^*) p(\mathbf{y}^*|\theta^*) d\mathbf{y}^*$$

$A_{\epsilon,\mathbf{y}^o}(\mathbf{y}^*) = \{\mathbf{y}^* \in \mathcal{Y}; \| S(\mathbf{y}^*) - S(\mathbf{y}^o) \| < \epsilon\}$.

[2] Pritchard et al. 1999, Molecular Biology and Evolution, 16:1791-1798.

Weibull example, reprise with $n = 200$

Set $S(\mathbf{y}) = (\text{sample mean of } \mathbf{y}, \text{sample SD of } \mathbf{y})$ and similarly for $\mathbf{y}^*$. Set $n = 200$. Use $\epsilon = 0.35$.



This time we have captured **both shape and scale** (with 1% acceptance).

Using summary statistics clearly introduces a further level of approximation. Except when $S(\cdot)$ is *sufficient* for θ (carries the same info about θ as the whole $\mathbf{y}^o$).

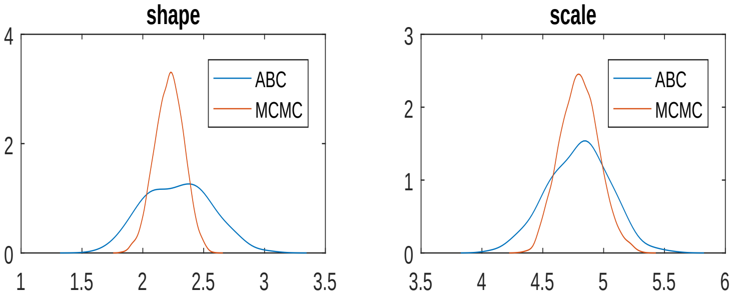
When $S(\cdot)$ is a set of sufficient statistics for θ,

$$\pi_\epsilon(\theta|S(\mathbf{y}^o)) = \pi_\epsilon(\theta|\mathbf{y}^o)$$

But then again when the distribution of $\mathbf{y}^o$ is not in the *exponential family*, we basically have no hope to construct sufficient statistics.

For this toy model, exact inference is possible.

Did we obtain an accurate approximation to the exact posterior?



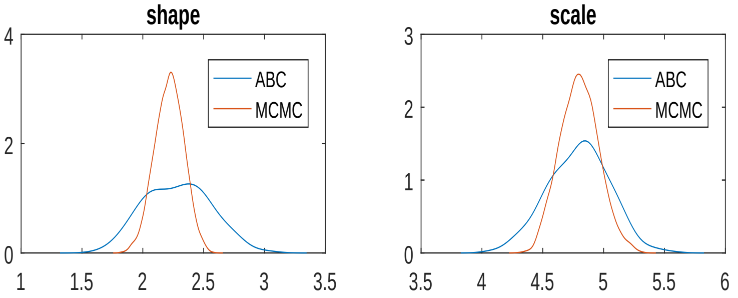
Exact posteriors (via MCMC) are in red.

ABC has 2 sources of approximation:

- we used arbitrary non-sufficient statistics.
- $\epsilon > 0$

For this toy model, exact inference is possible.

Did we obtain an accurate approximation to the exact posterior?



Exact posteriors (via MCMC) are in red.

ABC has 2 sources of approximation:

- we used arbitrary non-sufficient statistics.
- $\epsilon > 0$

Beyond ABC rejection

ABC rejection is the simplest example of ABC algorithm.

It generates independent draws and can be coded into an **embarrassingly parallel algorithm**. However in can be very inefficient.

Parameters are proposed from the prior $\pi(\theta)$. **A prior does not exploit the information of already accepted parameters.**

Unless $\pi(\theta)$ is somehow similar to the posterior, many proposals will be rejected for moderately small ϵ.

A natural approach is to consider ABC within an MCMC algorithm.

In a MCMC with random walk proposals, the proposed parameter explores a neighbourhood of the last accepted parameter.

Beyond ABC rejection

ABC rejection is the simplest example of ABC algorithm.

It generates independent draws and can be coded into an **embarrassingly parallel algorithm**. However in can be very inefficient.

Parameters are proposed from the prior $\pi(\theta)$. **A prior does not exploit the information of already accepted parameters.**

Unless $\pi(\theta)$ is somehow similar to the posterior, many proposals will be rejected for moderately small ϵ.

A natural approach is to consider ABC within an MCMC algorithm.

In a MCMC with random walk proposals, the proposed parameter explores a neighbourhood of the last accepted parameter.

Beyond ABC rejection

ABC rejection is the simplest example of ABC algorithm.

It generates independent draws and can be coded into an **embarrassingly parallel algorithm**. However in can be very inefficient.

Parameters are proposed from the prior $\pi(\theta)$. **A prior does not exploit the information of already accepted parameters.**

Unless $\pi(\theta)$ is somehow similar to the posterior, many proposals will be rejected for moderately small ϵ.

A natural approach is to consider ABC within an MCMC algorithm.

In a MCMC with random walk proposals, the proposed parameter explores a neighbourhood of the last accepted parameter.

ABC-MCMC

Integrating ABC within MCMC is **very simple** [Marjoram et al. 2003]

Notation: write $\mathbf{s}^* \equiv S(\mathbf{y}^*)$, $\mathbf{s}^o \equiv S(\mathbf{y}^o)$.
and $\mathbb{I}_\epsilon(\mathbf{s}^*, \mathbf{s}^o)$ equals 1 if $||\mathbf{s}^* - \mathbf{s}^o|| < \epsilon$, and 0 otherwise.

1. sample proposal $\theta^* \sim q(\theta^*|\theta^\#)$
2. plug $\theta^* \to \mathcal{M}(\theta^*) \to \mathbf{y}^* \to \mathbf{s}^*$
3. compute acceptance ratio:

$$\text{ratio} := \frac{\mathbb{I}_\epsilon(\mathbf{s}^*, \mathbf{s}^o)\pi(\theta^*)}{\mathbb{I}_\epsilon(\mathbf{s}^\#, \mathbf{s}^o)\pi(\theta^\#)} \times \frac{q(\theta^\#|\theta^*)}{q(\theta^*|\theta^\#)}$$

4. simulate $u \sim U(0,1)$, accept θ^* if $u <$ ratio.

ABC-MCMC

Integrating ABC within MCMC is **very simple** [Marjoram et al. 2003]

Notation: write $\mathbf{s}^* \equiv S(\mathbf{y}^*)$, $\mathbf{s}^o \equiv S(\mathbf{y}^o)$.
and $\mathbb{I}_\epsilon(\mathbf{s}^*, \mathbf{s}^o)$ equals 1 if $||\mathbf{s}^* - \mathbf{s}^o|| < \epsilon$, and 0 otherwise.

1. sample proposal $\theta^* \sim q(\theta^*|\theta^\#)$
2. plug $\theta^* \to \mathcal{M}(\theta^*) \to \mathbf{y}^* \to \mathbf{s}^*$
3. compute acceptance ratio:

$$\text{ratio} := \frac{\mathbb{I}_\epsilon(\mathbf{s}^*, \mathbf{s}^o)\pi(\theta^*)}{\mathbb{I}_\epsilon(\mathbf{s}^\#, \mathbf{s}^o)\pi(\theta^\#)} \times \frac{q(\theta^\#|\theta^*)}{q(\theta^*|\theta^\#)}$$

4. simulate $u \sim U(0,1)$, accept θ^* if $u <$ ratio.

The previous algorithm produces dependent samples from the "augmented" posterior $\pi_\epsilon(\theta, \mathbf{s}^*|\mathbf{s}^o)$.

This means that if we disregard $\mathbf{s}^*$, and retain only θ^*, we have

$$\theta^* \sim \pi_\epsilon(\theta|\mathbf{s}^o)$$

This is just another way to sample from an approximated posterior. Using a more informed proposal function than the prior.

The previous algorithm produces dependent samples from the "augmented" posterior $\pi_\epsilon(\theta, \mathbf{s}^*|\mathbf{s}^o)$.

This means that if we disregard $\mathbf{s}^*$, and retain only θ^*, we have

$$\theta^* \sim \pi_\epsilon(\theta|\mathbf{s}^o)$$

This is just another way to sample from an approximated posterior. Using a more informed proposal function than the prior.

Example: stochastic Ricker model

$$\begin{cases} \textbf{(observations):}\ y_t \sim \text{Poisson}(\phi N_t) \\ \textbf{(unobservable process):}\ N_t = r \cdot N_{t-1} \cdot e^{-N_{t-1}+e_t}, \quad e_t \sim_{iid} \mathcal{N}(0, \sigma^2) \end{cases}$$

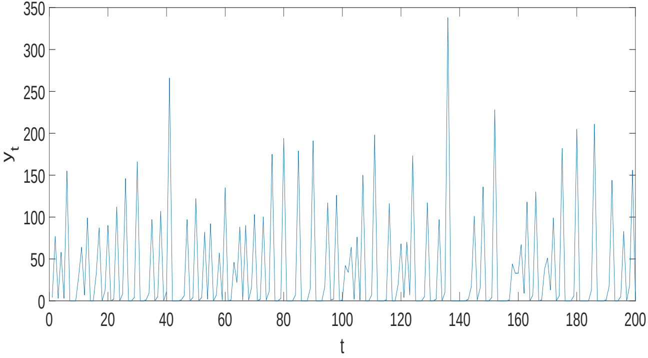
It can be used to describe the evolution in time of a population of size N_t.

- r is the intrinsic growth rate of the population;
- ϕ is a scale parameter
- e_t interpreted as environmental noise.

This is a hidden Markov model, as the dynamics of $\{N_t\}$ are Markovian and we assume measurements $y_{1:T}$ to be conditionally independent given $\{N_t\}$.

the data

We simulated 200 time points from the model, with $\log r = 3.8$, $\log \phi = 2.3$, $\log \sigma = -1.2$



Summary statistics

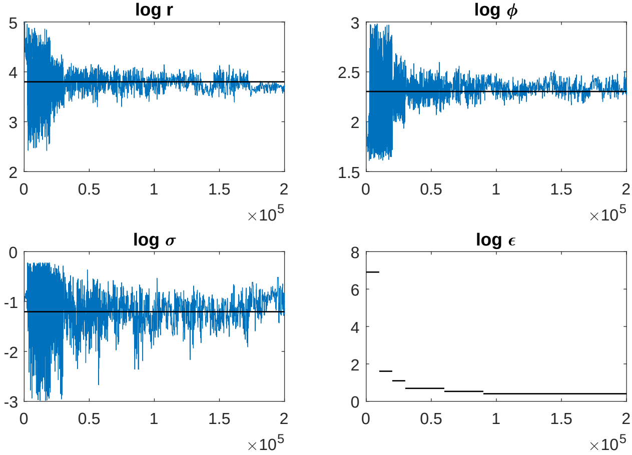
We used the 13 summary statistics suggested in Wood 2010. These include:

- the sample mean of observations $\bar{y}$;
- number of zeros in the dataset;
- autocovariances up to lag 5;
- and six more summaries...(not important to be mentioned here, see the reference above).

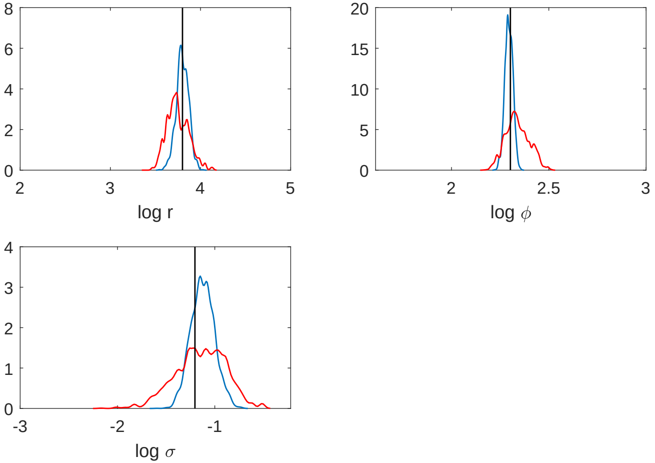
So we have $s(\mathbf{y}^o) = (\bar{y}, \#\text{zeros}, \text{autocov lag1}, ..., \text{autocov lag5}, ...)$.

ABC-MCMC traces

We performed 200,000 ABC-MCMC iterations with decreasing ϵ_t.



Using “particle marginal methods” we can obtain **exact inference** (blue) and compare it to ABC-MCMC (red)



Summary

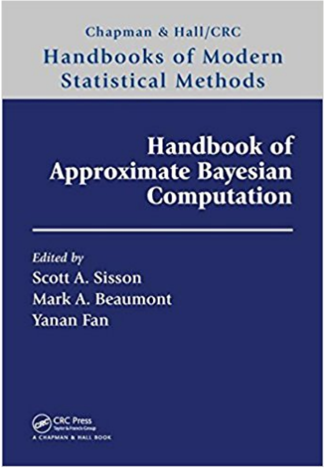
- ABC allows you to produce approximate inference for models having an intractable/unknown likelihood function.
- in our examples **we never needed to know the likelihood**;
- ABC is completely plug-and-play: it only requires the ability to computer-simulate artificial data.
- the main difficulty is to specify “informative” summary statistics.
- when summaries are not informative and ϵ is too large results will be poor.
- tuning ABC is not straightforward. Many available resources though...

Summary

- ABC allows you to produce approximate inference for models having an intractable/unknown likelihood function.
- in our examples **we never needed to know the likelihood**;
- ABC is completely plug-and-play: it only requires the ability to computer-simulate artificial data.
- the main difficulty is to specify "informative" summary statistics.
- when summaries are not informative and ϵ is too large results will be poor.
- tuning ABC is not straightforward. Many available resources though...

ABC book

Something to look forward to (promised for November 2018)


Chapman & Hall/CRC

Handbooks of Modern Statistical Methods

Handbook of Approximate Bayesian Computation

Edited by

Scott A. Sisson

Mark A. Beaumont

Yanan Fan

CRC Press

A CHAPMAN & HALL BOOK

Software (coloured links are clickable)

- EasyABC, R package. Research article.
- abc, R package. Research article
- abctools, R package. Research article. Focusses on tuning.
- A list with more options here .
- examples with implemented model simulators (useful to incorporate in your programs).

Reviews

Fairly extensive but accessible reviews:

1. Sisson and Fan 2010
2. (with applications in ecology) Beaumont 2010
3. Marin et al. 2010

Simpler introductions:

1. Sunnåker et al. 2013
2. (with applications in ecology) Hartig et al. 2013

Review specific for dynamical models:

1. Jasra 2015

Determination of summary statistics

1. review paper by Blum et al. 2013 on dimension reduction methods for ABC;
2. Fearnhed and Prangle 2012 (a JRSS-B discussion paper).

Blog posts and slides

1. Christian P. Robert often blogs about ABC (and beyond: it's a fantastic blog!)
2. an intro to ABC by Darren J. Wilkinson
3. Two posts by Rasmus Bååth here and here
4. Tons of slides at Slideshare.