An introduction to Approximate Bayesian Computation methods

M.E. Castellanos

maria.castellanos@urjc.es

(from several works with S. Cabras, E. Ruli and O. Ratmann)

Valencia, January 28, 2015

Valencia Bayesian Research Group

Departamento de Informática y Estadística

Introduction to Approximate Bayesian Computation (ABC)

- ABC is a relative recent computational technique to approximate posterior distributions with the only requirement of being able to sample from the model (likelihood):

$$f(\cdot \mid \boldsymbol{\theta})$$

- First ABC ideas where mentioned by Donal Rubin (Annals of Statistics, 1984), also Diggle and Gratton in 1984 (JRSS B) proposed to use systematic simulation to approximate the likelihood function.
- The first paper proposing ABC to approximate posterior distributions in a Bayesian context, was in the field of population genetics about 18 years ago (Tavaré et al., 1997, Genetics).

Computation in Bayesian Inference

- Bayesian inference involves the estimation of a conditional probability density.
- The expert defines a model for observable given parameters (parametric inference): $f(y \mid \boldsymbol{\theta})$, and a prior distribution for parameters, $\pi(\boldsymbol{\theta})$.
- Using Bayes Theorem, the aim is to compute the posterior distribution for $\boldsymbol{\theta}$

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(\mathbf{y})},$$

where $f(\mathbf{y}) = \int f(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$.
Such marginal density in general is difficult to be calculated, because it is a high dimensional integral.

Methods of computation in Bayesian Inference

Different computation/simulation methods have been proposed in literature to approximate posterior and marginal distributions*:

- Monte Carlo methods, such as Markov Chain Monte Carlo (MCMC);
- Importance sampling (IS);
- Sequential Monte Carlo (SMC)

When the likelihood is intractable, it is not possible to evaluate $L(\boldsymbol{\theta} \mid \mathbf{y}) = f(\mathbf{y} \mid \boldsymbol{\theta})$, these standard Monte Carlo techniques do not apply.
ABC methods are Monte Carlo techniques developed for use with completely intractable likelihood, that is when $f(\mathbf{y} \mid \boldsymbol{\theta})$ can not be evaluated.

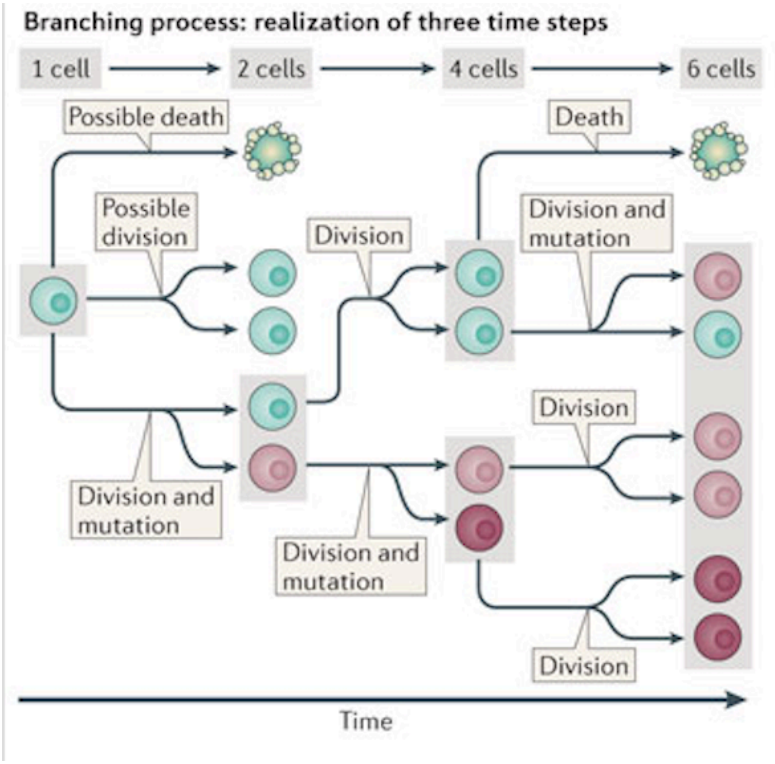
*Robert and Casella, 2004

Example: birth-death-mutation process

- Many epidemic transmission process can be represented by a birth-death-mutation process (Tanaka et al. 2006).
- It consists on a continuous-time stochastic model describing the growth in the number of infectious cases of a disease over time.

  Birth: represents the occurrence of a new infection;
  Death: corresponds to death of the host;
  Mutation: allows different genotypes of the pathogen.
- Assuming some epidemiological properties, it is possible to describe the probabilities of transition in the continuous-time process, using three rates: Birth rate, α; death rate, δ; and mutation rate, θ (per unit of time).

Example: birth-death-mutation process



Example: birth-death-mutation process

- Let the observable variable be $X_i(t)$ = number of infected with genotype i and:
- Let be $P_{i,x}(t) = P(X_i(t) = x)$
- It is possible to express the time evolution of $P_{i,x}(t)$ through the differential equation:

$$\frac{dP_{i,x}(t)}{dt} = -(\alpha+\delta+\theta)xP_{i,x}(t)+\alpha(x-1)P_{i,x-1}(t)+(\delta+\theta)(x+1)P_{i,x+1}(t)$$

- Similar equations account for the creation of new genotypes, or the total number of cases.

Example: Simulation of the birth-death-mutation process

- $G(t)$ is the number of distinct genotypes at current time t.
- $N(t) = \sum_{i=1}^{G(t)} X_i(t)$ is the total number of infected.
- Type of event simulation:

$$P(\text{ birth } | \text{ event }) = \frac{\alpha}{\alpha + \delta + \theta}$$

$$P(\text{ death } | \text{ event }) = \frac{\delta}{\alpha + \delta + \theta}$$

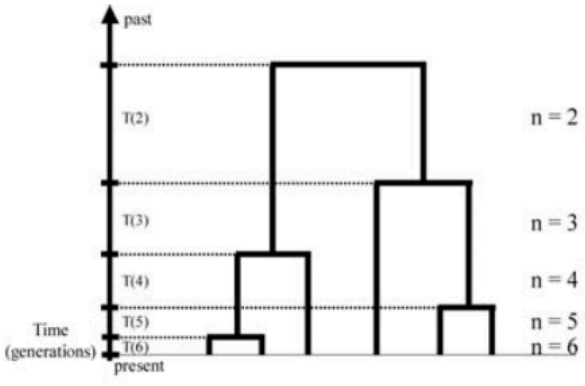
$$P(\text{ mutation } | \text{ event }) = \frac{\theta}{\alpha + \delta + \theta}$$

  If event= mutation, $G(t) = G(t) + 1$, and $X_{G(t)} = 1$.
- Given an event of the three types,

$$P(\text{ occurrence in genotype } i \mid \text{ event }) = \frac{X_i(t)}{N(t)}$$

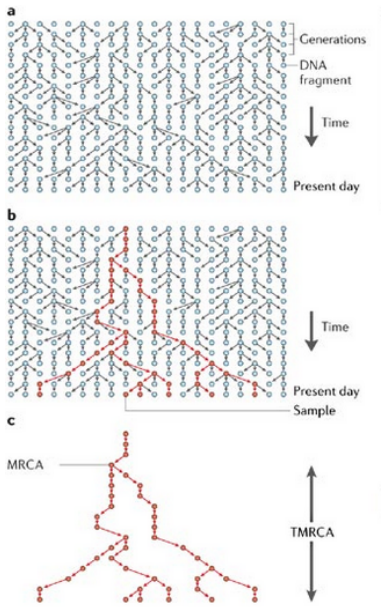
Other Example: Coalescent Model

- This is a model used in population genetics.
- Given a sample of n genes, this model could be use to know how long we must go backward in generations (6 in the figure) to share a common ancestor (TMRCA).



Example: Coalescent Model

This model is used to estimate the time to the common ancestor, but also other characteristics as the effective mutation rate, θ, from the observed data, as the number of segregating sites.



Estimation the mutation rate

- Using recombination rules from genetics, then random fluctuation in allele frequencies can be expressed in terms of probability.
- For a set of n DNA sequences, where the aim is the estimation of the effective mutation rate $\theta > 0$, under the infinitely-many-sites model assumption.
- In this model, mutations occur at rate θ at DNA sites that have not been hit by mutation before.
- If a site is affected by mutation, it is said to be segregating in the sample.
- Data consists in $s =$ number of segregating sites.

Estimation the mutation rate. Computer Model

- The generating mechanism for s, under the assumptions above, is the following:
  1. Generate L_n, the total length of the branches in the genealogical tree, with $L_n = \sum_{j=2}^{n} jT_j$.
  2. In a wide range of models in population genetics, the inter-coalescence times, T_j, can be expressed as independent random variables distributed exponential with rate $\mu_j = j(j-1)/2$, so L_n has

$$E(L_n) = 2\sum_{j=1}^{n-1} \frac{1}{j}$$

$$Var(L_n) = 4\sum_{j=1}^{n-1} \frac{1}{j^2}$$

  3. Generate $(S \mid \theta, L_n) \sim Poisson(\theta L_n/2)$.
     The election of this rate in the Poisson is to verify that $E(S \mid \theta) = \theta \sum_{j=1}^{n-1} \frac{1}{j}$

Example: Coalescent Model. Likelihood

- The likelihood $f(\cdot \mid \theta)$ is given by the marginal density of $(S \mid \theta)$ with respect to L_n, which has a closed form only for $n = 2$ as $T_2 \sim Exp(1/2)$.
- For large n, it is easy simulate from the model, but there is no closed expression of the likelihood.
- Once s is observed, we can do inference about θ, using ABC methods.

Original ABC works as follows:

Target: To approximate via simulation $\pi(\boldsymbol{\theta} \,|\, \mathbf{y}) \propto \pi(\boldsymbol{\theta}) f(\mathbf{y} \,|\, \boldsymbol{\theta})$.

When $f(\mathbf{y} \,|\, \boldsymbol{\theta})$ has not closed expression, then original version of ABC can be used:

ABC algorithm

Suppose data is $\mathbf{y}$ from model $f(\mathbf{y} \,|\, \boldsymbol{\theta})$. Under the prior $\pi(\boldsymbol{\theta})$, simulate jointly

$$\boldsymbol{\theta}^* \sim \pi(\boldsymbol{\theta}), \mathbf{z} \sim f(\mathbf{z} \,|\, \boldsymbol{\theta}^*)$$

until $\mathbf{z} = \mathbf{y}$.

(Tavaré et al. 1997)

ABC works:

It is based on Acception-Rejection:

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

Approximate in ABC because...

When $\mathbf{y}$ is a continuous random variable, the event $\mathbf{z} = \mathbf{y}$ has probability zero! So, the equality is replaced with a tolerance condition:

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

where ρ is a distance.
In this case, simulations are distributed according to:

$$\pi(\boldsymbol{\theta}) P(\rho(\mathbf{y}, \mathbf{z}) \leq \epsilon \,|\, \boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta} \,|\, \rho(\mathbf{y}, \mathbf{z}) \leq \epsilon)$$

(Pritchard et al. 1999)

ABC algorithm

1. $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$
2. $\mathbf{z} \,|\, \boldsymbol{\theta} \sim f(\mathbf{y} \,|\, \boldsymbol{\theta})$
3. if $\rho(\mathbf{z}, \mathbf{y}) < \epsilon$, retain $\boldsymbol{\theta}$ (indirect evaluation of the likelihood)

- If $\epsilon = 0$ this algorithm is exact and gives draws from the posterior distribution.
- Whereas as $\epsilon \to \infty$, the algorithm gives draws from the prior.
- Smaller values of ϵ produce samples that approximate better the posterior, but,
- It results in lower acceptance rates in step 3, that using larger values.

Extensions to use summary statistics

- When data is high dimensional, a standard change is to summarize the model output and data, using a summary statistic $s(\cdot)$ to work in a low dimensional space.
- In this case the step 3 is:
  3. if $\rho(s(\mathbf{z}), s(\mathbf{y})) < \epsilon$, retain $\boldsymbol{\theta}$.

How are distributed the simulations?...

Denote $\mathbf{s} = s(\mathbf{z})$ and the observed statistic $\mathbf{s}_{obs} = s(\mathbf{y})$.

The above ABC algorithm samples from the joint distribution:

$$\pi^{\epsilon}(\boldsymbol{\theta}, \mathbf{s} \mid \mathbf{s}_{obs}) \propto \pi(\boldsymbol{\theta}) f(\mathbf{s} \mid \boldsymbol{\theta}) \mathbb{I}_{\rho(\mathbf{s}, \mathbf{s}_{obs}) < \epsilon}$$

where $\mathbb{I}_{\rho(\mathbf{s}, \mathbf{s}_{obs}) < \epsilon}$ is the indicator for the event

$$\{\mathbf{s} \in \mathcal{S} \mid \rho(\mathbf{s}_{obs}, \mathbf{s}) < \epsilon\}$$

So ABC algorithm approximates the posterior for $\boldsymbol{\theta}$ using:

$$\pi^{\epsilon}(\boldsymbol{\theta} \mid \mathbf{s}_{obs}) = \int_{\mathcal{S}} \pi^{\epsilon}(\boldsymbol{\theta}, \mathbf{s} \mid \mathbf{s}_{obs}) d\mathbf{s} \approx \pi(\boldsymbol{\theta} \mid \mathbf{y})$$

The idea is that a small ϵ coupled with suitable summary statistics provide a good approximation of the posterior

Some comments about $s(\cdot)$

- Ideally, $s(\cdot)$ should be sufficient for $\boldsymbol{\theta}$.
- But, in real problems, if the likelihood is unknown, sufficient statistics cannot be identified.
- Summarizing the data and model output through non-sufficient summaries adds another layer of approximation.
- It is not known what effect any given choice for $s(\cdot)$ has on the approximation.

Extensions to basic ABC

- As simulating from the prior is often poor in efficiency, Marjoram et al, 2003 extend the rejection algorithm to MCMC algorithms.
- Sisson et al., 2007 propose the use of approximate sequential Monte Carlo algorithms.
- Beaumont et al., 2002 extend the ABC using a weighting scheme instead of the 0-1 of the acceptation-rejection method. Then the weighted sample is used to train a local-linear regression to model the posterior distribution.

Introduction: The MCMC ABC

The MCMC-ABC [**] works as follows for a certain proposal $q(\cdot)$ at step t:

1. $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)})$;
2. $\mathbf{s}|\boldsymbol{\theta}^* \sim f(\mathbf{s}(y) \mid \boldsymbol{\theta}^*)$
3. accept $\boldsymbol{\theta}^*$ with probability

$$\max \left\{ \frac{\pi(\boldsymbol{\theta}^*) q(\boldsymbol{\theta}^{(t-1)} \mid \boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(t-1)}) q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}_{t-1})} \mathbb{I}_{\rho(\mathbf{s}_{obs},\mathbf{s})<\epsilon}, 1 \right\},$$

which:

- does not involve direct evaluation of the likelihood $f(\mathbf{y} \mid \boldsymbol{\theta})$.
- works with $\pi(\boldsymbol{\theta})$ improper.

Our aim is to find *automatically* a *good* proposal $q(\cdot)$.

[**] Marjoram et al. (PNAS, 2003)

Our proposal[***]: Construction of an automatic proposal, using a type of pseudo-likelihood.

Set

$$q(\boldsymbol{\theta}) \propto L_Q(\boldsymbol{\theta})$$

where $L_Q(\boldsymbol{\theta})$ is the Quasi Likelihood function for $\boldsymbol{\theta}$.

[***] Cabras, Castellanos and Ruli (Bayesian Analysis, 2015)

Quasi Likelihood for $p = 1$

- Let $\Psi = \Psi(\mathbf{y}; \theta) = \sum_{i=1}^{n} \psi(y_i; \theta)$ be an unbiased estimating function $E(\Psi|\theta) = 0$
- A quasi likelihood is defined as

$$L_Q(\theta) = \exp\left\{\sum_{i=1}^{n} \int_{c_0}^{\theta} A(t)\psi(y_i; t)\, dt\right\},$$

where c_0 is an arbitrary constant and $A(\theta) = \Omega(\theta)^{-1} M(\theta)$ with:

- $M(\theta) = -E\left(\frac{\partial \Psi}{\partial \theta} \mid \theta\right)$;
- $\Omega(\theta) = E(\Psi^2 \mid \theta) = \text{Var}(\Psi|\theta)$.

Quasi Likelihood for $p = 1$ (Proposition)

Suppose:

- $f(\theta) = E(s|\theta)$ is a bounded regression function with $\mid f'(\theta) \mid < \infty$
- $\sigma_R^2 = Var(s|\theta)$ be the conditional variance.
- For $\psi(s_{obs}; \theta) = s_{obs} - f(\theta)$

THEN

$$L_Q(\theta) = \phi\left(\frac{f(\theta) - s_{obs}}{\sigma_R}\right),$$

where $\phi(\cdot)$ is the density of the standard normal.

Tools to estimate

In order to complete our definition we need to estimate:

- $f(\theta)$ with $\widehat{f}(\theta)$ (e.g. a spline, GAM, ...);
- σ_R^2 with the residual regression variance $\widehat{\sigma}_R^2$.
- In the algorithm, the variance could be also no constant, $\sigma_R^2(\theta)$, it can be estimated as well as $f(\theta)$.

We use a pilot run to calculte these estimators:

1. draw M values of $s \mid \theta \sim f(\mathbf{y} \mid \theta)$ for θ in a certain regular grid;
2. regres s on θ, obtaining $\widehat{f}(\boldsymbol{\theta})$ and $\widehat{\sigma}_R^2(\theta)$.

The ABC_{ql} is an ABC-MCMC algorithm with proposal:

$$q^Q(\theta \mid \theta^{(t-1)}) = \phi\left(\frac{f(\theta) - f(\theta^{(t-1)})}{\sigma_R(\theta^{(t-1)})}\right) \mid f'(\theta) \mid .$$

ABC-MCMC with Pseudo-likelihoods ($p = 1$)

Require: f, $f'(\theta)$, $\sigma_R^2(\theta)$, or their estimates ($\widehat{f}$, $\widehat{f}'(\theta)$, $\widehat{\sigma}_R^2(\theta)$).
For $t = 1$ to T

1. Simulate
$$f^* \sim N(f(\theta^{(t-1)}), \sigma_R^2(\theta^{(t-1)}));$$
2. Set $\theta^* = \{\theta : f^{-1}(f^*) = \theta\}$;
3. Generate $s \sim f(s(y) \mid \theta^*)$;
4. Calculate $\rho = \rho(s_{obs}, s)$;
5. Calculate the derivative, $f'(\theta)$, of $f(\theta)$, at $\theta^{(t-1)}$ and θ^*;
6. With probability
$$\min\left\{1, \frac{\pi(\theta^*) q^Q(\theta^{(t-1)} \mid \theta^*)}{\pi(\theta^{(t-1)}) q^Q(\theta^* \mid \theta^{(t-1)})} \mathbb{I}_{\rho<\epsilon}\right\}$$
accept θ^* and set $\theta^{(t)} = \theta^*$, otherwise $\theta^{(t)} = \theta^{(t-1)}$

Calculating estimators: $\widehat{f}$, $\widehat{f}'(\theta)$, $\widehat{\sigma}^2_R(\theta)$

1. Consider M values $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_M)$ taken in a regular spaced grid of a suitable large subset $\tilde{\Theta} \subseteq \Theta$;
2. Generate $\tilde{s} = (\tilde{s}_1, \ldots, \tilde{s}_M)$ where $\tilde{s}_m \sim f(s(y) \mid \tilde{\theta}_m)$;
3. Regress $\tilde{s}$ on $\tilde{\theta}$ obtaining $\widehat{f}(\theta)$ and $\widehat{f}'(\theta)$ (using Splines, GAM, etc.);
4. Regress $\left\{ \log\left(\widehat{f}(\tilde{\theta}_m) - \tilde{s}_m\right)^2 \right\}_{m=1,\ldots,M}$ on $\tilde{\theta}$ obtaining $\widehat{\sigma}^2_R(\theta)$.

Example: Coalescent Model (revisited)

This model assigns, for a certain DNA sequence of length n, the probability to have y mutations given an unknown mutation rate $\theta > 0$.
Remember that the simulation model (computer model) is:

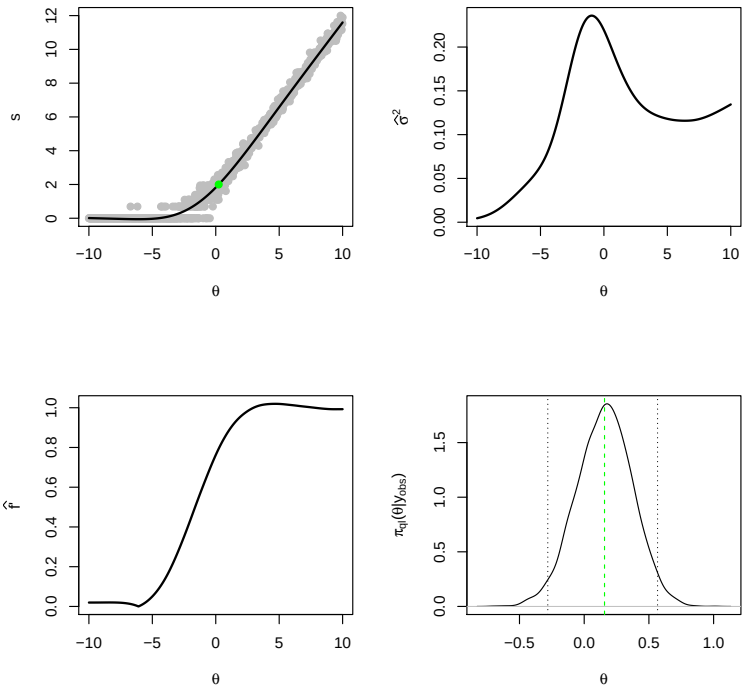
- $T_j \sim \text{Exp}(\text{mean} = 2/j(j-1))$ is the unobservable time;
- $L_n = \sum_{j=2}^{n} jT_j$ is the total length of the genealogical tree;
- $(Y \mid \theta, L_n) \sim \textit{Poisson}(\theta L_n/2)$.

$L_N(\theta)$ has a closed form only for $n = 2$.
We apply ABC_{ql} considering $s = \log(1 + y)$ and parametrization in $\log(\theta)$.
For purposes of comparison, with $n > 2$ we consider:

- A parametric approximation $\pi_{ap}(\theta|s)$ using the Poisson likelihood;
- $\pi(\theta) = \text{Exp}(1)$.

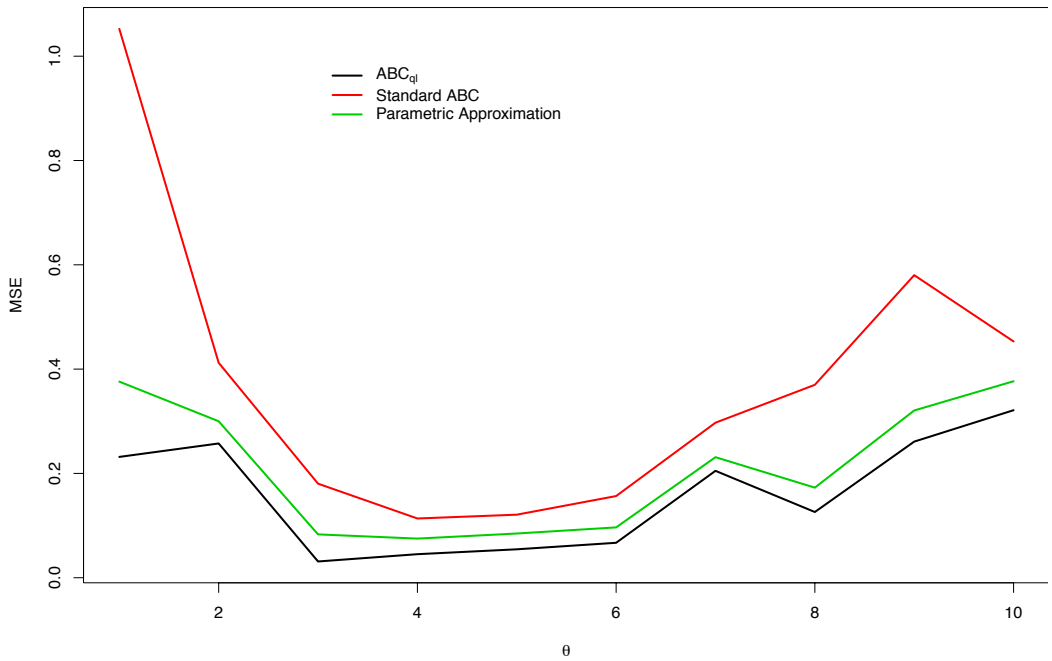
Example: Coalescent Model (cont.)

Estimation of $f(\theta)$, $f'(\theta)$, $\sigma_R^2(\theta)$:



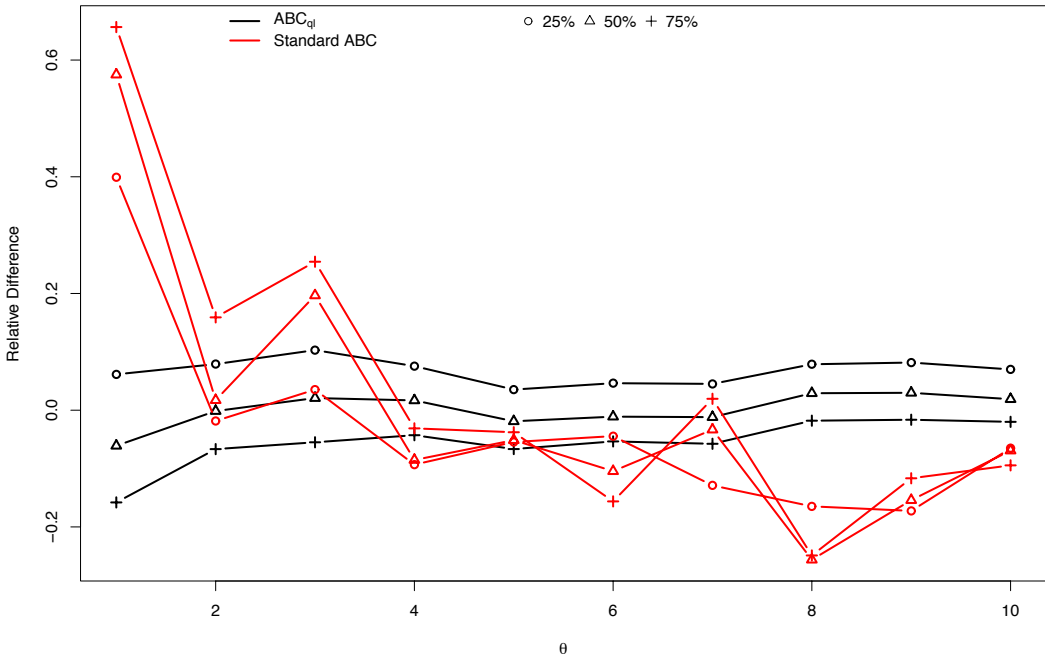
Example: Coalescent Model (cont.)

Comparison in terms of Mean Squared Error for θ for n $= 100$.



Example: Coalescent Model (cont.)

Comparison in terms of Quantile Relative Difference $(Q_p - Q_p^0)/Q_p^0$ where Q_p and Q_p^0 are the p-th quantiles of the ABC posterior and the parametric approximation.



Goodness of fit ABC models

1. GOF uses **calibrated** p-values,

$$(P - value|\text{Null Model}) \sim U(0,1).$$

2. GOF focuses on evaluating particular a given model feature: **diagnostic statistic** $T = t(\mathbf{y})$
(large values $\Rightarrow$ incompatibilities) and T possibly not ancillary w.r.t. $\boldsymbol{\theta}$.
3. The model under GOF is not $\pi(\boldsymbol{\theta}|\mathbf{y})$, but $\pi^{\epsilon}(\boldsymbol{\theta} \mid \mathbf{s}_{obs})$
(*it is the one we deal with*).

The GOF ABC evaluation: implementation

- Recall: a p-value is

$$p - value = Pr^{h(t)}(T \geq t_{obs}),$$

  - $H(T)$ is the sampling distribution of T under the model.
- $H(T)$ is usually not known exactly;
  - we approximate it by drawing T in ABC algorithm.
  - How ?

The GOF ABC evaluation: implementation

In the original ABC:

1. $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$;
2. $\mathbf{y}|\boldsymbol{\theta} \sim f(\mathbf{y} \mid \boldsymbol{\theta})$ and calculate $\mathbf{s}(\mathbf{y})$;
3. if $\rho(\mathbf{s}_{obs}, \mathbf{s}) < \epsilon$ retain $\boldsymbol{\theta}$ and $t(\mathbf{y})$.

The GOF ABC evaluation: rationale

The **conditional predictive** p-value[****]

$$p_{cpred} = Pr^{m(\cdot|\mathbf{s}_{obs})}(T(\mathbf{y}) \geq t_{obs}),$$

where

$$m(t \mid \mathbf{s}) = \int f(t \mid \mathbf{s}, \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathbf{s})d\boldsymbol{\theta},$$

$$\pi(\boldsymbol{\theta} \mid \mathbf{s}) = \frac{f(\mathbf{s} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{s} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

[****] Bayarri and Berger, (JASA, 2000)

The GOF ABC evaluation: rationale

Applying the above ABC algorithm, we are using

$$m(t \mid \mathbf{s})$$

where $\mathbf{s}$ are the statistics used in ABC. So, we are approximating p_{cpred}, and:

1. *Fact*: $p_{cpred} \sim U(0,1)$ for $n \to \infty$ if $\mathbf{s}_{obs} = \widehat{\boldsymbol{\theta}}$;[*****]
2. if $\mathbf{s}$ not ancillary $\Rightarrow$ $\mathbf{s}$ is sufficient for model $f(\mathbf{s}|\boldsymbol{\theta})$;
3. for $\epsilon \to 0$

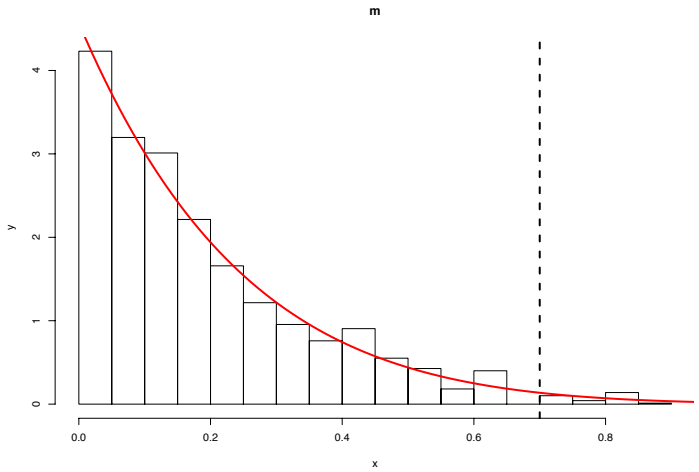
$$\Rightarrow \; f(\mathbf{s}|\boldsymbol{\theta})\mathbb{I}\{\mathbf{s} \in B_\epsilon(\mathbf{s}_{obs})\} \to f(\mathbf{s}_{obs}|\boldsymbol{\theta}).$$

4. $\Rightarrow \mathbf{s}_{obs} = \widehat{\boldsymbol{\theta}} \Rightarrow$ We are in step 1.

[*****] Fraser and Rousseau (Biometrika, 2008)

Exponential distribution

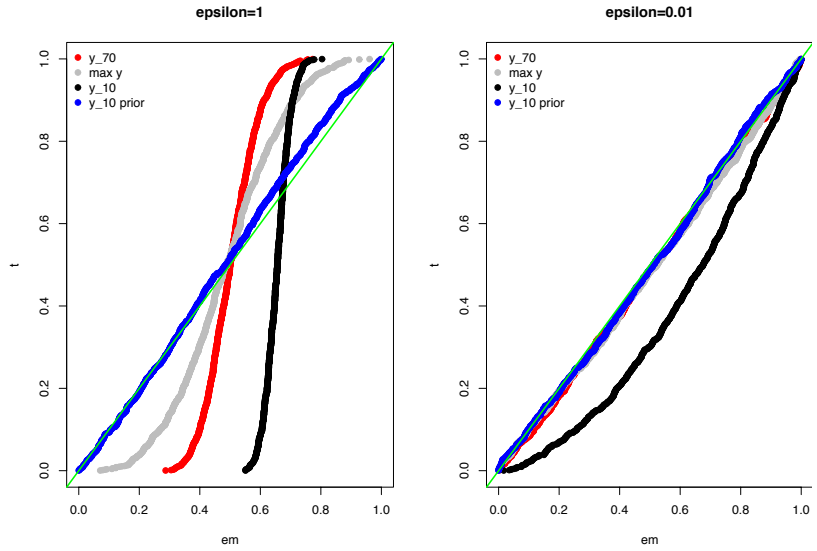
$Y \sim Exp(\theta)$, $\pi(\theta) \propto 1/\theta$, $S = 10 \cdot \bar{\mathbf{y}}$, $T = \min(\mathbf{y})$, $n = 10$
Exact $m(t|\widehat{\theta})$ (red line) and approximated $m(t|s)$ (simulations) with MCMC-ABC



ABC: p_{cpred}= 0.015 (Exact: 0.019), p_{post} = 0,048.

Uniform model. Effect of non sufficient statistics

$Y \sim U(0, \theta)$, $\pi(\theta) = U(0, 10)$, $T = \bar{\mathbf{y}}$, $n = 20$



Remarks (1/2)

1. Since we are able to simulate from $f(y \mid \theta)$ then $\widehat{f}(\theta)$ and $\widehat{\sigma}^2_R$ can be practically estimated at a desired precision;
2. More precision can be achieved by making wider the regular grid/lattice;
3. With $p > 1$ large values of M are needed because of the course of dimensionality;
4. The grid/lattice should be always enough to include the observed $\mathbf{s}_{obs}$
5. $\widehat{f}(\theta)$ can be any estimator which provides smooth functions.

Remarks (2/2)

1. For not injective $f(\theta)$ one could consider to estimate it, separately, on those subspaces of $\mathbb{R} \times \Theta$ in which it is monotone;
2. The inverse $\widehat{f}^{-1}(f^*)$ can be either obtained analytically or with the bisection method on $\widehat{f}'(\theta) = f^*$ or by numerical minimization of $(\widehat{f}'(\theta) - f^*)^2$, e.g. by a Newton-Rapson algorithm;
3. In oder to fix ϵ it would be enough to draw samples of $\boldsymbol{\theta}$ from $q(\boldsymbol{\theta})$ and set ϵ as some percentile of the empirical distribution of the distances $\rho(s_1, s_{obs}), \ldots, \rho(s_K, s_{obs})$.

Conclusions

- ABC_{ql} relies on the reparametrization $f(\boldsymbol{\theta})$ which that relates $\boldsymbol{\theta}$ with $\mathbf{s}$.
  - ...other authors****** suggest that $f(\boldsymbol{\theta})$ should be te posterior quantities of interest as $E_{\pi_N(\boldsymbol{\theta}|y)}(\boldsymbol{\theta})$.
- ABC_{ql} mitigates the importance of the choice of $\mathbf{s}$.
- To implement ABC_{ql} we need just basic knowledge of regression analysis to: have a sensible choice of $\mathbf{s}$ (at least avoid ancillarities).
- We need just standard numerical routines to calculate inverse and Jacobian.
- The only important tuning parameter remains ϵ.

******Fearnhead, P. and D. Prangle (JRRS-B, 2012).

Thanks !!!

Main References - ABC

- Tavaré S., D. J. Balding, R. C. Griffiths, and P. Donnelly (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145** (2), 505-18.
- Biau, G., Cérou, F., and Guyader, A. (2012). New insights into approximate bayesian computation. *arXiv preprint* arXiv:1207.6461.
- Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *JRRS B* **74** (3), 419-474.
- Mengersen, K., P. Pudlo, and C. Robert (2012). Approximate bayesian computation via empirical likelihood. *PNAS*, 2013