Asymptotics of ABC

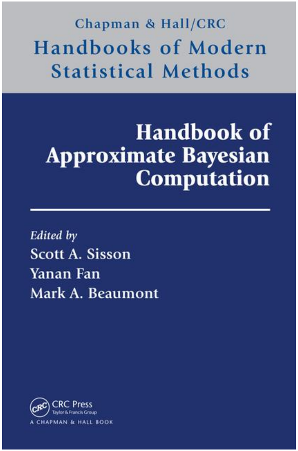
Christian P. Robert
U. Paris Dauphine PSL & Warwick U. & CREST
ENSAE

IAM–PIMS Colloquium, 9 Dec. 2019

Co-authors

- Jean-Marie Cornuet, Arnaud Estoup (INRA Montpellier)
- David Frazier, Gael Martin (Monash U)
- Jean-Michel Marin (U Montpellier)
- Kerrie Mengersen (QUT)
- Espen Bernton, Pierre Jacob, Natesh Pillai (Harvard U)
- Judith Rousseau (U Oxford)
- Grégoire Clarté, Robin Ryder, Julien Stoehr (U Paris Dauphine)

Outline


Chapman & Hall/CRC
Handbooks of Modern Statistical Methods
Handbook of Approximate Bayesian Computation
Edited by
Scott A. Sisson
Yanan Fan
Mark A. Beaumont
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Posterior distribution

Central concept of Bayesian inference:

$$\pi(\underbrace{\theta}_{\text{parameter}} \mid \underbrace{x^{\text{obs}}}_{\text{observation}}) \overbrace{\propto}^{\text{proportional}} \underbrace{\pi(\theta)}_{\text{prior}} \times \underbrace{L(\theta \mid x^{\text{obs}})}_{\text{likelihood}}$$

Drives

- derivation of optimal decisions
- assessment of uncertainty
- model selection
- prediction

[McElreath, 2015]

Monte Carlo representation

Exploration of Bayesian posterior $\pi(\theta|x^{\text{obs}})$ may (!) require to produce sample

$$\theta_1, \ldots, \theta_T$$

distributed from $\pi(\theta|x^{\text{obs}})$ (or asymptotically by Markov chain Monte Carlo aka MCMC)

[McElreath, 2015]

Difficulties

MCMC = workhorse of practical Bayesian analysis (BUGS, JAGS, Stan, &tc.), except when product

$$\pi(\theta) \times L(\theta \mid x^{\text{obs}})$$

well-defined but numerically unavailable or too costly to compute

Only partial solutions are available:

- demarginalisation (latent variables)
- exchange algorithm (auxiliary variables)
- pseudo-marginal (unbiased estimator)

Difficulties

MCMC = workhorse of practical Bayesian analysis (BUGS, JAGS, Stan, &tc.), except when product

$$\pi(\theta) \times L(\theta \mid x^{\text{obs}})$$

well-defined but numerically unavailable or too costly to compute

Only partial solutions are available:

- demarginalisation (latent variables)
- exchange algorithm (auxiliary variables)
- pseudo-marginal (unbiased estimator)

Example 1: Dynamic mixture

Mixture model

$$\{1 - w_{\mu,\tau}(x)\}f_{\beta,\lambda}(x) + w_{\mu,\tau}(x)g_{\epsilon,\sigma}(x) \qquad x > 0$$

where

- $f_{\beta,\lambda}$ Weibull density,
- $g_{\epsilon,\sigma}$ generalised Pareto density, and
- $w_{\mu,\tau}$ Cauchy (arctan) cdf

Intractable normalising constant

$$\mathfrak{C}(\mu,\tau,\beta,\lambda,\epsilon,\sigma) = \int_0^\infty \{(1 - w_{\mu,\tau}(x))f_{\beta,\lambda}(x) + w_{\mu,\tau}(x)g_{\epsilon,\sigma}(x)\}\,dx$$

[Frigessi, Haug & Rue, 2002]

Example 2: truncated Normal

Given set $\mathcal{A} \subset \mathbb{R}^k$ (k large), truncated Normal model

$$f(x \mid \mu, \Sigma, \mathcal{A}) \propto \exp\{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2\}\, \mathbb{I}_{\mathcal{A}}(x)$$

with intractable normalising constant

$$\mathfrak{C}(\mu, \Sigma, \mathcal{A}) = \int_{\mathcal{A}} \exp\{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2\} dx$$

Example 3: robust Normal statistics

Normal sample

$$x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma^2)$$

summarised into (insufficient)

$$\hat{\mu}_n = \text{med}(x_1, \ldots, x_n)$$

and

$$\hat{\sigma}_n = \text{mad}(x_1, \ldots, x_n) = \text{med}\,|x_i - \hat{\mu}_n|$$

Under a conjugate prior $\pi(\mu, \sigma^2)$, posterior close to intractable.

but simulation of $(\hat{\mu}_n, \hat{\sigma}_n)$ straightforward

Example 3: robust Normal statistics

Normal sample

$$x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma^2)$$

summarised into (insufficient)

$$\hat{\mu}_n = \text{med}(x_1, \ldots, x_n)$$

and

$$\hat{\sigma}_n = \text{mad}(x_1, \ldots, x_n) = \text{med}\,|x_i - \hat{\mu}_n|$$

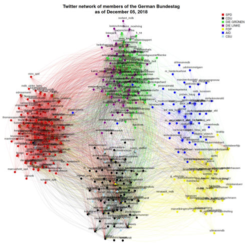
Under a conjugate prior $\pi(\mu, \sigma^2)$, posterior close to intractable.
but simulation of $(\hat{\mu}_n, \hat{\sigma}_n)$ straightforward

Example 4: exponential random graph

ERGM: binary random vector x indexed by all edges on set of nodes plus graph

$$f(x \mid \theta) = \frac{1}{\mathfrak{C}(\theta)} \exp(\theta^{\mathrm{T}} S(x))$$

with $S(x)$ vector of statistics and $\mathfrak{C}(\theta)$ intractable normalising constant



[Grelaud & al., 2009; Everitt, 2012; Bouranis & al., 2017]

Realistic[er] applications

- Kingman's coalescent in population genetics
  [Tavaré et al., 1997; Beaumont et al., 2003]
- α-stable distributions
  [Peters et al, 2012]
- complex networks
  [Dutta et al., 2018]
- astrostatistics & cosmostatistics
  [Cameron & Pettitt, 2012; Ishida et al., 2015]

Concept

A?B?C?

- **A** stands for approximate [wrong likelihood]
- **B** stands for Bayesian [right prior]
- **C** stands for computation [producing a parameter sample]

A?B?C?

- Rough version of the data [from dot to ball]
- Non-parametric approximation of the likelihood [near actual observation]
- Use of non-sufficient statistics [dimension reduction]
- Monte Carlo error [and no unbiasedness]

A seemingly naïve representation

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

ABC-AR algorithm

For an observation $x^{\text{obs}} \sim f(x|\theta)$, under the prior $\pi(\theta)$, keep *jointly* simulating

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

until the auxiliary variable z is equal to the observed value,

$$z = x^{\text{obs}}$$

[Diggle & Gratton, 1984; Rubin, 1984; Tavaré et al., 1997]

A seemingly naïve representation

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

ABC-AR algorithm

For an observation $x^{\text{obs}} \sim f(x|\theta)$, under the prior $\pi(\theta)$, keep *jointly* simulating

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

until the auxiliary variable z is equal to the observed value,

$$z = x^{\text{obs}}$$

[Diggle & Gratton, 1984; Rubin, 1984; Tavaré et al., 1997]

Why does it work?

The mathematical proof is trivial:

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

[Accept–Reject 101]

But very impractical when

$$\mathbb{P}_\theta(Z = x^{\text{obs}}) \approx 0$$

Why does it work?

The mathematical proof is trivial:

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

[Accept–Reject 101]

But very impractical when

$$\mathbb{P}_\theta(Z = x^{\text{obs}}) \approx 0$$

A as approximative

When y is a continuous random variable, strict equality

$$z = x^{\text{obs}}$$

is replaced with a tolerance zone

$$\rho(x^{\text{obs}}, z) \leq \epsilon$$

where ρ is a distance

Output distributed from

$$\pi(\theta)\, P_\theta\{\rho(x^{\text{obs}}, z) < \epsilon\} \overset{\text{def}}{\propto} \pi(\theta | \rho(x^{\text{obs}}, z) < \epsilon)$$

[Pritchard et al., 1999]

A as approximative

When y is a continuous random variable, strict equality

$$z = x^{\text{obs}}$$

is replaced with a tolerance zone

$$\rho(x^{\text{obs}}, z) \leq \epsilon$$

where ρ is a distance
Output distributed from

$$\pi(\theta)\, P_\theta\{\rho(x^{\text{obs}}, z) < \epsilon\} \overset{\text{def}}{\propto} \pi(\theta|\rho(x^{\text{obs}}, z) < \epsilon)$$

[Pritchard et al., 1999]

ABC algorithm

Algorithm 1 Likelihood-free rejection sampler

for $i = 1$ to N **do**
  repeat
    generate θ' from prior $\pi(\cdot)$
    generate z from sampling density $f(\cdot|\theta')$
  until $\rho\{\eta(z), \eta(x^{\text{obs}})\} \leq \epsilon$
  set $\theta_i = \theta'$
end for

where $\eta(x^{\text{obs}})$ defines a (not necessarily sufficient) statistic

Custom: $\eta(x^{\text{obs}})$ called summary statistic

ABC algorithm

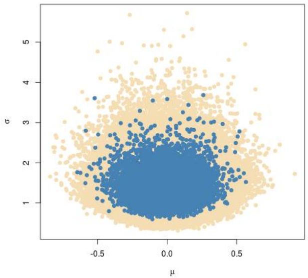
Algorithm 2 Likelihood-free rejection sampler

for $i = 1$ to N **do**
  repeat
    generate θ' from prior $\pi(\cdot)$
    generate z from sampling density $f(\cdot|\theta')$
  until $\rho\{\eta(z), \eta(x^{\text{obs}})\} \leq \epsilon$
  set $\theta_i = \theta'$
end for

where $\eta(x^{\text{obs}})$ defines a (not necessarily sufficient) statistic
Custom: $\eta(x^{\text{obs}})$ called summary statistic

Example 3: robust Normal statistics

```
mu=rnorm(N<-1e6)  #prior
sig=sqrt(rgamma(N,2,2))
medobs=median(obs)
madobs=mad(obs) #summary
for(t in diz<-1:N){
 psud=rnorm(1e2)/sig[t]+mu[t]
 medpsu=median(psud)-medobs
 madpsu=mad(psud)-madobs
 diz[t]=medpsu^2+madpsu^2}
#ABC subsample
subz=which(diz<quantile(diz,.1))
```



Exact ABC posterior

Algorithm samples from marginal in z of [exact] posterior

$$\pi_\varepsilon^{\text{ABC}}(\theta, z|x^{\text{obs}}) = \frac{\pi(\theta)f(z|\theta)\mathbb{I}_{A_{\varepsilon,x^{\text{obs}}}}(z)}{\int_{A_{\varepsilon,x^{\text{obs}}}\times\Theta}\pi(\theta)f(z|\theta)\text{d}z\text{d}\theta},$$

where $A_{\varepsilon,x^{\text{obs}}} = \{z \in \mathcal{D}|\rho\{\eta(z),\eta(x^{\text{obs}})\} < \varepsilon\}$.

Intuition that proper summary statistics coupled with small tolerance $\varepsilon = \varepsilon_\eta$ should provide good approximation of the posterior distribution:

$$\pi_\varepsilon^{\text{ABC}}(\theta|x^{\text{obs}}) = \int \pi_\varepsilon^{\text{ABC}}(\theta, z|x^{\text{obs}})\text{d}z \approx \pi\{\theta|\eta(x^{\text{obs}})\}$$

Exact ABC posterior

Algorithm samples from marginal in z of [exact] posterior

$$\pi_\varepsilon^{\text{ABC}}(\theta, z|x^{\text{obs}}) = \frac{\pi(\theta)f(z|\theta)\mathbb{I}_{A_{\varepsilon,x^{\text{obs}}}}(z)}{\int_{A_{\varepsilon,x^{\text{obs}}}\times\Theta}\pi(\theta)f(z|\theta)\text{d}z\text{d}\theta},$$

where $A_{\varepsilon,x^{\text{obs}}} = \{z \in \mathcal{D}|\rho\{\eta(z),\eta(x^{\text{obs}})\} < \varepsilon\}$.

Intuition that proper summary statistics coupled with small tolerance $\varepsilon = \varepsilon_\eta$ should provide good approximation of the posterior distribution:

$$\pi_\varepsilon^{\text{ABC}}(\theta|x^{\text{obs}}) = \int \pi_\varepsilon^{\text{ABC}}(\theta, z|x^{\text{obs}})\text{d}z \approx \pi\{\theta|\eta(x^{\text{obs}})\}$$

Why summaries?

- reduction of dimension
- improvement of signal to noise ratio
- reduce tolerance ε considerably
- whole data may be unavailable (as in Example 3)

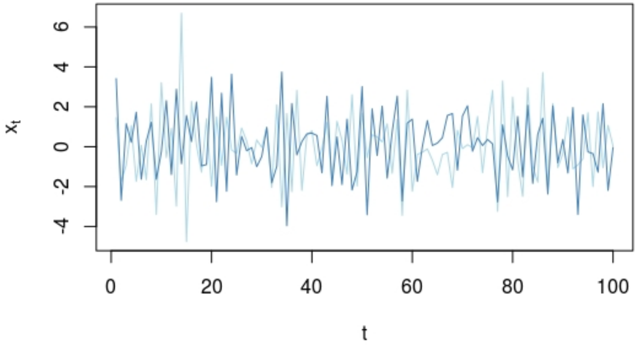
```
medobs=median(obs)
madobs=mad(obs) #summary
```

Example 6: MA inference

Moving average model MA(2):

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$$

Comparison of raw series:



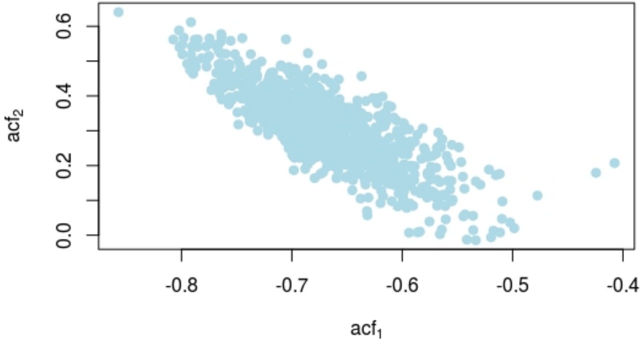
Example 6: MA inference

Moving average model MA(2):

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

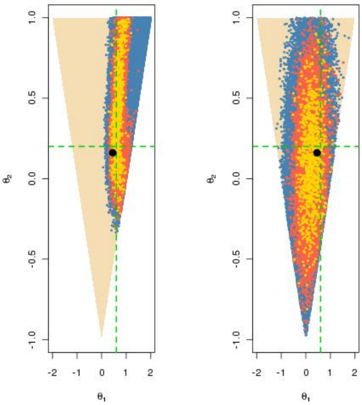
[Feller, 1970]

Comparison of acf's:



Example 6: MA inference

Summary vs. raw:



Why not summaries?

- loss of sufficient information when $\pi^{\text{ABC}}(\theta|x^{\text{obs}})$ replaced with $\pi^{\text{ABC}}(\theta|\eta(x^{\text{obs}}))$
- arbitrariness of summaries
- uncalibrated approximation
- whole data may be available (at same cost as summaries)
- (empirical) distributions may be compared (Wasserstein distances)

[Bernton et al., 2019]

Summary selection strategies

Fundamental difficulty of selecting summary statistics when there is no non-trivial sufficient statistic [except when done by experimenters from the field]

Summary selection strategies

Fundamental difficulty of selecting summary statistics when there is no non-trivial sufficient statistic [except when done by experimenters from the field]

1. Starting from large collection of summary statistics, Joyce and Marjoram (2008) consider the sequential inclusion into the ABC target, with a stopping rule based on a likelihood ratio test

Summary selection strategies

Fundamental difficulty of selecting summary statistics when there is no non-trivial sufficient statistic [except when done by experimenters from the field]

2. Based on decision-theoretic principles, Fearnhead and Prangle (2012) end up with $\mathbb{E}[\theta|x^{\text{obs}}]$ as the optimal summary statistic

Summary selection strategies

Fundamental difficulty of selecting summary statistics when there is no non-trivial sufficient statistic [except when done by experimenters from the field]

3. Use of indirect inference by Drovandi, Pettit, & Paddy (2011) with estimators of parameters of auxiliary model as summary statistics

Summary selection strategies

Fundamental difficulty of selecting summary statistics when there is no non-trivial sufficient statistic [except when done by experimenters from the field]

4. Starting from large collection of summary statistics, Raynal & al. (2018, 2019) rely on random forests to build estimators and select summaries

Summary selection strategies

Fundamental difficulty of selecting summary statistics when there is no non-trivial sufficient statistic [except when done by experimenters from the field]

5. Starting from large collection of summary statistics, Sedki & Pudlo (2012) use the Lasso to eliminate summaries

Semi-automated ABC

Use of summary statistic $\eta(\cdot)$, importance proposal $g(\cdot)$, kernel $K(\cdot) \leq 1$ with bandwidth $h \downarrow 0$ such that

$$(\theta, z) \sim g(\theta) f(z|\theta)$$

accepted with probability (hence the bound)

$$K[\{\eta(z) - \eta(x^{\text{obs}})\}/h]$$

and the corresponding importance weight defined by

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

Theorem Optimality of posterior expectation $\mathbb{E}[\theta|x^{\text{obs}}]$ of parameter of interest as summary statistics $\eta(x^{\text{obs}})$

[Fearnhead & Prangle, 2012; Sisson et al., 2019]

Semi-automated ABC

Use of summary statistic $\eta(\cdot)$, importance proposal $g(\cdot)$, kernel $K(\cdot) \leq 1$ with bandwidth $h \downarrow 0$ such that

$$(\theta, z) \sim g(\theta) f(z|\theta)$$

accepted with probability (hence the bound)

$$K[\{\eta(z) - \eta(x^{\text{obs}})\}/h]$$

and the corresponding importance weight defined by

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

Theorem Optimality of posterior expectation $\mathbb{E}[\theta|x^{\text{obs}}]$ of parameter of interest as summary statistics $\eta(x^{\text{obs}})$

[Fearnhead & Prangle, 2012; Sisson et al., 2019]

Random forests

Technique that stemmed from Leo Breiman's bagging (or *bootstrap aggregating*) machine learning algorithm for both classification [testing] and regression [estimation]

[Breiman, 1996]

Improved performances by averaging over classification schemes of randomly generated training sets, creating a "forest" of (CART) decision trees, inspired by Amit and Geman (1997) ensemble learning

[Breiman, 2001]

Growing the forest

Breiman's solution for inducing random features in the trees of the forest:

- boostrap resampling of the dataset and
- random subset-ing [of size $\sqrt{t}$] of the covariates driving the classification or regression at every node of each tree

Covariate (summary) x_τ that drives the node separation

$$x_\tau \gtrless c_\tau$$

and the separation bound c_τ chosen by minimising entropy or Gini index

ABC with random forests

Idea: Starting with

- possibly large collection of summary statistics $(\eta_1, \ldots, \eta_p)$ (from scientific theory input to available statistical softwares, methodologies, to machine-learning alternatives)
- ABC reference table involving model index, parameter values and summary statistics for the associated simulated pseudo-data

run R randomforest to infer $\mathfrak{M}$ or θ from $(\eta_{1i}, \ldots, \eta_{pi})$

ABC with random forests

Idea: Starting with

- possibly large collection of summary statistics $(\eta_1, \ldots, \eta_p)$ (from scientific theory input to available statistical softwares, methodologies, to machine-learning alternatives)
- ABC reference table involving model index, parameter values and summary statistics for the associated simulated pseudo-data

run R randomforest to infer $\mathfrak{M}$ or θ from $(\eta_{1i}, \ldots, \eta_{pi})$

Average of the trees is resulting summary statistics, highly non-linear predictor of the model index

ABC with random forests

Idea: Starting with

- possibly large collection of summary statistics $(\eta_1, \ldots, \eta_p)$ (from scientific theory input to available statistical softwares, methodologies, to machine-learning alternatives)
- ABC reference table involving model index, parameter values and summary statistics for the associated simulated pseudo-data

run R randomforest to infer $\mathfrak{M}$ or θ from $(\eta_{1i}, \ldots, \eta_{pi})$

Potential selection of most active summaries, calibrated against pure noise

Classification of summaries by random forests

Given large collection of summary statistics, rather than selecting a subset and excluding the others, estimate each parameter by random forests

- handles thousands of predictors
- ignores useless components
- fast estimation with good local properties
- automatised with few calibration steps
- substitute to Fearnhead and Prangle (2012) preliminary estimation of $\hat{\theta}(y^{\text{obs}})$
- includes a natural (classification) distance measure that avoids choice of either distance or tolerance

[Marin et al., 2016, 2018]

Calibration of tolerance

Calibration of threshold ε

- from scratch [how small is small?]
- from k-nearest neighbour perspective [quantile of prior predictive]

  ```
  subz=which(diz<quantile(diz,.1))
  ```
- from asymptotics [convergence speed]
- related with choice of distance [automated selection by random forests]

[Fearnhead & Prangle, 2012; Biau et al., 2013; Liu & Fearnhead 2018]

Implementation

Sofware

Several **ABC R packages** for performing parameter estimation and model selection

Name	References	Stand-alone	Platform	Models
abc	Csilléry et al. (2012)	No (R package)	All	General
ABCreg	Thornton (2009)	Yes	Linux, OS X	General
easyABC	Jabot et al. (2013)	No (R package)	All	General
ABCtoolbox	Wegmann et al. (2010)	Yes	Linux, Windows	Genetics
Bayes-SSC	Anderson et al. (2005)	Yes	All	Genetics
DIY-ABC	Cornuet et al. (2008, 2010, 2014)	Yes	All	Genetics
msBayes	Hickerson et al. (2007)	Yes	Linux, OS X	Genetics
MTML-msBayes	Huang et al. (2011)	Yes	Linux, OS X	Genetics
onesamp	Tallmon et al. (2008)	Yes (web interface)	All	Genetics
PopABC	Lopes et al. (2009)	Yes	All	Genetics
REJECTOR	Jobin and Mountain (2008)	Yes	All	Genetics
EP-ABC	Barthelmé and Chopin (2014)	No (MATLAB toolbox)	All	State space models (and related)
ABC-SDE	Picchini (2013)	No (MATLAB toolbox)	All	Stochastic differential equations
ABC-SysBio	Liepe et al. (2010)	Yes (Python scripts)	All	Systems biology

Table 1: Software for ABC. "All" regarding platform refers to Linux, OS X (Mac) and Windows.

[Nunes & Prangle, 2017]

Sofware

Several **ABC R packages** for performing parameter estimation and model selection

Name	References	Stand-alone	Platform	Models
abc	Csilléry et al. (2012)	No (R package)	All	General
ABCreg	Thornton (2009)	Yes	Linux, OS X	General
easyABC	Jabot et al. (2013)	No (R package)	All	General
ABCtoolbox	Wegmann et al. (2010)	Yes	Linux, Windows	Genetics
Bayes-SSC	Anderson et al. (2005)	Yes	All	Genetics
DIY-ABC	Cornuet et al. (2008, 2010, 2014)	Yes	All	Genetics
msBayes	Hickerson et al. (2007)	Yes	Linux, OS X	Genetics
MTML-msBayes	Huang et al. (2011)	Yes	Linux, OS X	Genetics
onesamp	Tallmon et al. (2008)	Yes (web interface)	All	Genetics
PopABC	Lopes et al. (2009)	Yes	All	Genetics
REJECTOR	Jobin and Mountain (2008)	Yes	All	Genetics
EP-ABC	Barthelmé and Chopin (2014)	No (MATLAB toolbox)	All	State space models (and related)
ABC-SDE	Picchini (2013)	No (MATLAB toolbox)	All	Stochastic differential equations
ABC-SysBio	Liepe et al. (2010)	Yes (Python scripts)	All	Systems biology

Table 1: Software for ABC. "All" regarding platform refers to Linux, OS X (Mac) and Windows.

[Nunes & Prangle, 2017]

abctools R package tuning ABC analyses

Sofware

Several ABC R packages for performing parameter estimation and model selection

Name	References	Stand-alone	Platform	Models
abc	Csilléry et al. (2012)	No (R package)	All	General
ABCreg	Thornton (2009)	Yes	Linux, OS X	General
easyABC	Jabot et al. (2013)	No (R package)	All	General
ABCtoolbox	Wegmann et al. (2010)	Yes	Linux, Windows	Genetics
Bayes-SSC	Anderson et al. (2005)	Yes	All	Genetics
DIY-ABC	Cornuet et al. (2008, 2010, 2014)	Yes	All	Genetics
msBayes	Hickerson et al. (2007)	Yes	Linux, OS X	Genetics
MTML-msBayes	Huang et al. (2011)	Yes	Linux, OS X	Genetics
onesamp	Tallmon et al. (2008)	Yes (web interface)	All	Genetics
PopABC	Lopes et al. (2009)	Yes	All	Genetics
REJECTOR	Jobin and Mountain (2008)	Yes	All	Genetics
EP-ABC	Barthelmé and Chopin (2014)	No (MATLAB toolbox)	All	State space models (and related)
ABC-SDE	Picchini (2013)	No (MATLAB toolbox)	All	Stochastic differential equations
ABC-SysBio	Liepe et al. (2010)	Yes (Python scripts)	All	Systems biology

Table 1: Software for ABC. "All" regarding platform refers to Linux, OS X (Mac) and Windows.

[Nunes & Prangle, 2017]

abcrf R package ABC via random forests

Sofware

Several **ABC R packages** for performing parameter estimation and model selection

Name	References	Stand-alone	Platform	Models
abc	Csilléry et al. (2012)	No (R package)	All	General
ABCreg	Thornton (2009)	Yes	Linux, OS X	General
easyABC	Jabot et al. (2013)	No (R package)	All	General
ABCtoolbox	Wegmann et al. (2010)	Yes	Linux, Windows	Genetics
Bayes-SSC	Anderson et al. (2005)	Yes	All	Genetics
DIY-ABC	Cornuet et al. (2008, 2010, 2014)	Yes	All	Genetics
msBayes	Hickerson et al. (2007)	Yes	Linux, OS X	Genetics
MTML-msBayes	Huang et al. (2011)	Yes	Linux, OS X	Genetics
onesamp	Tallmon et al. (2008)	Yes (web interface)	All	Genetics
PopABC	Lopes et al. (2009)	Yes	All	Genetics
REJECTOR	Jobin and Mountain (2008)	Yes	All	Genetics
EP-ABC	Barthelmé and Chopin (2014)	No (MATLAB toolbox)	All	State space models (and related)
ABC-SDE	Picchini (2013)	No (MATLAB toolbox)	All	Stochastic differential equations
ABC-SysBio	Liepe et al. (2010)	Yes (Python scripts)	All	Systems biology

Table 1: Software for ABC. “All” regarding platform refers to Linux, OS X (Mac) and Windows.

[Nunes & Prangle, 2017]

EasyABC R package several algorithms for performing efficient ABC sampling schemes, including four sequential sampling schemes and 3 MCMC schemes

Sofware

Several **ABC R packages** for performing parameter estimation and model selection

Name	References	Stand-alone	Platform	Models
abc	Csilléry et al. (2012)	No (R package)	All	General
ABCreg	Thornton (2009)	Yes	Linux, OS X	General
easyABC	Jabot et al. (2013)	No (R package)	All	General
ABCtoolbox	Wegmann et al. (2010)	Yes	Linux, Windows	Genetics
Bayes-SSC	Anderson et al. (2005)	Yes	All	Genetics
DIY-ABC	Cornuet et al. (2008, 2010, 2014)	Yes	All	Genetics
msBayes	Hickerson et al. (2007)	Yes	Linux, OS X	Genetics
MTML-msBayes	Huang et al. (2011)	Yes	Linux, OS X	Genetics
onesamp	Tallmon et al. (2008)	Yes (web interface)	All	Genetics
PopABC	Lopes et al. (2009)	Yes	All	Genetics
REJECTOR	Jobin and Mountain (2008)	Yes	All	Genetics
EP-ABC	Barthelmé and Chopin (2014)	No (MATLAB toolbox)	All	State space models (and related)
ABC-SDE	Picchini (2013)	No (MATLAB toolbox)	All	Stochastic differential equations
ABC-SysBio	Liepe et al. (2010)	Yes (Python scripts)	All	Systems biology

Table 1: Software for ABC. "All" regarding platform refers to Linux, OS X (Mac) and Windows.

[Nunes & Prangle, 2017]

DIYABC non R software for population genetics

ABC-IS

Basic ABC algorithm limitations

- blind [no learning]
- inefficient [curse of dimension]
- inapplicable to improper priors

Importance sampling version

- importance density $g(\theta)$
- bounded kernel function K_h with bandwidth h
- acceptance probability of

$$K_h\{\rho[\eta(x^{obs}), \eta(x\{\theta\})]\} \pi(\theta) \big/ g(\theta) \max_{\theta} A_\theta$$

[Fearnhead & Prangle, 2012]

ABC-IS

Basic ABC algorithm limitations

- blind [no learning]
- inefficient [curse of dimension]
- inapplicable to improper priors

Importance sampling version

- importance density $g(\theta)$
- bounded kernel function K_h with bandwidth h
- acceptance probability of

$$K_h\{\rho[\eta(x^{\text{obs}}),\eta(x\{\theta\})]\}\,\pi(\theta)\big/g(\theta)\,\max_{\theta} A_\theta$$

[Fearnhead & Prangle, 2012]

ABC-MCMC

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

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

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

[Marjoram et al, 2003]

ABC-MCMC

Algorithm 3 Likelihood-free MCMC sampler

get $(\theta^{(0)}, z^{(0)})$ by Algorithm **??**

for $t = 1$ to N **do**

generate θ' from $K_\omega\left(\cdot|\theta^{(t-1)}\right)$, z' from $f(\cdot|\theta')$, u from $\mathcal{U}_{[0,1]}$,

if $u \leq \frac{\pi(\theta')K_\omega(\theta^{(t-1)}|\theta')}{\pi(\theta^{(t-1)}K_\omega(\theta'|\theta^{(t-1)})}\mathbb{I}_{A_{\varepsilon,x^{\text{obs}}}}(z')$ **then**

set $(\theta^{(t)}, z^{(t)}) = (\theta', z')$

else

$(\theta^{(t)}, z^{(t)})) = (\theta^{(t-1)}, z^{(t-1)})$,

end if

end for

ABC-PMC

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, with tolerance $\varepsilon_t \downarrow$, and use importance weight associated with accepted simulation $\theta_i^{(t)}$

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

© **Still likelihood free**

[Sisson et al., 2007; Beaumont et al., 2009]

ABC-SMC

Use of a kernel K_t associated with target π_{ε_t} and derivation of the backward kernel

$$L_{t-1}(z, z') = \frac{\pi_{\varepsilon_t}(z') K_t(z', z)}{\pi_{\varepsilon_t}(z)}$$

Update of the weights

$$\omega_i^{(t)} \propto \omega_i^{(t-1)} \frac{\sum_{m=1}^{M} \mathbb{I}_{A_{\varepsilon_t}}(x_{im}^{(t)})}{\sum_{m=1}^{M} \mathbb{I}_{A_{\varepsilon_{t-1}}}(x_{im}^{(t-1)})}$$

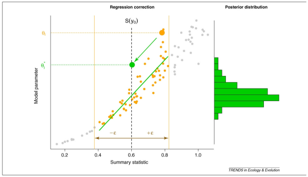
when $x_{im}^{(t)} \sim K_t(x_i^{(t-1)}, \cdot)$

[Del Moral, Doucet & Jasra, 2009]

ABC-NP

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

$$\theta^* = \theta - \{\eta(z) - \eta(x^{\text{obs}})\}^{\mathrm{T}} \hat{\beta}$$



[Csilléry et al., TEE, 2010]

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

$$K_\delta \left\{ \rho(\eta(z), \eta(x^{\text{obs}})) \right\}$$

[Beaumont et al., 2002, Genetics]

ABC-NN

Incorporating non-linearities and heterocedasticities:

$$\theta^* = \hat{m}(\eta(x^{\text{obs}})) + [\theta - \hat{m}(\eta(z))] \frac{\hat{\sigma}(\eta(x^{\text{obs}}))}{\hat{\sigma}(\eta(z))}$$

where

- $\hat{m}(\eta)$ estimated by non-linear regression (e.g., neural network)
- $\hat{\sigma}(\eta)$ estimated by non-linear regression on residuals

$$\log\{\theta_i - \hat{m}(\eta_i)\}^2 = \log \sigma^2(\eta_i) + \xi_i$$

[Blum & François, 2009]

ABC-NN

Incorporating non-linearities and heterocedasticities:

$$\theta^* = \hat{m}(\eta(x^{\text{obs}})) + [\theta - \hat{m}(\eta(z))] \frac{\hat{\sigma}(\eta(x^{\text{obs}}))}{\hat{\sigma}(\eta(z))}$$

where

- $\hat{m}(\eta)$ estimated by non-linear regression (e.g., neural network)
- $\hat{\sigma}(\eta)$ estimated by non-linear regression on residuals

$$\log\{\theta_i - \hat{m}(\eta_i)\}^2 = \log \sigma^2(\eta_i) + \xi_i$$

[Blum & François, 2009]

Convergence

Asymptotics of ABC

Since $\pi^{\text{ABC}}(\cdot \mid x^{\text{obs}})$ is an approximation of $\pi(\cdot \mid x^{\text{obs}})$ or $\pi(\cdot \mid \eta(x^{\text{obs}}))$ **coherence of ABC-based inference need be established on its own**

[Li & Fearnhead, 2018a,b; Frazier et al., 2018,2020]

Meaning

- establishing large sample (n) properties of ABC posteriors and ABC procedures
- finding sufficient conditions and checks on summary statistics $\eta(\cdot)$
- determining proper rate $\varepsilon = \varepsilon_n$ of convergence of tolerance to 0
- [mostly] ignoring Monte Carlo errors

Asymptotics of ABC

Since $\pi^{\mathrm{ABC}}(\cdot \mid x^{\mathrm{obs}})$ is an approximation of $\pi(\cdot \mid x^{\mathrm{obs}})$ or $\pi(\cdot \mid \eta(x^{\mathrm{obs}}))$ **coherence of ABC-based inference need be established on its own**

[Li & Fearnhead, 2018a,b; Frazier et al., 2018,2020]

Meaning

- establishing large sample (n) properties of ABC posteriors and ABC procedures
- finding sufficient conditions and checks on summary statistics $\eta(\cdot)$
- determining proper rate $\varepsilon = \varepsilon_n$ of convergence of tolerance to 0
- [mostly] ignoring Monte Carlo errors

Consistency of ABC posteriors

ABC algorithm **Bayesian consistent** at θ_0 if for any $\delta > 0$,

$$\Pi\left(\|\theta - \theta_0\| > \delta \mid \rho\{\eta(x^{\text{obs}}), \eta(Z)\} \leq \varepsilon\right) \to 0$$

as $n \to +\infty, \varepsilon \to 0$

Bayesian consistency implies that sets containing θ_0 have posterior probability tending to one as $n \to +\infty$, with implication being the existence of a specific rate of concentration

Consistency of ABC posteriors

ABC algorithm **Bayesian consistent** at θ_0 if for any $\delta > 0$,

$$\Pi\left(\|\theta - \theta_0\| > \delta \mid \rho\{\eta(x^{\text{obs}}), \eta(Z)\} \leq \varepsilon\right) \to 0$$

as $n \to +\infty, \varepsilon \to 0$

- Concentration around true value and Bayesian consistency impose less stringent conditions on the convergence speed of tolerance ε_n to zero, when compared with asymptotic normality of ABC posterior
- asymptotic normality of ABC posterior mean does not require asymptotic normality of ABC posterior

Asymptotic setup

Assumptions:

- asymptotic: $x^{\text{obs}} = x^{\text{obs}(n)} \sim \mathbb{P}^n_{\theta_0}$ and $\varepsilon = \varepsilon_n$, $n \to +\infty$
- parametric: $\theta \in \mathbb{R}^k$, k fixed concentration of summary statistic $\eta(z^n)$:

$$\exists b : \theta \to b(\theta) \quad \eta(z^n) - b(\theta) = o_{\mathbb{P}_\theta}(1), \quad \forall \theta$$

- identifiability of parameter $b(\theta) \neq b(\theta')$ when $\theta \neq \theta'$

Consistency of ABC posteriors

- Concentration of summary $\eta(z)$: there exists $b(\theta)$ such that

$$\eta(z) - b(\theta) = o_{\mathbb{P}_\theta}(1)$$

- Consistency:

$$\Pi_{\varepsilon_n}\left(\|\theta - \theta_0\| \leq \delta \mid \eta(x^{\text{obs}})\right) = 1 + o_p(1)$$

- Convergence rate: there exists $\delta_n = o(1)$ such that

$$\Pi_{\varepsilon_n}\left(\|\theta - \theta_0\| \leq \delta_n \mid \eta(x^{\text{obs}})\right) = 1 + o_p(1)$$

Consistency of ABC posteriors

- Consistency:

$$\Pi_{\varepsilon_n}\left(\|\theta-\theta_0\| \le \delta \mid \eta(x^{obs})\right) = 1 + o_p(1)$$

- Convergence rate: there exists $\delta_n = o(1)$ such that

$$\Pi_{\varepsilon_n}\left(\|\theta-\theta_0\| \le \delta_n \mid \eta(x^{obs})\right) = 1 + o_p(1)$$

- Point estimator consistency

$$\hat{\theta}_\varepsilon = \mathbb{E}_{ABC}[\theta|\eta(x^{obs^{(n)}})], \quad \mathbb{E}_{ABC}[\theta \mid \eta(x^{obs^{(n)}})] - \theta_0 = o_p(1)$$

$$v_n(\mathbb{E}_{ABC}[\theta \mid \eta(x^{obs^{(n)}})] - \theta_0) \Rightarrow N(0, v)$$

Asymptotic shape of posterior distribution

Shape of

$$\Pi\left(\cdot \mid \|\eta(x^{\text{obs}}), \eta(z)\| \leq \varepsilon_n\right)$$

depending on relation between ε_n and rate σ_n at which $\eta(x^{\text{obs}^n})$ satisfy CLT

Three different regimes:

1. $\sigma_n = o(\varepsilon_n) \longrightarrow$ Uniform limit
2. $\sigma_n \asymp \varepsilon_n \longrightarrow$ perturbated Gaussian limit
3. $\sigma_n \gg \varepsilon_n \longrightarrow$ Gaussian limit

Asymptotic behaviour of posterior mean

When $k_\eta = \dim(\eta(x^{\text{obs}})) = k_\theta = \dim(\theta)$ and $\varepsilon_n = o(n^{-3/10})$

$$\mathbb{E}_{ABC}[\nu_n(\theta - \theta_0) \mid x^{\text{obs}}] \Rightarrow N(0, \left\{(\nabla b^o)^T \Sigma^{-1} \nabla b^o\right\}^{-1}$$

[Li & Fearnhead (2018a)]

In fact, if $\varepsilon_n^{\beta+1}\sqrt{n} = o(1)$, with β Hölder-smoothness of π

$$\mathbb{E}_{ABC}[(\theta - \theta_0) \mid x^{\text{obs}}] = \frac{(\nabla b^o)^{-1} Z^o}{\sqrt{n}} + \sum_{j=1}^{k} h_j(\theta_0)\varepsilon_n^{2j} + o_p(1), \quad 2k = \lfloor \beta \rfloor$$

[Fearnhead & Prangle, 2012]

Asymptotic behaviour of posterior mean

When $k_\eta = \dim(\eta(x^{\text{obs}})) = k_\theta = \dim(\theta)$ and $\varepsilon_n = o(n^{-3/10})$

$$\mathbb{E}_{ABC}[\nu_n(\theta - \theta_0) \mid x^{\text{obs}}] \Rightarrow N(0, \left\{(\nabla b^o)^T \Sigma^{-1} \nabla b^o\right\}^{-1}$$

[Li & Fearnhead (2018a)]

Iterating for fixed k_θ mildly interesting: if

$$\tilde{\eta}(x^{\text{obs}}) = \mathbb{E}_{ABC}[\theta \mid x^{\text{obs}}]$$

then

$$\mathbb{E}_{ABC}[\theta|\tilde{\eta}(x^{\text{obs}})] = \theta_0 + \frac{(\nabla b^o)^{-1} Z^o}{\sqrt{n}} + \frac{\pi'(\theta_0)}{\pi(\theta_0)} \varepsilon_n^2 + o()$$

[Fearnhead & Prangle, 2012]

Curse of dimension

- for reasonable statistical behavior, decline of acceptance α_n the faster the larger the dimension of θ, k_θ, but unaffected by dimension of η, k_η
- theoretical justification for dimension reduction methods that process parameter components individually and independently of other components

  [Fearnhead & Prangle, 2012; Martin & al., 2016]

- importance sampling approach of Li & Fearnhead (2018a) yields acceptance rates $\alpha_n = O(1)$, when $\varepsilon_n = O(1/v_n)$

Curse of dimension

- for reasonable statistical behavior, decline of acceptance α_n the faster the larger the dimension of θ, k_θ, but unaffected by dimension of η, k_η
- theoretical justification for dimension reduction methods that process parameter components individually and independently of other components
  [Fearnhead & Prangle, 2012; Martin & al., 2016]
- importance sampling approach of Li & Fearnhead (2018a) yields acceptance rates $\alpha_n = O(1)$, when $\varepsilon_n = O(1/v_n)$

Monte Carlo error

Link the choice of ε_n to Monte Carlo error associated with N_n draws in ABC Algorithm

Conditions (on ε_n) under which

$$\hat{\alpha}_n = \alpha_n\{1 + o_p(1)\}$$

where $\hat{\alpha}_n = \sum_{i=1}^{N_n} \mathbb{I}\left[d\{\eta(y), \eta(z)\} \leq \varepsilon_n\right] / N_n$ proportion of accepted draws from N_n simulated draws of θ

Either

(i) $\varepsilon_n = o(\nu_n^{-1})$ and $(\nu_n \varepsilon_n)^{-k_\eta} \varepsilon_n^{-k_\theta} \leq M N_n$

or

(ii) $\varepsilon_n \gtrsim \nu_n^{-1}$ and $\varepsilon_n^{-k_\theta} \leq M N_n$

for M large enough

Bayesian model choice

Model candidates $M_1, M_2, \ldots$ to be compared for dataset x^{obs} making model index $\mathcal{M}$ part of inference

Use of a prior distribution. $\pi(\mathcal{M} = m)$, plus a prior distribution on the parameter conditional on the value m of the model index, $\pi_m(\theta_m)$

Goal to derive the posterior distribution of M, challenging computational target when models are complex

[Savage, 1964; Berger, 1980]

Generic ABC for model choice

Algorithm 4 Likelihood-free model choice sampler (ABC-MC)

for $t = 1$ to T **do**
  repeat
    Generate m from the prior $\pi(\mathcal{M} = m)$
    Generate θ_m from the prior $\pi_m(\theta_m)$
    Generate z from the model $f_m(z|\theta_m)$
  until $\rho\{\eta(z), \eta(x^{\text{obs}})\} < \epsilon$
  Set $m^{(t)} = m$ and $\theta^{(t)} = \theta_m$
end for

[Cornuet et al., DIYABC, 2009]

ABC model choice consistency

Leaving approximations aside, limiting ABC procedure is Bayes factor based on $\eta(x^{\text{obs}})$

$$B_{12}(\eta(x^{\text{obs}}))$$

Potential loss of information at the testing level

[Robert et al., 2010]

When is Bayes factor based on insufficient statistic $\eta(x^{\text{obs}})$ consistent?

[Marin et al., 2013]

ABC model choice consistency

Leaving approximations aside, limiting ABC procedure is Bayes factor based on $\eta(x^{\text{obs}})$

$$B_{12}(\eta(x^{\text{obs}}))$$

Potential loss of information at the testing level

[Robert et al., 2010]

When is Bayes factor based on insufficient statistic $\eta(x^{\text{obs}})$ consistent?

[Marin et al., 2013]

Example 7: Gauss versus Laplace

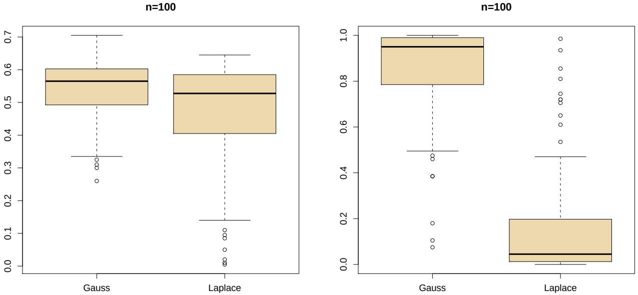
Model $\mathfrak{M}_1$: $x^{\text{obs}} \sim \mathcal{N}(\theta_1, 1)^{\otimes n}$ opposed to model $\mathfrak{M}_2$: $x^{\text{obs}} \sim \mathcal{L}(\theta_2, 1/\sqrt{2})^{\otimes n}$, Laplace distribution with mean θ_2 and variance one

Four possible statistics $\eta(x^{\text{obs}})$

1. sample mean $\overline{x^{\text{obs}}}$ (sufficient for $\mathfrak{M}_1$ if not $\mathfrak{M}$);
2. sample median $\text{med}(x^{\text{obs}})$ (insufficient);
3. sample variance $\text{var}(x^{\text{obs}})$ (ancillary);
4. median absolute deviation $\text{mad}(x^{\text{obs}}) = \text{med}(|x^{\text{obs}} - \text{med}(x^{\text{obs}})|)$;

Example 7: Gauss versus Laplace

Model $\mathfrak{M}_1$: $x^{\text{obs}} \sim \mathcal{N}(\theta_1, 1)^{\otimes n}$ opposed to model $\mathfrak{M}_2$: $x^{\text{obs}} \sim \mathcal{L}(\theta_2, 1/\sqrt{2})^{\otimes n}$, Laplace distribution with mean θ_2 and variance one



Consistency

Summary statistics

$$\eta(x^{obs}) = (\tau_1(x^{obs}), \tau_2(x^{obs}), \cdots, \tau_d(x^{obs})) \in \mathbb{R}^d$$

with

- true distribution $\eta \sim G_n$, true mean μ_0,
- distribution $G_{i,n}$ under model $\mathfrak{M}_i$, corresponding posteriors $\pi_i(\cdot \mid \eta^n)$

Assumptions of central limit theorem and large deviations for $\eta(x^{obs})$ under true, plus usual Bayesian asymptotics with d_i effective dimension of the parameter)

[Pillai et al., 2013]

Asymptotic marginals

Asymptotically

$$m_{i,n}(t) = \int_{\Theta_i} g_{i,n}(t|\theta_i)\, \pi_i(\theta_i)\, d\theta_i$$

such that
(i) if $\inf\{|\mu_i(\theta_i) - \mu_0|; \theta_i \in \Theta_i\} = 0$,

$$C_l \sqrt{n}^{d-d_i} \leq m_{i,n}(\eta^n) \leq C_u \sqrt{n}^{d-d_i}$$

and
(ii) if $\inf\{|\mu_i(\theta_i) - \mu_0|; \theta_i \in \Theta_i\} > 0$

$$m_{i,n}(\eta^n) = o_{P^n}[\sqrt{n}^{d-\tau_i} + \sqrt{n}^{d-\alpha_i}].$$

Between-model consistency

Consequence of above is that asymptotic behaviour of the Bayes factor is driven by the asymptotic mean value $\mu(\theta)$ of η^n under both models. **And only by this mean value!**

Between-model consistency

Consequence of above is that asymptotic behaviour of the Bayes factor is driven by the asymptotic mean value $\mu(\theta)$ of η^n under both models. **And only by this mean value!**

Indeed, if

$$\inf\{|\mu_0 - \mu_2(\theta_2)|; \theta_2 \in \Theta_2\} = \inf\{|\mu_0 - \mu_1(\theta_1)|; \theta_1 \in \Theta_1\} = 0$$

then

$$C_l \sqrt{n}^{-(d_1-d_2)} \leq m_{1,n}(\eta^n) / m_2(\eta^n) \leq C_u \sqrt{n}^{-(d_1-d_2)},$$

where $C_l, C_u = O_{P^n}(1)$, irrespective of the true model.
© Only depends on the difference $d_1 - d_2$: no consistency

Between-model consistency

Consequence of above is that asymptotic behaviour of the Bayes factor is driven by the asymptotic mean value $\mu(\theta)$ of η^n under both models. **And only by this mean value!**

Else, if

$$\inf\{|\mu_0 - \mu_2(\theta_2)|; \theta_2 \in \Theta_2\} > \inf\{|\mu_0 - \mu_1(\theta_1)|; \theta_1 \in \Theta_1\} = 0$$

then

$$\frac{m_{1,n}(\eta^n)}{m_{2,n}(\eta^n)} \geq C_u \min\left(\sqrt{n}^{-(d_1-\alpha_2)}, \sqrt{n}^{-(d_1-\tau_2)}\right)$$

Checking for adequate statistics

Run a practical check of the relevance (or non-relevance) of η^n

null hypothesis that both models are compatible with the statistic η^n

$$\mathfrak{H}_0 : \inf\{|\mu_2(\theta_2) - \mu_0|; \theta_2 \in \Theta_2\} = 0$$

against

$$\mathfrak{H}_1 : \inf\{|\mu_2(\theta_2) - \mu_0|; \theta_2 \in \Theta_2\} > 0$$

testing procedure provides estimates of mean of η^n under each model and checks for equality

ABC under misspecification

ABC methods rely on simulations $z(\theta)$ from the model to identify those close to x^{obs}

What is happening when the model is wrong?

- for some tolerance sequences $\varepsilon_n \downarrow \varepsilon^*$, well-behaved ABC posteriors that concentrate posterior mass on pseudo-true value
- if ε_n too large, asymptotic limit of ABC posterior uniform with radius of order $\varepsilon_n - \varepsilon^*$
- even if $\sqrt{n}\{\varepsilon_n - \varepsilon^*\} \to 2c \in \mathbb{R}$, limiting distribution no longer Gaussian
- ABC credible sets invalid confidence sets

[Frazier et al., 2020]

ABC under misspecification

ABC methods rely on simulations $z(\theta)$ from the model to identify those close to x^{obs}

What is happening when the model is wrong?

- for some tolerance sequences $\epsilon_n \downarrow \epsilon^*$, well-behaved ABC posteriors that concentrate posterior mass on pseudo-true value
- if ϵ_n too large, asymptotic limit of ABC posterior uniform with radius of order $\epsilon_n - \epsilon^*$
- even if $\sqrt{n}\{\epsilon_n - \epsilon^*\} \to 2c \in \mathbb{R}$, limiting distribution no longer Gaussian
- ABC credible sets invalid confidence sets

[Frazier et al., 2020]

WARWICK THE UNIVERSITY OF WARWICK DAUPHINE UNIVERSITÉ PARIS | PSL

Example 8: Normal model with wrong variance

Assumed data generating process (DGP) is $z \sim \mathcal{N}(\theta, 1)$ but actual DGP is $x^{\text{obs}} \sim \mathcal{N}(\theta, \tilde{\sigma}^2)$

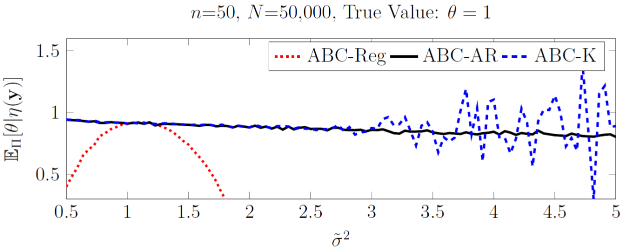
Use of summaries

- sample mean $\eta_1(x^{\text{obs}}) = \frac{1}{n}\sum_{i=1}^{n} x_i$
- centered summary $\eta_2(x^{\text{obs}}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \eta_1(x^{\text{obs}})^2 - 1$

Three ABC:

- ABC-AR: accept/reject approach with $K_\varepsilon(d\{\eta(z), \eta(x^{\text{obs}})\}) = \mathbb{I}\left[d\{\eta(z), \eta(x^{\text{obs}})\} \leq \varepsilon\right]$ and $d\{x, y\} = \|x - y\|$
- ABC-K: smooth rejection approach, with $K_\varepsilon(d\{\eta(z), \eta(x^{\text{obs}})\})$ univariate Gaussian kernel
- ABC-Reg: post-processing ABC approach with weighted linear regression adjustment

Example 8: Normal model with wrong variance



- posterior means for ABC-AR, ABC-K and ABC-Reg as σ^2 increases ($N = 50,000$ simulated data sets)
- $\alpha_n = n^{-5/9}$ quantile for ABC-AR
- ABC-K and ABC-Reg bandwidth of $n^{-5/9}$

[Frazier et al., 2020]

WARWICK THE UNIVERSITY OF WARWICK | DAUPHINE UNIVERSITÉ PARIS | PSL

ABC misspecification

- data x^{obs} with true distribution P_0 assumed issued from model P_θ $\theta \in \Theta \subset \mathbb{R}^{k_\theta}$ and summary statistic $\eta(x^{\text{obs}}) = (\eta_1(x^{\text{obs}}), ..., \eta_{k_\eta}(x^{\text{obs}}))$
- misspecification

$$\inf_{\theta \in \Theta} \mathcal{D}(P_0 \| P_\theta) = \inf_{\theta \in \Theta} \int \log \left\{ \frac{dP_0(x)}{dP_\theta(x)} \right\} dP_0(y) > 0,$$

with

$$\theta^* = \arg \inf_{\theta \in \Theta} \mathcal{D}(P_0 \| P_\theta)$$

[Muller, 2013]

- **ABC misspecification:**
for b_0 (resp. $b(\theta)$) limit of $\eta(x^{\text{obs}})$ (resp. $\eta(z)$)

$$\inf_{\theta \in \Theta} d\{b_0, b(\theta)\} > 0$$

ABC misspecification

- data x^{obs} with true distribution P_0 assumed issued from model P_θ $\theta \in \Theta \subset \mathbb{R}^{k_\theta}$ and summary statistic $\eta(x^{obs}) = (\eta_1(x^{obs}), ..., \eta_{k_\eta}(x^{obs}))$
- **ABC misspecification:**
  for b_0 (resp. $b(\theta)$) limit of $\eta(x^{obs})$ (resp. $\eta(z)$)

$$\inf_{\theta \in \Theta} d\{b_0, b(\theta)\} > 0$$

- **ABC pseudo-true value:**

$$\theta^* = \arg \inf_{\theta \in \Theta} d\{b_0, b(\theta)\}.$$

Minimum tolerance

Under identification conditions on $b(\cdot) \in \mathbb{R}^{k_\eta}$, there exists ε^* such that

$$\varepsilon^* = \inf_{\theta \in \Theta} d\{b_0, b(\theta)\} > 0$$

Once $\varepsilon_n < \varepsilon^*$ no draw of θ to be selected and posterior $\Pi_\varepsilon[A|\eta(x^{obs})]$ ill-conditioned

But appropriately chosen tolerance sequence $(\varepsilon_n)_n$ allows ABC-based posterior to concentrate on distance-dependent pseudo-true value θ^*

Minimum tolerance

Under identification conditions on $b(\cdot) \in \mathbb{R}^{k_\eta}$, there exists ε^* such that

$$\varepsilon^* = \inf_{\theta \in \Theta} d\{b_0, b(\theta)\} > 0$$

Once $\varepsilon_n < \varepsilon^*$ no draw of θ to be selected and posterior $\Pi_\varepsilon[A|\eta(x^{\text{obs}})]$ ill-conditioned

But appropriately chosen tolerance sequence $(\varepsilon_n)_n$ allows ABC-based posterior to concentrate on distance-dependent pseudo-true value θ^*

Minimum tolerance

Under identification conditions on $b(\cdot) \in \mathbb{R}^{k_\eta}$, there exists ε^* such that

$$\varepsilon^* = \inf_{\theta \in \Theta} d\{b_0, b(\theta)\} > 0$$

Once $\varepsilon_n < \varepsilon^*$ no draw of θ to be selected and posterior $\Pi_\varepsilon[A|\eta(x^{obs})]$ ill-conditioned

But appropriately chosen tolerance sequence $(\varepsilon_n)_n$ allows ABC-based posterior to concentrate on distance-dependent pseudo-true value θ^*

ABC concentration under misspecification

Assumptions

[A0] Existence of unique b_0 such that $d(\eta(x^{\text{obs}}), b_0) = o_{P_0}(1)$ and of sequence $v_{0,n} \to +\infty$ such that

$$\liminf_{n\to+\infty} P_0\left[d(\eta(x_n^{\text{obs}}), b_0) \geq v_{0,n}^{-1}\right] = 1.$$

ABC concentration under misspecification

Assumptions

[A1] Existence of injective map $b : \Theta \to \mathcal{B} \subset \mathbb{R}^{k_\eta}$ and function ρ_n with $\rho_n(\cdot) \downarrow 0$ as $n \to +\infty$, and $\rho_n(\cdot)$ non-increasing, such that

$$P_\theta\left[d(\eta(Z), b(\theta)) > u\right] \leq c(\theta)\rho_n(u), \quad \int_\Theta c(\theta)d\Pi(\theta) < \infty$$

and assume either

(i) *Polynomial deviations:* existence of $v_n \uparrow +\infty$ and $u_0, \kappa > 0$ such that $\rho_n(u) = v_n^{-\kappa}u^{-\kappa}$, for $u \leq u_0$

(ii) *Exponential deviations:*

ABC concentration under misspecification

Assumptions

[A1] Existence of injective map $b : \Theta \to \mathcal{B} \subset \mathbb{R}^{k_\eta}$ and function ρ_n with $\rho_n(\cdot) \downarrow 0$ as $n \to +\infty$, and $\rho_n(\cdot)$ non-increasing, such that

$$P_\theta \left[d(\eta(Z), b(\theta)) > u \right] \leq c(\theta)\rho_n(u), \quad \int_\Theta c(\theta) d\Pi(\theta) < \infty$$

and assume either

(i) *Polynomial deviations:*

(ii) *Exponential deviations:* existence of $h_\theta(\cdot) > 0$ such that $P_\theta[d(\eta(z), b(\theta)) > u] \leq c(\theta)e^{-h_\theta(uv_n)}$ and existence of $m, C > 0$ such that

$$\int_\Theta c(\theta)e^{-h_\theta(uv_n)} d\Pi(\theta) \leq Ce^{-m\cdot(uv_n)^\tau}, \quad \text{for } u \leq u_0.$$

ABC concentration under misspecification

Assumptions

[A2] existence of $D > 0$ and $M_0, \delta_0 > 0$ such that, for all $\delta_0 \geq \delta > 0$ and $M \geq M_0$, existence of $S_\delta \subset \{\theta \in \Theta : d(b(\theta), b_0) - \varepsilon^* \leq \delta\}$ for which

- **(i)** In case (i), $D < \kappa$ and $\int_{S_\delta} \left(1 - \frac{c(\theta)}{M}\right) d\Pi(\theta) \gtrsim \delta^D$.
- **(ii)** In case (ii), $\int_{S_\delta} \left(1 - c(\theta) e^{-h_\theta(M)}\right) d\Pi(\theta) \gtrsim \delta^D$.

Consistency

Assume **[A0]** – **[A2]**, with $\varepsilon_n \downarrow \varepsilon^*$ with

$$\varepsilon_n \geq \varepsilon^* + M v_n^{-1} + v_{0,n}^{-1},$$

for M large enough. Let $M_n \uparrow \infty$ and $\delta_n \geq M_n(\varepsilon_n - \varepsilon^*)$, then

$$\Pi_\varepsilon \left[d(b(\theta), b_0) \geq \varepsilon^* + \delta_n \mid \eta(x^{\text{obs}}) \right] = o_{P_0}(1),$$

1. if $\delta_n \geq M_n v_n^{-1} u_n^{-D/\kappa} = o(1)$ in case (i)
2. if $\delta_n \geq M_n v_n^{-1} |\log(u_n)|^{1/\tau} = o(1)$ in case (ii)

with $u_n = \varepsilon_n - (\varepsilon^* + M v_n^{-1} + v_{0,n}^{-1}) \geq 0$.

[Bernton et al., 2017; Frazier et al., 2020]

Regression adjustement under misspecification

Accepted value θ artificially related to $\eta(x^{\text{obs}})$ and $\eta(z)$ through local linear regression model

$$\theta' = \mu + \beta^{\mathsf{T}}\{\eta(x^{\text{obs}}) - \eta(z)\} + \nu,$$

where ν_i model residual

[Beaumont et al., 2003]

Asymptotic behavior of ABC-Reg posterior

$$\widetilde{\Pi}_{\varepsilon}[\cdot \mid \eta(x^{\text{obs}})]$$

determined by behavior of

$$\Pi_{\varepsilon}[\cdot \mid \eta(x^{\text{obs}})],\ \hat{\beta},\ \text{and}\ \{\eta(x^{\text{obs}}) - \eta(z)\}$$

Regression adjustement under misspecification

Accepted value θ artificially related to $\eta(x^{obs})$ and $\eta(z)$ through local linear regression model

$$\theta' = \mu + \beta^\intercal\{\eta(x^{obs}) - \eta(z)\} + \nu,$$

where ν_i model residual

[Beaumont et al., 2003]

Asymptotic behavior of ABC-Reg posterior

$$\widetilde{\Pi}_\varepsilon[\cdot \mid \eta(x^{obs})]$$

determined by behavior of

$$\Pi_\varepsilon[\cdot \mid \eta(x^{obs})], \ \hat{\beta}, \ \text{and } \{\eta(x^{obs}) - \eta(z)\}$$

Regression adjustement under misspecification

- ABC-Reg takes draws of asymptotically optimal θ, perturbed in a manner that need not preserve original optimality
- for $\|\beta_0\|$ large, pseudo-true value $\tilde{\theta}^*$ possibly outside Θ
- extends to nonlinear regression adjustments [Blum & François, 2010]
- potential correction of the adjustment [Frazier et al., 2020]
- local regression adjustments with smaller posterior variability than ABC-AR but fake precision

Example 9: misspecified g-&-k

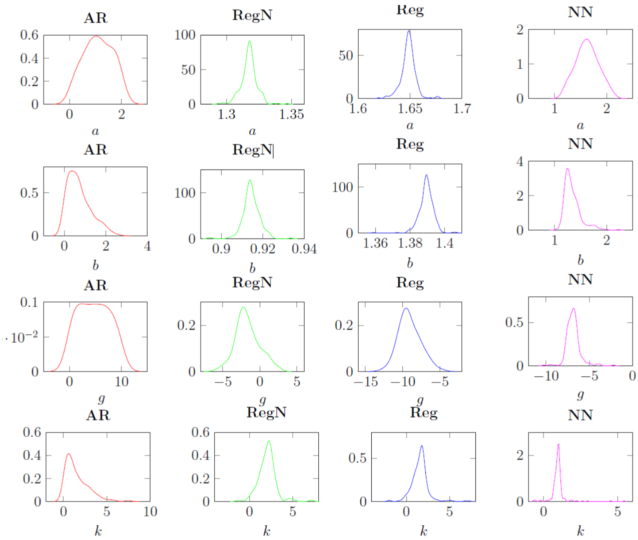
Quantile function of Tukey's g-&-k model:

$$F^{-1}(q) = a + b\left(1 + 0.8\frac{1-\exp(-gz(q))}{1+\exp(-gz(q))}\right)\left(1+z(q)^2\right)^k z(q),$$

where $z(q)$ q-th $\mathcal{N}(0,1)$ quantile

But data generated from a mixture distribution with minor bi-modality

Example 9: misspecified g-&-k



Advanced topics

Computational bottleneck

Time per iteration increases with sample size n of the data: cost of sampling $O(n^{1+?})$ associated with a reasonable acceptance probability makes ABC infeasible for large datasets

- surrogate models to get samples (e.g., using copulas)
- direct sampling of summary statistics (e.g., synthetic likelihood)

  [Wood, 2010]

- borrow from proposals for scalable MCMC (e.g., divide & conquer)

Approximate ABC [AABC]

Idea approximations on both parameter and model spaces by resorting to bootstrap techniques.

[Buzbas & Rosenberg, 2015]

Procedure scheme

1. Sample (θ_i, x_i), $i = 1, \ldots, m$, from prior predictive
2. Simulate $\theta^* \sim \pi(\cdot)$ and assign weight w_i to dataset $x_{(i)}$ simulated under k-closest θ_i to θ^*
3. Generate dataset x^* as bootstrap weighted sample from $(x_{(1)}, \ldots, x_{(k)})$

Drawbacks

- If m too small, prior predictive sample may miss informative parameters
- large error and misleading representation of true posterior
- only suited for models with very few parameters

Approximate ABC [AABC]

Procedure scheme

1. Sample (θ_i, x_i), $i = 1, \ldots, m$, from prior predictive
2. Simulate $\theta^* \sim \pi(\cdot)$ and assign weight w_i to dataset $x_{(i)}$ simulated under k-closest θ_i to θ^*
3. Generate dataset x^* as bootstrap weighted sample from $(x_{(1)}, \ldots, x_{(k)})$

Drawbacks

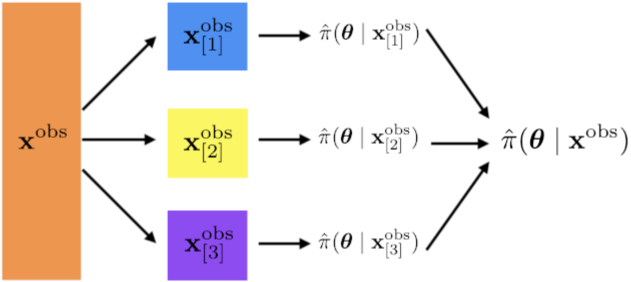
- If m too small, prior predictive sample may miss informative parameters
- large error and misleading representation of true posterior
- only suited for models with very few parameters

Divide-and-conquer perspectives

1. divide the large dataset into smaller batches
2. sample from the batch posterior
3. combine the result to get a sample from the targeted posterior

Alternative via ABC-EP

[Barthelmé & Chopin, 2014]

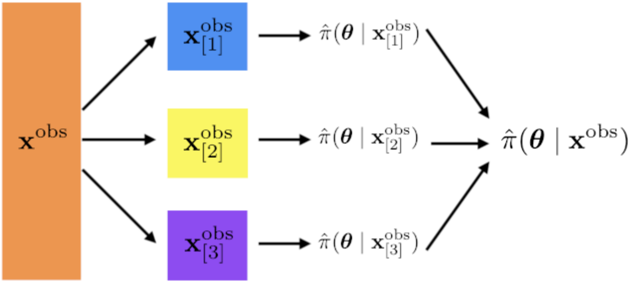


Divide-and-conquer perspectives

1. divide the large dataset into smaller batches
2. sample from the batch posterior
3. combine the result to get a sample from the targeted posterior

Alternative via ABC-EP

[Barthelmé & Chopin, 2014]



Geometric combination: WASP

Subset posteriors given partition $x^{obs}_{[1]}, \ldots, x^{obs}_{[B]}$ of observed data x^{obs}, let define

$$\pi(\theta \mid x^{obs}_{[b]}) \propto \pi(\theta) \prod_{j \in [b]} f(x^{obs}_j \mid \theta)^B.$$

[Srivastava et al., 2015]

Subset posteriors are combined via Wasserstein barycenter

[Cuturi, 2014]

Geometric combination: WASP

Subset posteriors given partition $x^{\text{obs}}_{[1]}, \ldots, x^{\text{obs}}_{[B]}$ of observed data x^{obs}, let define

$$\pi(\theta \mid x^{\text{obs}}_{[b]}) \propto \pi(\theta) \prod_{j \in [b]} f(x^{\text{obs}}_j \mid \theta)^B.$$

[Srivastava et al., 2015]

Subset posteriors are combined via Wasserstein barycenter

[Cuturi, 2014]

Drawback require sampling from $f(\cdot \mid \theta)^B$ by ABC means. Should be feasible for latent variable (z) representations when $f(x \mid z, \theta)$ available in closed form

[Doucet & Robert, 2001]

Geometric combination: WASP

Subset posteriors given partition $x^{\text{obs}}_{[1]}, \ldots, x^{\text{obs}}_{[B]}$ of observed data x^{obs}, let define

$$\pi(\theta \mid x^{\text{obs}}_{[b]}) \propto \pi(\theta) \prod_{j \in [b]} f(x^{\text{obs}}_j \mid \theta)^B.$$

[Srivastava et al., 2015]

Subset posteriors are combined via Wasserstein barycenter

[Cuturi, 2014]

Alternative backfeed subset posteriors as priors to other subsets, partitioning summaries

Consensus ABC

Naïve scheme

- For each data batch $b = 1, \ldots, B$
  1. Sample $(\theta_1^{[b]}, \ldots, \theta_n^{[b]})$ from diffused prior $\tilde{\pi}(\cdot) \propto \pi(\cdot)^{1/B}$
  2. Run ABC to sample from batch posterior $\hat{\pi}(\cdot \mid d(S(x_{[b]}^{obs}), S(x_{[b]})) < \varepsilon)$
  3. Compute sample posterior variance Σ_b^{-1}
- Combine batch posterior approximations

$$\theta_j = \sum_{b=1}^{B} \Sigma_b \theta_j^{[b]} \Big/ \sum_{b=1}^{B} \Sigma_b$$

Remark Diffuse prior $\tilde{\pi}(\cdot)$ non informative calls for ABC-MCMC steps

Consensus ABC

Naïve scheme

- For each data batch $b = 1, \ldots, B$
  1. Sample $(\theta_1^{[b]}, \ldots, \theta_n^{[b]})$ from diffused prior $\tilde{\pi}(\cdot) \propto \pi(\cdot)^{1/B}$
  2. Run ABC to sample from batch posterior $\hat{\pi}(\cdot \mid d(S(x_{[b]}^{obs}), S(x_{[b]})) < \varepsilon)$
  3. Compute sample posterior variance Σ_b^{-1}
- Combine batch posterior approximations

$$\theta_j = \sum_{b=1}^{B} \Sigma_b \theta_j^{[b]} \Big/ \sum_{b=1}^{B} \Sigma_b$$

Remark Diffuse prior $\tilde{\pi}(\cdot)$ non informative calls for ABC-MCMC steps

Big parameter issues

Curse of dimension: as $\dim(\Theta) = k_\theta$ increases

- exploration of parameter space gets harder
- summary statistic η forced to increase, since at least of dimension $k_\eta \geq \dim(\Theta)$

Some solutions

- adopt more local algorithms like ABC-MCMC or ABC-SMC
- aim at posterior marginals and approximate joint posterior by copula

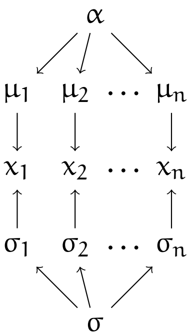
[Li et al., 2016]

- run ABC-Gibbs

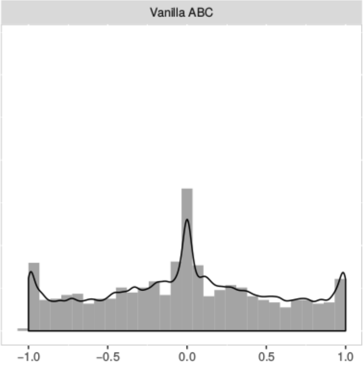
[Clarté et al., 2016]

Example 11: Hierarchical MA(2)

- $x_i \overset{\text{iid}}{\sim} \text{MA}_2(\mu_i, \sigma_i)$
- $\mu_i = (\beta_{i,1} - \beta_{i,2}, 2(\beta_{i,1} + \beta_{i,2}) - 1)$ with $(\beta_{i,1}, \beta_{i,2}, \beta_{i,3}) \overset{\text{iid}}{\sim} \mathcal{D}ir(\alpha_1, \alpha_2, \alpha_3)$
- $\sigma_i \overset{\text{iid}}{\sim} \mathcal{IG}(\sigma_1, \sigma_2)$
- $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, with prior $\mathcal{E}(1)^{\otimes 3}$
- $\sigma = (\sigma_1, \sigma_2)$ with prior $\mathcal{C}(1)^{+\otimes 2}$



Example 11: Hierarchical MA(2)



ⓒ Based on 10^6 prior and 10^3 posteriors simulations, $3n$ summary statistics, and series of length 100, ABC-Rej posterior hardly distinguishable from prior!

ABC-Gibbs

When parameter decomposed into $\theta = (\theta_1, \dots, \theta_n)$

Algorithm 5 ABC-Gibbs sampler

starting point $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$, observation x^{obs}
for $i = 1, \dots, N$ **do**
  for $j = 1, \dots, n$ **do**
    $\theta_j^{(i)} \sim \pi_{\varepsilon_j}(\cdot \mid x^\star, s_j, \theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_n^{(i-1)})$
  end for
end for

Divide & conquer:

- one tolerance ε_j for each parameter θ_j
- one statistic s_j for each parameter θ_j

[Clarté et al., 2019]

ABC-Gibbs

When parameter decomposed into $\theta = (\theta_1, \dots, \theta_n)$

Algorithm 6 ABC-Gibbs sampler

starting point $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$, observation x^{obs}
for $i = 1, \dots, N$ **do**
  for $j = 1, \dots, n$ **do**
    $\theta_j^{(i)} \sim \pi_{\varepsilon_j}(\cdot \mid x^\star, s_j, \theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_n^{(i-1)})$
  end for
end for

Divide & conquer:

- one tolerance ε_j for each parameter θ_j
- one statistic s_j for each parameter θ_j

[Clarté et al., 2019]

Compatibility

When using ABC-Gibbs conditionals with different acceptance events, e.g., different statistics

$$\pi(\alpha)\pi(s_\alpha(\mu) \mid \alpha) \text{ and } \pi(\mu)f(s_\mu(x^\star) \mid \alpha, \mu).$$

conditionals are incompatible

- ABC-Gibbs does not necessarily converge (even for tolerance equal to zero)
- potential limiting distribution
  - not a genuine posterior (double use of data)
  - unknown [except for a specific version]
  - possibly far from genuine posterior(s)

[Clarté et al., 2016]

Compatibility

When using ABC-Gibbs conditionals with different acceptance events, e.g., different statistics

$$\pi(\alpha)\pi(s_\alpha(\mu) \mid \alpha) \text{ and } \pi(\mu)f(s_\mu(x^\star) \mid \alpha, \mu).$$

conditionals are incompatible

- ABC-Gibbs does not necessarily converge (even for tolerance equal to zero)
- potential limiting distribution
  - not a genuine posterior (double use of data)
  - unknown [except for a specific version]
  - possibly far from genuine posterior(s)

[Clarté et al., 2016]

Convergence

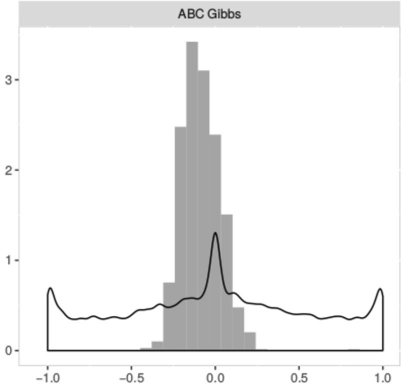
In hierarchical case $n = 2$,

Theorem If there exists $0 < \kappa < 1/2$ such that

$$\sup_{\theta_1, \tilde{\theta}_1} \|\pi_{\varepsilon_2}(\cdot \mid x^\star, s_2, \theta_1) - \pi_{\varepsilon_2}(\cdot \mid x^\star, s_2, \tilde{\theta}_1)\|_{TV} = \kappa$$

ABC-Gibbs Markov chain geometrically converges in total variation to stationary distribution ν_ε, with geometric rate $1 - 2\kappa$.

Example 11: Hierarchical MA(2)



Separation from the prior for identical number of simulations

Explicit limiting distribution

For model

$$x_j \mid \mu_j \sim \pi(x_j \mid \mu_j)\,, \qquad \mu_j \mid \alpha \overset{\text{i.i.d.}}{\sim} \pi(\mu_j \mid \alpha)\,, \qquad \alpha \sim \pi(\alpha)$$

alternative ABC based on:

$$\tilde{\pi}(\alpha, \mu \mid x^\star) \propto \pi(\alpha) q(\mu) \int \overbrace{\pi(\tilde{\mu} \mid \alpha) \mathbb{1}_{\eta(s_\alpha(\mu), s_\alpha(\tilde{\mu})) < \epsilon_\alpha} \, \mathrm{d}\tilde{\mu}}^{\text{generate a new } \mu}$$
$$\times \int f(\tilde{x} \mid \mu) \pi(x^\star \mid \mu) \mathbb{1}_{\eta(s_\mu(x^\star), s_\mu(\tilde{x})) < \epsilon_\mu} \, \mathrm{d}\tilde{x},$$

with q arbitrary distribution on μ

Explicit limiting distribution

For model

$$x_j \mid \mu_j \sim \pi(x_j \mid \mu_j)\,, \qquad \mu_j \mid \alpha \overset{\text{i.i.d.}}{\sim} \pi(\mu_j \mid \alpha)\,, \qquad \alpha \sim \pi(\alpha)$$

induces full conditionals

$$\tilde{\pi}(\alpha \mid \mu) \propto \pi(\alpha) \int \pi(\tilde{\mu} \mid \alpha) \mathbb{1}_{\eta(s_\alpha(\mu), s_\alpha(\tilde{\mu})) < \epsilon_\alpha} \, \mathrm{d}\tilde{x}$$

and

$$\tilde{\pi}(\mu \mid \alpha, x^\star) \propto q(\mu) \int \pi(\tilde{\mu} \mid \alpha) \mathbb{1}_{\eta(s_\alpha(\mu), s_\alpha(\tilde{\mu})) < \epsilon_\alpha} \, \mathrm{d}\tilde{\mu}$$
$$\times \int f(\tilde{x} \mid \mu) \pi(x^\star \mid \mu) \mathbb{1}_{\eta(s_\mu(x^\star), s_\mu(\tilde{x})) < \epsilon_\mu} \, \mathrm{d}\tilde{x}$$

now compatible with new artificial joint

Explicit limiting distribution

For model

$$x_j \mid \mu_j \sim \pi(x_j \mid \mu_j)\,, \qquad \mu_j \mid \alpha \overset{\text{i.i.d.}}{\sim} \pi(\mu_j \mid \alpha)\,, \qquad \alpha \sim \pi(\alpha)$$

that is,

- prior simulations of $\alpha \sim \pi(\alpha)$ and of $\tilde{\mu} \sim \pi(\tilde{\mu} \mid \alpha)$ until $\eta(s_\alpha(\mu), s_\alpha(\tilde{\mu})) < \epsilon_\alpha$
- simulation of μ from instrumental $q(\mu)$ and of auxiliary variables $\tilde{\mu}$ and $\tilde{x}$ until both constraints satisfied

Explicit limiting distribution

For model

$$x_j \mid \mu_j \sim \pi(x_j \mid \mu_j)\,, \qquad \mu_j \mid \alpha \overset{\text{i.i.d.}}{\sim} \pi(\mu_j \mid \alpha)\,, \qquad \alpha \sim \pi(\alpha)$$

Resulting Gibbs sampler stationary for posterior proportional to

$$\pi(\alpha, \mu) \underbrace{q(s_\alpha(\mu))}_{\text{projection}} \underbrace{f(s_\mu(x^\star) \mid \mu)}_{\text{projection}}$$

that is, for likelihood associated with $s_\mu(x^\star)$ and prior distribution proportional to $\pi(\alpha, \mu) q(s_\alpha(\mu))$ [exact!]

Incoming ABC workshops

- **[A]BayesComp**, Gainesville, Florida, Jan 7-10 2020
- **ABC in Grenoble**, France, March 18-19 2020
- **ISBA(BC)**, Kunming, China, June 26-30 2020
- **ABC in Longyearbyen**, Svalbard, April 11-13 2021

ABC postdoc positions

2 post-doc positions with the ABSint research grant:

- Focus on approximate Bayesian techniques like ABC, variational Bayes, PAC-Bayes, Bayesian non-parametrics, scalable MCMC, and related topics. A potential direction of research would be the derivation of new Bayesian tools for model checking in such complex environments.
- Terms: up to 24 months, no teaching duty attached, primarily located in Université Paris-Dauphine, with supported periods in Oxford (J. Rousseau) and visits to Montpellier (J.-M. Marin). **No hard deadline**.
- If interested, send application to me: bayesianstatistics@gmail.com