Approximate Bayesian Computations (ABC): advances & questions (A&Q)

Christian P. Robert

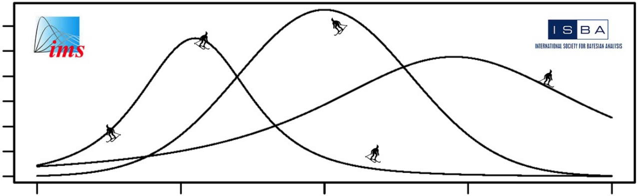
Université Paris-Dauphine, IuF, & CREST

Advertisement

MCMSki IV to be held in Chamonix M^{t} Blanc, France, from Monday, Jan. 6 to Wed., Jan. 8, 2014

All aspects of MCMC^{++} theory and methodology

Parallel (invited and contributed) sessions: call for proposals on website http://www.pages.drexel.edu/ mwl25/mcmski/



Econ'ections

Usages of simulation in Econometrics

Similar exploration of simulation-based techniques in Econometrics

- Simulated method of moments
- Method of simulated moments
- Simulated pseudo-maximum-likelihood
- Indirect inference

[Gouriéroux & Monfort, 1996]

Simulated method of moments

Given observations $y^o_{1:n}$ from a model

$$y_t = r(y_{1:(t-1)}, \epsilon_t, \theta) \,, \quad \epsilon_t \sim g(\cdot)$$

simulate $\epsilon^\star_{1:n}$, derive

$$y^\star_t(\theta) = r(y_{1:(t-1)}, \epsilon^\star_t, \theta)$$

and estimate θ by

$$\arg\min_\theta \sum_{t=1}^n (y^o_t - y^\star_t(\theta))^2$$

[Pakes & Pollard, 1989]

Simulated method of moments

Given observations $y^o_{1:n}$ from a model

$$y_t = r(y_{1:(t-1)}, \epsilon_t, \theta)\,, \quad \epsilon_t \sim g(\cdot)$$

simulate $\epsilon^\star_{1:n}$, derive

$$y^\star_t(\theta) = r(y_{1:(t-1)}, \epsilon^\star_t, \theta)$$

and estimate θ by

$$\arg\min_\theta \left\{ \sum_{t=1}^n y^o_t - \sum_{t=1}^n y^\star_t(\theta) \right\}^2$$

[Pakes & Pollard, 1989]

Indirect inference

Minimise (in θ) the distance between estimators $\hat{\beta}$ based on pseudo-models for genuine observations and for observations simulated under the true model and the parameter θ.

[Gouriéroux, Monfort, & Renault, 1993; Smith, 1993; Gallant & Tauchen, 1996]

Indirect inference

Example of the pseudo-maximum-likelihood (PML)

$$\hat{\beta}(\mathbf{y}) = \arg\max_{\beta} \sum_t \log f^\star(y_t|\beta, y_{1:(t-1)})$$

leading to

$$\arg\min_{\theta} ||\hat{\beta}(\mathbf{y}^o) - \hat{\beta}(\mathbf{y}_1(\theta), \ldots, \mathbf{y}_S(\theta))||^2$$

when

$$\mathbf{y}_s(\theta) \sim f(\mathbf{y}|\theta) \qquad s = 1, \ldots, S$$

Consistent indirect inference

> *...in order to get a unique solution the dimension of the auxiliary parameter β must be larger than or equal to the dimension of the initial parameter θ. If the problem is just identified the different methods become easier...*

Consistent indirect inference

> *...in order to get a unique solution the dimension of the auxiliary parameter β must be larger than or equal to the dimension of the initial parameter θ. If the problem is just identified the different methods become easier...*

Consistency depending on the criterion and on the asymptotic identifiability of θ

[Gouriéroux, Monfort, 1996, p. 66]

Consistent indirect inference

> *...in order to get a unique solution the dimension of the auxiliary parameter β must be larger than or equal to the dimension of the initial parameter θ. If the problem is just identified the different methods become easier...*

Consistency depending on the criterion and on the asymptotic identifiability of θ

[Gouriéroux, Monfort, 1996, p. 66]

© Indirect inference provides estimates rather than global inference...

Genetics of ABC

Genetic background of ABC

ABC is a recent computational technique that only requires being able to sample from the likelihood $f(\cdot|\theta)$

This technique stemmed from population genetics models, about 15 years ago, and population geneticists still contribute significantly to methodological developments of ABC.

[Griffith & al., 1997; Tavaré & al., 1999]

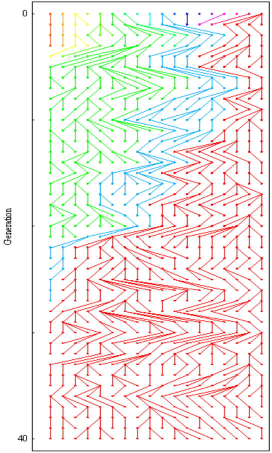
Population genetics

[Part derived from the teaching material of Raphael Leblois, ENS Lyon, November 2010]

- Describe the genotypes, estimate the alleles frequencies, determine their distribution among individuals, populations and between populations;
- Predict and understand the evolution of gene frequencies in populations as a result of various factors.

© Analyses the effect of various evolutive forces (mutation, drift, migration, selection) on the evolution of gene frequencies in time and space.

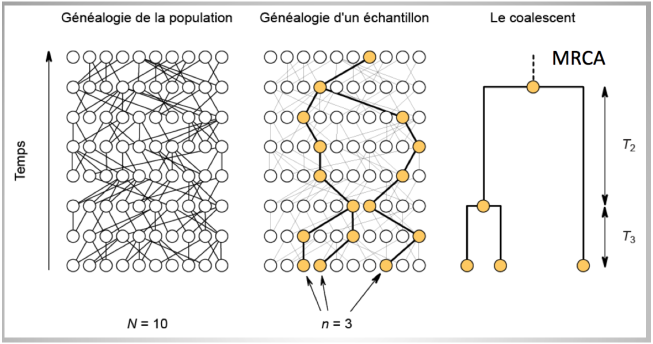
Wright-Fisher model



- A population of constant size, in which individuals reproduce at the same time.
- Each gene in a generation is a copy of a gene of the previous generation.
- In the absence of mutation and selection, allele frequencies derive inevitably until the fixation of an allele.

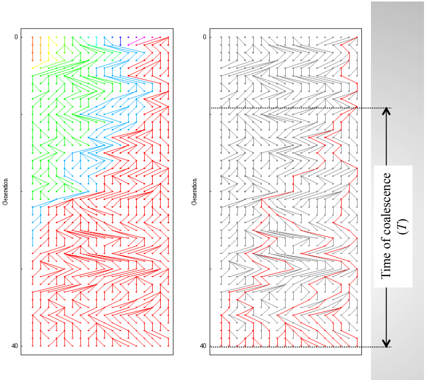
Coalescent theory

[Kingman, 1982; Tajima, Tavaré, &tc]



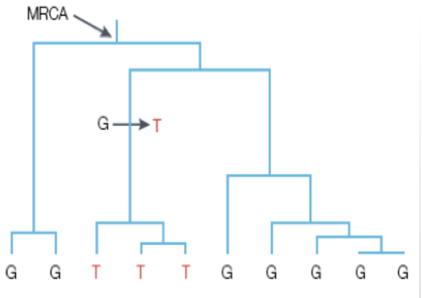
Coalescence theory interested in the genealogy of a sample of genes back in time to the common ancestor of the sample.

Common ancestor



The different lineages merge when we go back in the past.

Neutral mutations



- Under the assumption of neutrality, the mutations are independent of the genealogy.
- We construct the genealogy according to the demographic parameters, then we add a posteriori the mutations.

Demo-genetic inference

Each model is characterized by a set of parameters $\boldsymbol{\theta}$ that cover historical (time divergence, admixture time ...), demographics (population sizes, admixture rates, migration rates, ...) and genetic (mutation rate, ...) factors

The goal is to estimate these parameters from a dataset of polymorphism (DNA sample) $\mathbf{y}$ observed at the present time

Problem: most of the time, we can not calculate the likelihood of the polymorphism data $f(\mathbf{y}|\boldsymbol{\theta})$.

Untractable likelihood

Missing (too missing!) data structure:

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int_G f(\mathbf{y}|G, \boldsymbol{\theta}) f(G|\boldsymbol{\theta}) \mathrm{d}G$$

The genealogies are considered as nuisance parameters.

Warnin: problematic differs from the phylogenetic approach where the tree is the parameter of interesst.

ABC basics

simulation-based methods in Econometrics

Genetics of ABC

ABC basics

Alphabet soup

Summary statistic selection

Untractable likelihoods

Cases when the likelihood function $f(\mathbf{y}|\theta)$ is unavailable and when the completion step

$$f(\mathbf{y}|\theta) = \int_{\mathscr{Z}} f(\mathbf{y}, \mathbf{z}|\theta)\,\mathrm{d}\mathbf{z}$$

is impossible or too costly because of the dimension of $\mathbf{z}$
© MCMC cannot be implemented!

Untractable likelihoods

Cases when the likelihood function $f(\mathbf{y}|\theta)$ is unavailable and when the completion step

$$f(\mathbf{y}|\theta) = \int_{\mathscr{Z}} f(\mathbf{y}, \mathbf{z}|\theta)\, \mathrm{d}\mathbf{z}$$

is impossible or too costly because of the dimension of $\mathbf{z}$
© MCMC cannot be implemented!

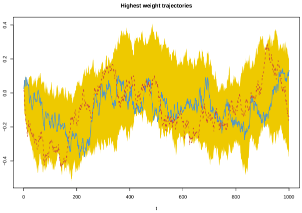
Illustrations

Example ()

Stochastic volatility model: for $t = 1, \dots, T$,

$$y_t = \exp(z_t)\epsilon_t \,, \; z_t = a + bz_{t-1} + \sigma\eta_t \,,$$

T very large makes it difficult to include $\mathbf{z}$ within the simulated parameters



Illustrations

Example ()

Potts model: if $\mathbf{y}$ takes values on a grid $\mathfrak{Y}$ of size k^n and

$$f(\mathbf{y}|\theta) \propto \exp\left\{\theta \sum_{l \sim i} \mathbb{I}_{y_l = y_i}\right\}$$

where $l \sim i$ denotes a neighbourhood relation, n moderately large prohibits the computation of the normalising constant

Illustrations

Example (Genesis)

The ABC method

Bayesian setting: target is $\pi(\theta)f(x|\theta)$

The ABC method

Bayesian setting: target is $\pi(\theta)f(x|\theta)$
When likelihood $f(x|\theta)$ not in closed form, likelihood-free rejection technique:

The ABC method

Bayesian setting: target is $\pi(\theta)f(x|\theta)$
When likelihood $f(x|\theta)$ not in closed form, likelihood-free rejection technique:

ABC algorithm

For an observation $\mathbf{y} \sim f(\mathbf{y}|\theta)$, under the prior $\pi(\theta)$, keep *jointly* simulating

$$\theta' \sim \pi(\theta)\,, \mathbf{z} \sim f(\mathbf{z}|\theta')\,,$$

until the auxiliary variable $\mathbf{z}$ is equal to the observed value, $\mathbf{z} = \mathbf{y}$.

[Tavaré et al., 1997]

Why does it work?!

The proof is trivial:

$$\begin{aligned} f(\theta_i) &\propto \sum_{\mathbf{z}\in\mathcal{D}} \pi(\theta_i) f(\mathbf{z}|\theta_i) \mathbb{I}_{\mathbf{y}}(\mathbf{z}) \\ &\propto \pi(\theta_i) f(\mathbf{y}|\theta_i) \\ &= \pi(\theta_i|\mathbf{y})\,. \end{aligned}$$

[Accept–Reject 101]

A as A...pproximative

When y is a continuous random variable, equality $\mathbf{z} = \mathbf{y}$ is replaced with a tolerance condition,

$$\varrho(\mathbf{y}, \mathbf{z}) \leq \epsilon$$

where ϱ is a distance

A as A...pproximative

When y is a continuous random variable, equality $\mathbf{z} = \mathbf{y}$ is replaced with a tolerance condition,

$$\varrho(\mathbf{y}, \mathbf{z}) \leq \epsilon$$

where ϱ is a distance
Output distributed from

$$\pi(\theta)\, P_\theta\{\varrho(\mathbf{y}, \mathbf{z}) < \epsilon\} \propto \pi(\theta | \varrho(\mathbf{y}, \mathbf{z}) < \epsilon)$$

[Pritchard et al., 1999]

ABC algorithm

Algorithm 1 Likelihood-free rejection sampler 2

for $i = 1$ to N **do**
  repeat
    generate θ' from the prior distribution $\pi(\cdot)$
    generate $\mathbf{z}$ from the likelihood $f(\cdot|\theta')$
  until $\rho\{\eta(\mathbf{z}), \eta(\mathbf{y})\} \leq \epsilon$
  set $\theta_i = \theta'$
end for

where $\eta(\mathbf{y})$ defines a (not necessarily sufficient) statistic

Output

The likelihood-free algorithm samples from the marginal in $\mathbf{z}$ of:

$$\pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) = \frac{\pi(\theta) f(\mathbf{z}|\theta) \mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z})}{\int_{A_{\epsilon,\mathbf{y}} \times \Theta} \pi(\theta) f(\mathbf{z}|\theta) \mathrm{d}\mathbf{z} \mathrm{d}\theta},$$

where $A_{\epsilon,\mathbf{y}} = \{\mathbf{z} \in \mathcal{D} | \rho(\eta(\mathbf{z}), \eta(\mathbf{y})) < \epsilon\}$.

Output

The likelihood-free algorithm samples from the marginal in $\mathbf{z}$ of:

$$\pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) = \frac{\pi(\theta) f(\mathbf{z}|\theta) \mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z})}{\int_{A_{\epsilon,\mathbf{y}} \times \Theta} \pi(\theta) f(\mathbf{z}|\theta) \mathrm{d}\mathbf{z} \mathrm{d}\theta},$$

where $A_{\epsilon,\mathbf{y}} = \{\mathbf{z} \in \mathcal{D} | \rho(\eta(\mathbf{z}), \eta(\mathbf{y})) < \epsilon\}$.

The idea behind ABC is that the summary statistics coupled with a small tolerance should provide a good approximation of the posterior distribution:

$$\pi_\epsilon(\theta|\mathbf{y}) = \int \pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) \mathrm{d}\mathbf{z} \approx \pi(\theta|\eta(\mathbf{y})).$$

MA example

MA(2) model

$$x_t = \epsilon_t + \sum_{i=1}^{2} \vartheta_i \epsilon_{t-i}$$

Simple prior: uniform prior over the identifiability zone, e.g. triangle for MA(2)

MA example (2)

ABC algorithm thus made of

1. picking a new value $(\vartheta_1, \vartheta_2)$ in the triangle
2. generating an iid sequence $(\epsilon_t)_{-q<t\leq T}$
3. producing a simulated series $(x'_t)_{1\leq t\leq T}$

MA example (2)

ABC algorithm thus made of

1. picking a new value $(\vartheta_1, \vartheta_2)$ in the triangle
2. generating an iid sequence $(\epsilon_t)_{-q<t\leq T}$
3. producing a simulated series $(x'_t)_{1\leq t\leq T}$

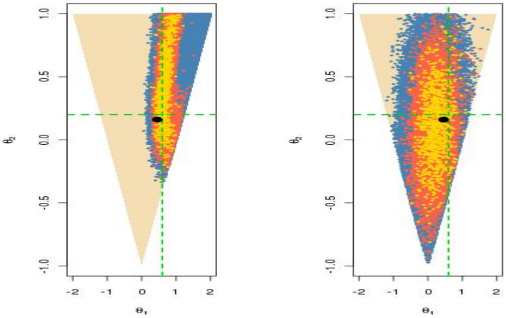
Distance: basic distance between the series

$$\rho((x'_t)_{1\leq t\leq T}, (x_t)_{1\leq t\leq T}) = \sum_{t=1}^{T} (x_t - x'_t)^2$$

or distance between summary statistics like the q autocorrelations

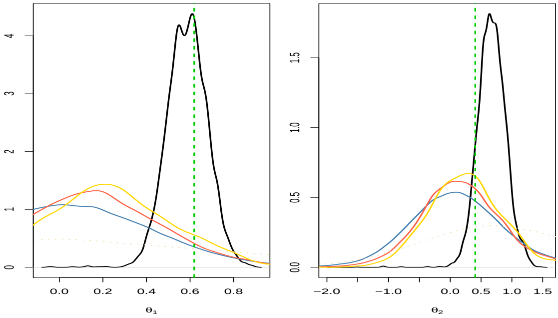
$$\tau_j = \sum_{t=j+1}^{T} x_t x_{t-j}$$

Comparison of distance impact



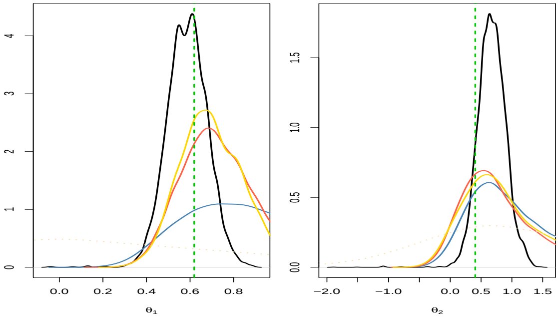
Evaluation of the tolerance on the ABC sample against both distances ($\epsilon = 100\%, 10\%, 1\%, 0.1\%$) for an MA(2) model

Comparison of distance impact



Evaluation of the tolerance on the ABC sample against both distances ($\epsilon = 100\%, 10\%, 1\%, 0.1\%$) for an MA(2) model

Comparison of distance impact



Evaluation of the tolerance on the ABC sample against both distances ($\epsilon = 100\%, 10\%, 1\%, 0.1\%$) for an MA(2) model

ABC advances

Simulating from the prior is often poor in efficiency

ABC advances

Simulating from the prior is often poor in efficiency
Either modify the proposal distribution on θ to increase the density of x's within the vicinity of y...

[Marjoram et al, 2003; Bortot et al., 2007, Sisson et al., 2007]

ABC advances

Simulating from the prior is often poor in efficiency
Either modify the proposal distribution on θ to increase the density of x's within the vicinity of y...

[Marjoram et al, 2003; Bortot et al., 2007, Sisson et al., 2007]

...or by viewing the problem as a conditional density estimation and by developing techniques to allow for larger ϵ

[Beaumont et al., 2002]

ABC advances

Simulating from the prior is often poor in efficiency
Either modify the proposal distribution on θ to increase the density of x's within the vicinity of y...

[Marjoram et al, 2003; Bortot et al., 2007, Sisson et al., 2007]

...or by viewing the problem as a conditional density estimation and by developing techniques to allow for larger ϵ

[Beaumont et al., 2002]

.....or even by including ϵ in the inferential framework [ABC$_\mu$]

[Ratmann et al., 2009]

Alphabet soup

simulation-based methods in Econometrics

Genetics of ABC

ABC basics

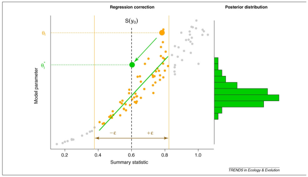
Alphabet soup

Summary statistic selection

ABC-NP

Better usage of [prior] simulations by adjustement: instead of throwing away θ' such that $\rho(\eta(\mathbf{z}),\eta(\mathbf{y})) > \epsilon$, replace θ's with locally regressed transforms

$$\theta^* = \theta - \{\eta(\mathbf{z}) - \eta(\mathbf{y})\}^\mathsf{T} \hat{\beta}$$



[Csilléry et al., TEE, 2010]

where $\hat{\beta}$ is obtained by [NP] weighted least square regression on $(\eta(\mathbf{z}) - \eta(\mathbf{y}))$ with weights

$$K_\delta \left\{\rho(\eta(\mathbf{z}),\eta(\mathbf{y}))\right\}$$

[Beaumont et al., 2002, Genetics]

ABC-NP (regression)

Also found in the subsequent literature, e.g. in Fearnhead-Prangle (2012): weight directly simulation by

$$K_\delta\left\{\rho(\eta(\mathbf{z}(\theta)),\eta(\mathbf{y}))\right\}$$

or

$$\frac{1}{S}\sum_{s=1}^{S}K_\delta\left\{\rho(\eta(\mathbf{z}^s(\theta)),\eta(\mathbf{y}))\right\}$$

[consistent estimate of $f(\eta|\theta)$]

ABC-NP (regression)

Also found in the subsequent literature, e.g. in ▸ Fearnhead-Prangle (2012) : weight directly simulation by

$$K_\delta \left\{ \rho(\eta(\mathbf{z}(\theta)), \eta(\mathbf{y})) \right\}$$

or

$$\frac{1}{S} \sum_{s=1}^{S} K_\delta \left\{ \rho(\eta(\mathbf{z}^s(\theta)), \eta(\mathbf{y})) \right\}$$

[consistent estimate of $f(\eta|\theta)$]

Curse of dimensionality: poor estimate when $d = \dim(\eta)$ is large...

ABC-NP (density estimation)

Use of the kernel weights

$$K_\delta \left\{\rho(\eta(\mathbf{z}(\theta)), \eta(\mathbf{y}))\right\}$$

leads to the NP estimate of the posterior conditional density

$$\frac{\sum_i \tilde{K}_b(\theta_i - \theta) K_\delta \left\{\rho(\eta(\mathbf{z}(\theta_i)), \eta(\mathbf{y}))\right\}}{\sum_i K_\delta \left\{\rho(\eta(\mathbf{z}(\theta_i)), \eta(\mathbf{y}))\right\}}$$

[Blum, JASA, 2010]

ABC-MCMC

Markov chain $(\theta^{(t)})$ created via the transition function

$$\theta^{(t+1)} = \begin{cases} \theta' \sim K_\omega(\theta'|\theta^{(t)}) & \text{if } x \sim f(x|\theta') \text{ is such that } x = y \\ & \text{ and } u \sim \mathcal{U}(0,1) \leq \frac{\pi(\theta')K_\omega(\theta^{(t)}|\theta')}{\pi(\theta^{(t)})K_\omega(\theta'|\theta^{(t)})}, \\ \theta^{(t)} & \text{otherwise,} \end{cases}$$

ABC-MCMC

Markov chain $(\theta^{(t)})$ created via the transition function

$$\theta^{(t+1)} = \begin{cases} \theta' \sim K_\omega(\theta'|\theta^{(t)}) & \text{if } x \sim f(x|\theta') \text{ is such that } x = y \\ & \text{and } u \sim \mathcal{U}(0,1) \leq \frac{\pi(\theta')K_\omega(\theta^{(t)}|\theta')}{\pi(\theta^{(t)})K_\omega(\theta'|\theta^{(t)})}, \\ \theta^{(t)} & \text{otherwise,} \end{cases}$$

has the posterior $\pi(\theta|y)$ as stationary distribution

[Marjoram et al, 2003]

ABC-PMC

Another sequential version producing a sequence of Markov transition kernels K_t and of samples $(\theta_1^{(t)}, \ldots, \theta_N^{(t)})$ $(1 \leq t \leq T)$
Generate a sample at iteration t by

$$\hat{\pi}_t(\theta^{(t)}) \propto \sum_{j=1}^{N} \omega_j^{(t-1)} K_t(\theta^{(t)}|\theta_j^{(t-1)})$$

modulo acceptance of the associated x_t, and use an importance weight associated with an accepted simulation $\theta_i^{(t)}$

$$\omega_i^{(t)} \propto \pi(\theta_i^{(t)})/\hat{\pi}_t(\theta_i^{(t)}) \,.$$

© **Still likelihood free**

[Beaumont et al., 2009]

Sequential Monte Carlo

SMC is a simulation technique to approximate a sequence of related probability distributions π_n with π_0 "easy" and π_T as target.
Iterated IS as PMC: particles moved from time n to time n via kernels K_n and use of a sequence of extended targets $\tilde{\pi}_n$

$$\tilde{\pi}_n(\mathbf{z}_{0:n}) = \pi_n(z_n) \prod_{j=0}^{n} L_j(z_{j+1}, z_j)$$

where the L_j's are backward Markov kernels [check that $\pi_n(z_n)$ is a marginal]

[Del Moral, Doucet & Jasra, Series B, 2006]

ABC-SMC

True derivation of an SMC-ABC algorithm
Use of a kernel K_n associated with target π_{ϵ_n} and derivation of the backward kernel

$$L_{n-1}(z, z') = \frac{\pi_{\epsilon_n}(z')K_n(z', z)}{\pi_n(z)}$$

Update of the weights

$$w_{in} \propto w_{i(n-1)} \frac{\sum_{m=1}^{M} \mathbb{I}_{A_{\epsilon_n}}(x_{in}^m)}{\sum_{m=1}^{M} \mathbb{I}_{A_{\epsilon_{n-1}}}(x_{i(n-1)}^m)}$$

when $x_{in}^m \sim K(x_{i(n-1)}, \cdot)$

[Del Moral, Doucet & Jasra, 2009]

Properties of ABC-SMC

The ABC-SMC method properly uses a backward kernel $L(z, z')$ to simplify the importance weight and to remove the dependence on the unknown likelihood from this weight. Update of importance weights is reduced to the ratio of the proportions of surviving particles

Major assumption: the forward kernel K is supposed to be invariant against the true target [tempered version of the true posterior]

Properties of ABC-SMC

The ABC-SMC method properly uses a backward kernel $L(z, z')$ to simplify the importance weight and to remove the dependence on the unknown likelihood from this weight. Update of importance weights is reduced to the ratio of the proportions of surviving particles

Major assumption: the forward kernel K is supposed to be invariant against the true target [tempered version of the true posterior]

Adaptivity in ABC-SMC algorithm only found in on-line construction of the thresholds ϵ_t, slowly enough to keep a large number of accepted transitions

Wilkinson's exact BC

ABC approximation error (i.e. non-zero tolerance) replaced with exact simulation from a controlled approximation to the target, convolution of true posterior with kernel function

$$\pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) = \frac{\pi(\theta) f(\mathbf{z}|\theta) K_\epsilon(\mathbf{y} - \mathbf{z})}{\int \pi(\theta) f(\mathbf{z}|\theta) K_\epsilon(\mathbf{y} - \mathbf{z}) \mathrm{d}\mathbf{z}\mathrm{d}\theta},$$

with K_ϵ kernel parameterised by bandwidth ϵ.

[Wilkinson, 2008]

Wilkinson's exact BC

ABC approximation error (i.e. non-zero tolerance) replaced with exact simulation from a controlled approximation to the target, convolution of true posterior with kernel function

$$\pi_\epsilon(\theta, \mathbf{z}|\mathbf{y}) = \frac{\pi(\theta) f(\mathbf{z}|\theta) K_\epsilon(\mathbf{y}-\mathbf{z})}{\int \pi(\theta) f(\mathbf{z}|\theta) K_\epsilon(\mathbf{y}-\mathbf{z}) \mathrm{d}\mathbf{z}\mathrm{d}\theta},$$

with K_ϵ kernel parameterised by bandwidth ϵ.

[Wilkinson, 2008]

Theorem

The ABC algorithm based on the assumption of a randomised observation $\mathbf{y} = \tilde{\mathbf{y}} + \xi$, $\xi \sim K_\epsilon$, *and an acceptance probability of*

$$K_\epsilon(\mathbf{y}-\mathbf{z})/M$$

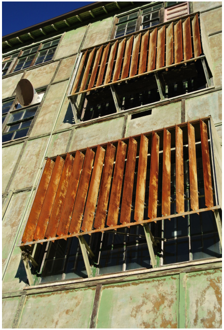
gives draws from the posterior distribution $\pi(\theta|\mathbf{y})$.

Consistent noisy ABC-MLE

- Degrading the data improves the estimation performances:
  - Noisy ABC-MLE is asymptotically (in n) consistent
  - under further assumptions, the noisy ABC-MLE is asymptotically normal
  - increase in variance of order ϵ^{-2}
- likely degradation in precision or computing time due to the lack of summary statistic [curse of dimensionality]

[Jasra, Singh, Martin, & McCoy, 2010]

Summary statistic selection



Semi-automatic ABC

Fearnhead and Prangle (2010) study ABC and the selection of the summary statistic in close proximity to Wilkinson's proposal

ABC then considered from a purely inferential viewpoint and calibrated for estimation purposes

Use of a randomised (or 'noisy') version of the summary statistics

$$\tilde{\eta}(\mathbf{y}) = \eta(\mathbf{y}) + \tau\epsilon$$

Summary statistics

- Optimality of the posterior expectation $\mathbb{E}[\theta|\mathbf{y}]$ of the parameter of interest as summary statistics $\eta(\mathbf{y})$!

Summary statistics

- Optimality of the posterior expectation $\mathbb{E}[\theta|\mathbf{y}]$ of the parameter of interest as summary statistics $\eta(\mathbf{y})$!
- Use of the standard quadratic loss function

$$(\theta - \theta_0)^\mathsf{T} A(\theta - \theta_0)\,.$$

Details on Fearnhead and Prangle (F&P) ABC

Use of a summary statistic $S(\cdot)$, an importance proposal $g(\cdot)$, a kernel $K(\cdot) \leq 1$ and a bandwidth $h > 0$ such that

$$(\theta, \mathbf{y}_{\text{sim}}) \sim g(\theta) f(\mathbf{y}_{\text{sim}}|\theta)$$

is accepted with probability (hence the bound)

$$K[\{S(\mathbf{y}_{\text{sim}}) - \mathbf{s}_{\text{obs}}\}/h]$$

and the corresponding importance weight defined by

$$\pi(\theta)/g(\theta)$$

[Fearnhead & Prangle, 2012]

Errors, errors, and errors

Three levels of approximation

- $\pi(\theta|\mathbf{y}_{\text{obs}})$ by $\pi(\theta|\mathbf{s}_{\text{obs}})$ loss of information

  [ignored]

- $\pi(\theta|\mathbf{s}_{\text{obs}})$ by

$$\pi_{\text{ABC}}(\theta|\mathbf{s}_{\text{obs}}) = \frac{\int \pi(\mathbf{s})K[\{\mathbf{s}-\mathbf{s}_{\text{obs}}\}/h]\pi(\theta|\mathbf{s})\,\text{d}\mathbf{s}}{\int \pi(\mathbf{s})K[\{\mathbf{s}-\mathbf{s}_{\text{obs}}\}/h]\,\text{d}\mathbf{s}}$$

  noisy observations

- $\pi_{\text{ABC}}(\theta|\mathbf{s}_{\text{obs}})$ by importance Monte Carlo based on N simulations, represented by $\text{var}(a(\theta)|\mathbf{s}_{\text{obs}})/N_{\text{acc}}$ [expected number of acceptances]

Optimal summary statistic

> *"We take a different approach, and weaken the requirement for π_{ABC} to be a good approximation to $\pi(\theta|\mathbf{y}_{obs})$. We argue for π_{ABC} to be a good approximation solely in terms of the accuracy of certain estimates of the parameters." (F&P, p.5)*

From this result, F&P

- derive their choice of summary statistic,

$$S(\mathbf{y}) = \mathbb{E}(\theta|\mathbf{y})$$

[almost sufficient]

- suggest

$$h = O(N^{-1/(2+d)}) \quad \text{and} \quad h = O(N^{-1/(4+d)})$$

as optimal bandwidths for noisy and standard ABC.

Optimal summary statistic

> *"We take a different approach, and weaken the requirement for π_{ABC} to be a good approximation to $\pi(\theta|\mathbf{y}_{obs})$. We argue for π_{ABC} to be a good approximation solely in terms of the accuracy of certain estimates of the parameters." (F&P, p.5)*

From this result, F&P

- derive their choice of summary statistic,

$$S(\mathbf{y}) = \mathbb{E}(\theta|\mathbf{y})$$

[wow! $\mathbb{E}_{\cancel{ABC}}[\theta|S(\mathbf{y}_{\text{obs}})] = \mathbb{E}[\theta|\mathbf{y}_{\text{obs}}]$]

- suggest

$$h = O(N^{-1/(2+d)}) \quad \text{and} \quad h = O(N^{-1/(4+d)})$$

as optimal bandwidths for noisy and standard ABC.

Caveat

Since $\mathbb{E}(\theta|\mathbf{y}_{\text{obs}})$ is most usually unavailable, F&P suggest

(i) use a pilot run of ABC to determine a region of non-negligible posterior mass;

(ii) simulate sets of parameter values and data;

(iii) use the simulated sets of parameter values and data to estimate the summary statistic; and

(iv) run ABC with this choice of summary statistic.

How much Bayesian is aBc..?

- maybe another convergent type of inference (meaningful? sufficient?)
- approximation error unknown (w/o simulation)
- pragmatic Bayes (there is no other solution!)
- noisy Bayes (exact with dirty data)
- exhibition of pseudo-sufficient statistics (coherent? constructive?)

How much Bayesian is aBc..?

- maybe another convergent type of inference (meaningful? sufficient?)
- approximation error unknown (w/o simulation)
- pragmatic Bayes (there is no other solution!)
- noisy Bayes (exact with dirty data)
- exhibition of pseudo-sufficient statistics (coherent? constructive?)