ABC Methods for Bayesian Model Choice

Christian P. Robert

Université Paris-Dauphine, IuF, & CREST
`http://www.ceremade.dauphine.fr/~xian`

Bayes-250, Edinburgh, September 6, 2011

Approximate Bayesian computation

Regular Bayesian computation issues

When faced with a non-standard posterior distribution

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) L(\theta|\mathbf{y})$$

the standard solution is to use simulation (Monte Carlo) to produce a sample

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

from $\pi(\theta|\mathbf{y})$ (or approximately by Markov chain Monte Carlo methods)

[Robert & Casella, 2004]

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

© MCMC cannot be implemented!

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]

A as approximative

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 approximative

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

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

ABC algorithm

Algorithm 1 Likelihood-free rejection sampler

```
for $i = 1$ to $N$ do
    repeat
        generate $\theta'$ 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 (maybe in-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:

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

MA example

Consider the MA(q) model

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

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 $(x'_t)_{1\leq t\leq T}$

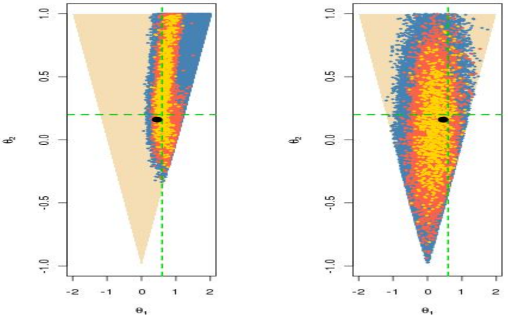
Distance: basic distance between the series

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

or between summary statistics like the first q autocorrelations

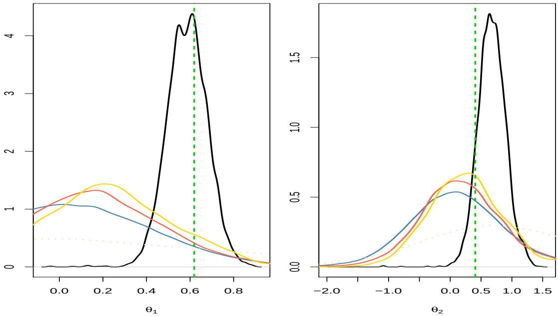
$$\tau_j = \sum_{t=j+1}^{T} x_t x_{t-j}$$

Comparison of distance impact



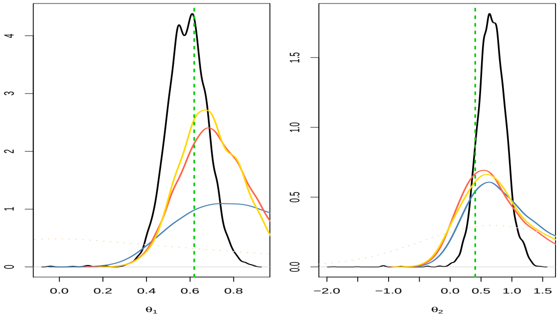
Evaluation of the tolerance on the ABC sample against both distances ($\epsilon = 100\%, 10\%, 1\%, 0.1\%$) for an MA$(2)$ model

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 for model choice

Approximate Bayesian computation

ABC for model choice

Gibbs random fields

Generic ABC model choice

Bayesian model choice

Principle

Several models $M_1, M_2, \ldots$ are considered simultaneously for dataset $\mathbf{y}$ and model index $\mathcal{M}$ central to inference.
Use of a prior $\pi(\mathcal{M} = m)$, plus a prior distribution on the parameter conditional on the value m of the model index, $\pi_m(\boldsymbol{\theta}_m)$
Goal is to derive the posterior distribution of $\mathcal{M}$, a challenging computational target when models are complex.

Generic ABC for model choice

Algorithm 2 Likelihood-free model choice sampler (ABC-MC)

```
for $t = 1$ to $T$ do
    repeat
        Generate $m$ from the prior $\pi(\mathcal{M} = m)$
        Generate $\boldsymbol{\theta}_m$ from the prior $\pi_m(\boldsymbol{\theta}_m)$
        Generate $\mathbf{z}$ from the model $f_m(\mathbf{z}|\boldsymbol{\theta}_m)$
    until $\rho\{\eta(\mathbf{z}), \eta(\mathbf{y})\} < \epsilon$
    Set $m^{(t)} = m$ and $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}_m$
end for
```

[Toni, Welch, Strelkowa, Ipsen & Stumpf, 2009]

ABC estimates

Posterior probability $\pi(\mathcal{M} = m|\mathbf{y})$ approximated by the frequency of acceptances from model m

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{I}_{m^{(t)}=m} \,.$$

Early issues with implementation:

- should tolerances ϵ be the same for all models?
- should summary statistics vary across models? incl. their dimension?
- should the distance measure ρ vary across models?

ABC estimates

Posterior probability $\pi(\mathcal{M}=m|\mathbf{y})$ approximated by the frequency of acceptances from model m

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{m^{(t)}=m}\,.$$

Early issues with implementation:

- ϵ then needs to become part of the model
- Varying statistics incompatible with Bayesian model choice proper
- ρ then part of the model

Extension to a weighted polychotomous logistic regression estimate of $\pi(\mathcal{M}=m|\mathbf{y})$, with non-parametric kernel weights

[Cornuet et al., DIYABC, 2009]

The great ABC controversy

On-going controvery in phylogeographic genetics about the validity of using ABC for testing

Against: Templeton, 2008, 2009, 2010a, 2010b, 2010c, &tc argues that nested hypotheses cannot have higher probabilities than nesting hypotheses (!)

The great ABC controversy

On-going controvery in phylogeographic genetics about the validity of using ABC for testing

Against: Templeton, 2008, 2009, 2010a, 2010b, 2010c, &tc argues that nested hypotheses cannot have higher probabilities than nesting hypotheses (!)

Replies: Fagundes et al., 2008, Beaumont et al., 2010, Berger et al., 2010, Csillèry et al., 2010 point out that the criticisms are addressed at [Bayesian] model-based inference and have nothing to do with ABC...

Potts model

Potts model

$\sum_{c\in\mathscr{C}} V_c(\mathbf{y})$ is of the form

$$\sum_{c\in\mathscr{C}} V_c(\mathbf{y}) = \theta S(\mathbf{y}) = \theta \sum_{l\sim i} \delta_{y_l=y_i}$$

where $l\sim i$ denotes a neighbourhood structure

Potts model

Potts model

$\sum_{c\in\mathscr{C}} V_c(\mathbf{y})$ is of the form

$$\sum_{c\in\mathscr{C}} V_c(\mathbf{y}) = \theta S(\mathbf{y}) = \theta \sum_{l\sim i} \delta_{y_l=y_i}$$

where $l{\sim}i$ denotes a neighbourhood structure

In most realistic settings, summation

$$Z_{\boldsymbol{\theta}} = \sum_{\mathbf{x}\in\mathcal{X}} \exp\{\boldsymbol{\theta}^{\mathsf{T}} S(\mathbf{x})\}$$

involves too many terms to be manageable and numerical approximations cannot always be trusted

[Cucala et al., 2009]

Neighbourhood relations

Setup

Choice to be made between M neighbourhood relations

$$i \overset{m}{\sim} i' \qquad (0 \leq m \leq M-1)$$

with

$$S_m(\mathbf{x}) = \sum_{i \overset{m}{\sim} i'} \mathbb{I}_{\{x_i = x_{i'}\}}$$

driven by the posterior probabilities of the models.

Model index

Computational target:

$$\mathbb{P}(\mathcal{M}=m|\mathbf{x}) \propto \int_{\Theta_m} f_m(\mathbf{x}|\theta_m)\pi_m(\theta_m)\,\mathrm{d}\theta_m\,\pi(\mathcal{M}=m)$$

Model index

Computational target:

$$\mathbb{P}(\mathcal{M}=m|\mathbf{x}) \propto \int_{\Theta_m} f_m(\mathbf{x}|\theta_m)\pi_m(\theta_m)\,\mathrm{d}\theta_m\,\pi(\mathcal{M}=m)$$

If $S(\mathbf{x})$ sufficient statistic for the joint parameters $(\mathcal{M},\theta_0,\ldots,\theta_{M-1})$,

$$\mathbb{P}(\mathcal{M}=m|\mathbf{x}) = \mathbb{P}(\mathcal{M}=m|S(\mathbf{x}))\,.$$

Sufficient statistics in Gibbs random fields

Sufficient statistics in Gibbs random fields

Each model m has its own sufficient statistic $S_m(\cdot)$ and $S(\cdot) = (S_0(\cdot), \ldots, S_{M-1}(\cdot))$ is also (model-)sufficient.

Sufficient statistics in Gibbs random fields

Each model m has its own sufficient statistic $S_m(\cdot)$ and $S(\cdot) = (S_0(\cdot), \ldots, S_{M-1}(\cdot))$ is also (model-)sufficient.
Explanation: For Gibbs random fields,

$$\begin{aligned} x|\mathcal{M} = m \sim f_m(\mathbf{x}|\theta_m) &= f_m^1(\mathbf{x}|S(\mathbf{x}))f_m^2(S(\mathbf{x})|\theta_m) \\ &= \frac{1}{n(S(\mathbf{x}))} f_m^2(S(\mathbf{x})|\theta_m) \end{aligned}$$

where

$$n(S(\mathbf{x})) = \sharp\left\{\tilde{\mathbf{x}} \in \mathcal{X} : S(\tilde{\mathbf{x}}) = S(\mathbf{x})\right\}$$

ⓒ $S(\mathbf{x})$ is therefore also sufficient for the joint parameters

Toy example

iid Bernoulli model versus two-state first-order Markov chain, i.e.

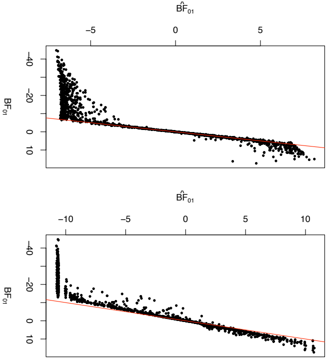
$$f_0(\mathbf{x}|\theta_0) = \exp\left(\theta_0 \sum_{i=1}^{n} \mathbb{I}_{\{x_i=1\}}\right) \bigg/ \{1+\exp(\theta_0)\}^n \,,$$

versus

$$f_1(\mathbf{x}|\theta_1) = \frac{1}{2} \exp\left(\theta_1 \sum_{i=2}^{n} \mathbb{I}_{\{x_i=x_{i-1}\}}\right) \bigg/ \{1+\exp(\theta_1)\}^{n-1} \,,$$

with priors $\theta_0 \sim \mathcal{U}(-5,5)$ and $\theta_1 \sim \mathcal{U}(0,6)$ (inspired by "phase transition" boundaries).

Toy example (2)



(left) Comparison of the true $BF_{m_0/m_1}(\mathbf{x}^0)$ with $\widehat{BF}_{m_0/m_1}(\mathbf{x}^0)$ (in logs) over $2,000$ simulations and 4.10^6 proposals from the prior. *(right)* Same when using tolerance ϵ corresponding to the 1% quantile on the distances.

Back to sufficiency

> *'Sufficient statistics for individual models are unlikely to be very informative for the model probability. This is already well known and understood by the ABC-user community.'*
>
> [Scott Sisson, Jan. 31, 2011, 'Og]

Back to sufficiency

> *'Sufficient statistics for individual models are unlikely to be very informative for the model probability. This is already well known and understood by the ABC-user community.'*

[Scott Sisson, Jan. 31, 2011, 'Og]

If $\eta_1(\mathbf{x})$ sufficient statistic for model $m = 1$ and parameter θ_1 and $\eta_2(\mathbf{x})$ sufficient statistic for model $m = 2$ and parameter θ_2, $(\eta_1(\mathbf{x}), \eta_2(\mathbf{x}))$ is not always sufficient for (m, θ_m)

Back to sufficiency

> *'Sufficient statistics for individual models are unlikely to be very informative for the model probability. This is already well known and understood by the ABC-user community.'*

[Scott Sisson, Jan. 31, 2011, 'Og]

If $\eta_1(\mathbf{x})$ sufficient statistic for model $m = 1$ and parameter θ_1 and $\eta_2(\mathbf{x})$ sufficient statistic for model $m = 2$ and parameter θ_2, $(\eta_1(\mathbf{x}), \eta_2(\mathbf{x}))$ is not always sufficient for (m, θ_m)

ⓒ **Potential loss of information at the testing level**

Limiting behaviour of B_{12} $(T \to \infty)$

ABC approximation

$$\widehat{B_{12}}(\mathbf{y}) = \frac{\sum_{t=1}^{T} \mathbb{I}_{m^t=1}\, \mathbb{I}_{\rho\{\eta(\mathbf{z}^t),\eta(\mathbf{y})\}\leq\epsilon}}{\sum_{t=1}^{T} \mathbb{I}_{m^t=2}\, \mathbb{I}_{\rho\{\eta(\mathbf{z}^t),\eta(\mathbf{y})\}\leq\epsilon}}\,,$$

where the (m^t, z^t)'s are simulated from the (joint) prior

Limiting behaviour of B_{12} $(T \to \infty)$

ABC approximation

$$\widehat{B_{12}}(\mathbf{y}) = \frac{\sum_{t=1}^{T} \mathbb{I}_{m^t=1}\, \mathbb{I}_{\rho\{\eta(\mathbf{z}^t),\eta(\mathbf{y})\}\leq\epsilon}}{\sum_{t=1}^{T} \mathbb{I}_{m^t=2}\, \mathbb{I}_{\rho\{\eta(\mathbf{z}^t),\eta(\mathbf{y})\}\leq\epsilon}},$$

where the (m^t, z^t)'s are simulated from the (joint) prior
As T go to infinity, limit

$$\begin{aligned} B_{12}^{\epsilon}(\mathbf{y}) &= \frac{\int \mathbb{I}_{\rho\{\eta(\mathbf{z}),\eta(\mathbf{y})\}\leq\epsilon} \pi_1(\boldsymbol{\theta}_1) f_1(\mathbf{z}|\boldsymbol{\theta}_1)\, \mathrm{d}\mathbf{z}\, \mathrm{d}\boldsymbol{\theta}_1}{\int \mathbb{I}_{\rho\{\eta(\mathbf{z}),\eta(\mathbf{y})\}\leq\epsilon} \pi_2(\boldsymbol{\theta}_2) f_2(\mathbf{z}|\boldsymbol{\theta}_2)\, \mathrm{d}\mathbf{z}\, \mathrm{d}\boldsymbol{\theta}_2} \\ &= \frac{\int \mathbb{I}_{\rho\{\eta,\eta(\mathbf{y})\}\leq\epsilon} \pi_1(\boldsymbol{\theta}_1) f_1^{\eta}(\eta|\boldsymbol{\theta}_1)\, \mathrm{d}\eta\, \mathrm{d}\boldsymbol{\theta}_1}{\int \mathbb{I}_{\rho\{\eta,\eta(\mathbf{y})\}\leq\epsilon} \pi_2(\boldsymbol{\theta}_2) f_2^{\eta}(\eta|\boldsymbol{\theta}_2)\, \mathrm{d}\eta\, \mathrm{d}\boldsymbol{\theta}_2}, \end{aligned}$$

where $f_1^{\eta}(\eta|\boldsymbol{\theta}_1)$ and $f_2^{\eta}(\eta|\boldsymbol{\theta}_2)$ distributions of $\eta(\mathbf{z})$

Limiting behaviour of B_{12} $(\epsilon \to 0)$

When ϵ goes to zero,

$$B^{\eta}_{12}(\mathbf{y}) = \frac{\int \pi_1(\boldsymbol{\theta}_1) f^{\eta}_1(\eta(\mathbf{y})|\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{\int \pi_2(\boldsymbol{\theta}_2) f^{\eta}_2(\eta(\mathbf{y})|\boldsymbol{\theta}_2)\,\mathrm{d}\boldsymbol{\theta}_2}$$

Limiting behaviour of B_{12} $(\epsilon \to 0)$

When ϵ goes to zero,

$$B^{\eta}_{12}(\mathbf{y}) = \frac{\int \pi_1(\boldsymbol{\theta}_1) f^{\eta}_1(\eta(\mathbf{y})|\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{\int \pi_2(\boldsymbol{\theta}_2) f^{\eta}_2(\eta(\mathbf{y})|\boldsymbol{\theta}_2)\,\mathrm{d}\boldsymbol{\theta}_2}$$

Bayes factor based on the sole observation of $\eta(\mathbf{y})$

Limiting behaviour of B_{12} (under sufficiency)

If $\eta(\mathbf{y})$ sufficient statistic in both models,

$$f_i(\mathbf{y}|\boldsymbol{\theta}_i) = g_i(\mathbf{y}) f_i^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_i)$$

Thus

$$\begin{aligned}
B_{12}(\mathbf{y}) &= \frac{\int_{\Theta_1} \pi(\boldsymbol{\theta}_1) g_1(\mathbf{y}) f_1^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{\int_{\Theta_2} \pi(\boldsymbol{\theta}_2) g_2(\mathbf{y}) f_2^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_2)\,\mathrm{d}\boldsymbol{\theta}_2} \\
&= \frac{g_1(\mathbf{y}) \int \pi_1(\boldsymbol{\theta}_1) f_1^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{g_2(\mathbf{y}) \int \pi_2(\boldsymbol{\theta}_2) f_2^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_2)\,\mathrm{d}\boldsymbol{\theta}_2} = \frac{g_1(\mathbf{y})}{g_2(\mathbf{y})} B_{12}^\eta(\mathbf{y})\,.
\end{aligned}$$

[Didelot, Everitt, Johansen & Lawson, 2011]

Limiting behaviour of B_{12} (under sufficiency)

If $\eta(\mathbf{y})$ sufficient statistic in both models,

$$f_i(\mathbf{y}|\boldsymbol{\theta}_i) = g_i(\mathbf{y}) f_i^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_i)$$

Thus

$$\begin{aligned} B_{12}(\mathbf{y}) &= \frac{\int_{\Theta_1} \pi(\boldsymbol{\theta}_1) g_1(\mathbf{y}) f_1^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{\int_{\Theta_2} \pi(\boldsymbol{\theta}_2) g_2(\mathbf{y}) f_2^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_2)\,\mathrm{d}\boldsymbol{\theta}_2} \\ &= \frac{g_1(\mathbf{y})\ \int \pi_1(\boldsymbol{\theta}_1) f_1^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{g_2(\mathbf{y})\ \int \pi_2(\boldsymbol{\theta}_2) f_2^\eta(\eta(\mathbf{y})|\boldsymbol{\theta}_2)\,\mathrm{d}\boldsymbol{\theta}_2} = \frac{g_1(\mathbf{y})}{g_2(\mathbf{y})} B_{12}^\eta(\mathbf{y})\,. \end{aligned}$$

[Didelot, Everitt, Johansen & Lawson, 2011]

© No discrepancy only when cross-model sufficiency

Poisson/geometric example

Sample

$$\mathbf{x} = (x_1, \ldots, x_n)$$

from either a Poisson $\mathcal{P}(\lambda)$ or from a geometric $\mathcal{G}(p)$
Sum

$$S = \sum_{i=1}^{n} x_i = \eta(\mathbf{x})$$

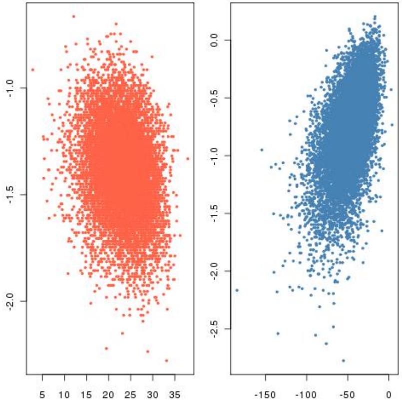
sufficient statistic for either model but not simultaneously

Discrepancy ratio

$$\frac{g_1(\mathbf{x})}{g_2(\mathbf{x})} = \frac{S! n^{-S} / \prod_i x_i!}{1 \big/ \binom{n+S-1}{S}}$$

Poisson/geometric discrepancy

Range of $B_{12}(\mathbf{x})$ versus $B^{\eta}_{12}(\mathbf{x})$: The values produced have nothing in common.



Formal recovery

Creating an encompassing exponential family

$$f(\mathbf{x}|\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\{\theta_1^{\mathsf{T}}\eta_1(\mathbf{x}) + \theta_1^{\mathsf{T}}\eta_1(\mathbf{x}) + \alpha_1 t_1(\mathbf{x}) + \alpha_2 t_2(\mathbf{x})\}$$

leads to a sufficient statistic $(\eta_1(\mathbf{x}),\eta_2(\mathbf{x}),t_1(\mathbf{x}),t_2(\mathbf{x}))$
[Didelot, Everitt, Johansen & Lawson, 2011]

Formal recovery

Creating an encompassing exponential family

$$f(\mathbf{x}|\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\{\theta_1^\mathsf{T}\eta_1(\mathbf{x})+\theta_1^\mathsf{T}\eta_1(\mathbf{x})+\alpha_1 t_1(\mathbf{x})+\alpha_2 t_2(\mathbf{x})\}$$

leads to a sufficient statistic $(\eta_1(\mathbf{x}),\eta_2(\mathbf{x}),t_1(\mathbf{x}),t_2(\mathbf{x}))$
[Didelot, Everitt, Johansen & Lawson, 2011]

In the Poisson/geometric case, if $\prod_i x_i!$ is added to S, no discrepancy

Formal recovery

Creating an encompassing exponential family

$$f(\mathbf{x}|\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\{\theta_1^{\mathsf{T}}\eta_1(\mathbf{x}) + \theta_1^{\mathsf{T}}\eta_1(\mathbf{x}) + \alpha_1 t_1(\mathbf{x}) + \alpha_2 t_2(\mathbf{x})\}$$

leads to a sufficient statistic $(\eta_1(\mathbf{x}), \eta_2(\mathbf{x}), t_1(\mathbf{x}), t_2(\mathbf{x}))$
[Didelot, Everitt, Johansen & Lawson, 2011]

Only applies in genuine sufficiency settings...

© **Inability to evaluate loss brought by summary statistics**

The Pitman–Koopman lemma

Efficient sufficiency is not such a common occurrence:

Lemma

A necessary and sufficient condition for the existence of a sufficient statistic with fixed dimension whatever the sample size is that the sampling distribution belongs to an exponential family.

[Pitman, 1933; Koopman, 1933]

The Pitman–Koopman lemma

Efficient sufficiency is not such a common occurrence:

Lemma

A necessary and sufficient condition for the existence of a sufficient statistic with fixed dimension whatever the sample size is that the sampling distribution belongs to an exponential family.

[Pitman, 1933; Koopman, 1933]

Provision of fixed support (consider $\mathcal{U}(0, \theta)$ counterexample)

Meaning of the ABC-Bayes factor

> *'This is also why focus on model discrimination typically (...) proceeds by (...) accepting that the Bayes Factor that one obtains is only derived from the summary statistics and may in no way correspond to that of the full model.'*
>
> [Scott Sisson, Jan. 31, 2011, 'Og]

Meaning of the ABC-Bayes factor

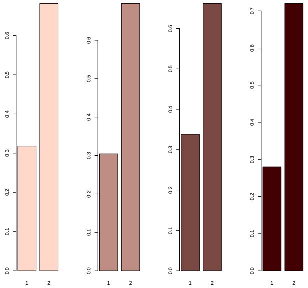
> *'This is also why focus on model discrimination typically (...) proceeds by (...) accepting that the Bayes Factor that one obtains is only derived from the summary statistics and may in no way correspond to that of the full model.'*

[Scott Sisson, Jan. 31, 2011, 'Og]

In the Poisson/geometric case, if $\mathbb{E}[y_i] = \theta_0 > 0$,

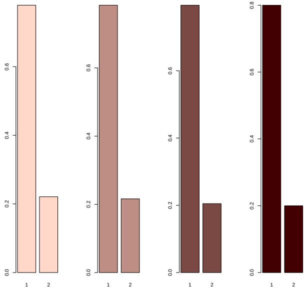
$$\lim_{n\to\infty} B_{12}^{\eta}(\mathbf{y}) = \frac{(\theta_0+1)^2}{\theta_0} e^{-\theta_0}$$

MA example



Evolution [against ϵ] of ABC Bayes factor, in terms of frequencies of visits to models MA(1) (left) and MA(2) (right) when ϵ equal to $10, 1, .1, .01\%$ quantiles on insufficient autocovariance distances. Sample of 50 points from a MA(2) with $\theta_1 = 0.6$, $\theta_2 = 0.2$. True Bayes factor equal to 17.71.

MA example



Evolution [against ϵ] of ABC Bayes factor, in terms of frequencies of visits to models MA(1) (left) and MA(2) (right) when ϵ equal to $10, 1, .1, .01\%$ quantiles on insufficient autocovariance distances. Sample of 50 points from a MA(1) model with $\theta_1 = 0.6$. True Bayes factor B_{21} equal to $.004$.

Further comments

> *'There should be the possibility that for the same model, but different (non-minimal)* [summary] *statistics (so different η's: η_1 and η_1^*) the ratio of evidences may no longer be equal to one.'*
>
> [Michael Stumpf, Jan. 28, 2011, 'Og]

Using different summary statistics [on different models] may indicate the loss of information brought by each set but agreement does not lead to trustworthy approximations.

A population genetics evaluation

Population genetics example with

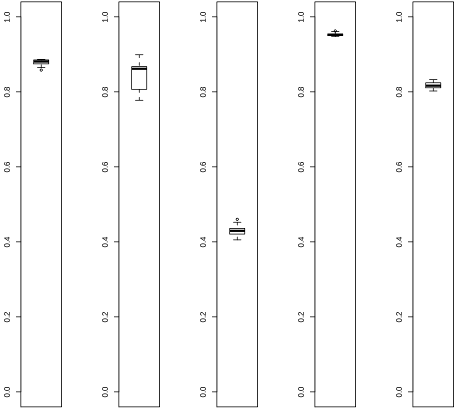
- 3 populations
- 2 scenari
- 15 individuals
- 5 loci
- single mutation parameter

A population genetics evaluation

Population genetics example with

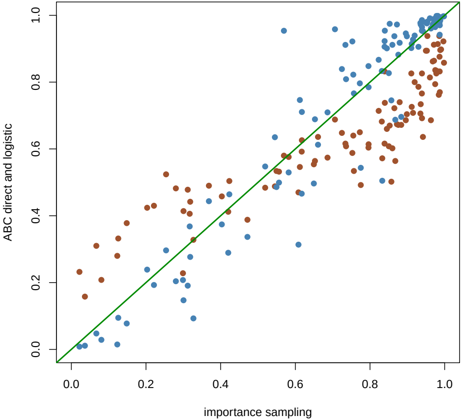
- 3 populations
- 2 scenari
- 15 individuals
- 5 loci
- single mutation parameter
- 24 summary statistics
- 2 million ABC proposal
- importance [tree] sampling alternative

Stability of importance sampling



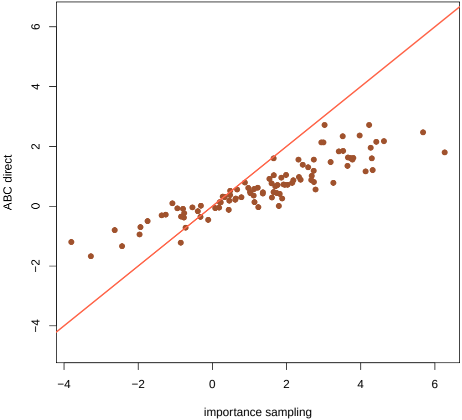
Comparison with ABC

Use of 24 summary statistics and DIY-ABC logistic correction



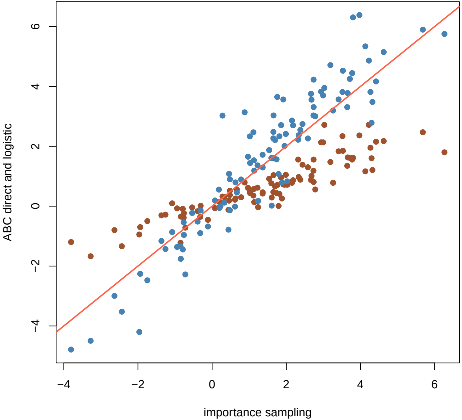
Comparison with ABC

Use of 15 summary statistics and DIY-ABC logistic correction



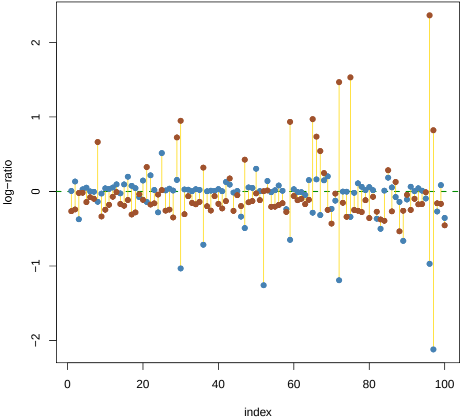
Comparison with ABC

Use of 15 summary statistics and DIY-ABC logistic correction



Comparison with ABC

Use of 15 summary statistics and DIY-ABC logistic correction



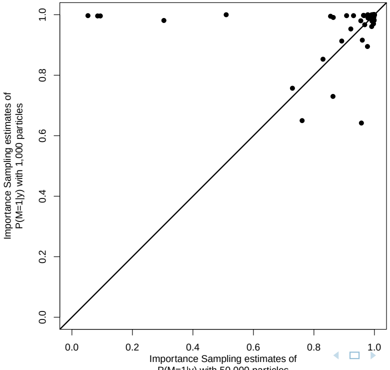
A second population genetics experiment

- three populations, two divergent 100 gen. ago
- two scenarios [third pop. recent admixture between first two pop. / diverging from pop. 1 5 gen. ago]
- In scenario 1, admixture rate 0.7 from pop. 1
- 100 datasets with 100 diploid individuals per population, 50 independent microsatellite loci.
- Effective population size of 1000 and mutation rates of 0.0005.
- 6 parameters: admixture rate ($U[0.1, 0.9]$), three effective population sizes ($U[200, 2000]$), the time of admixture/second divergence ($U[1, 10]$) and time of first divergence ($U[50, 500]$).

Results

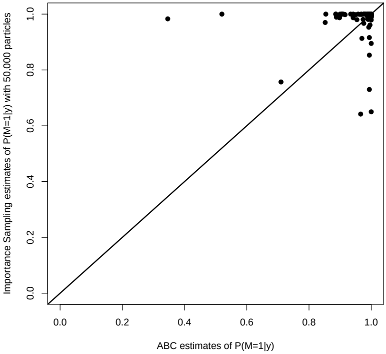
IS algorithm performed with 100 coalescent trees per particle and 50,000 particles [12 calendar days using 376 processors]

Using ten times as many loci and seven times as many individuals degrades the confidence in the importance sampling approximation because of an increased variability in the likelihood.



Results

Blurs potential divergence between ABC and genuine posterior probabilities because both are overwhelmingly close to one, due to the high information content of the data.



The only safe cases???

Besides specific models like Gibbs random fields,

using distances over the data itself escapes the discrepancy...

[Toni & Stumpf, 2010;Sousa et al., 2009]

The only safe cases???

Besides specific models like Gibbs random fields,

using distances over the data itself escapes the discrepancy...
[Toni & Stumpf, 2010;Sousa et al., 2009]

...and so does the use of more informal model fitting measures
[Ratmann, Andrieu, Richardson and Wiujf, 2009]