A modelling approach to ABC

Richard Wilkinson

School of Mathematical Sciences
University of Nottingham

r.d.wilkinson@nottingham.ac.uk

Sustain - April 2012

Talk Plan

1. Introduction - the need for simulation based methods
2. (Generalised) ABC as continuation of the modelling process
   - Uniform ABC
   - Difficulties with ABC
   - Generalised ABC
3. Relationship of ABC to other methods
   - history-matching
   - Simon Wood's 2010 approach
   - Paul Fearnhead and Dennis Prangle's noisy-ABC
4. Concluding remarks

The need for simulation based methods

Rohrlich (1991): Computer simulation is

> *'a key milestone somewhat comparable to the milestone that started the empirical approach (Galileo) and the deterministic mathematical approach to dynamics (Newton and Laplace)'*

The need for simulation based methods

Rohrlich (1991): Computer simulation is

> *'a key milestone somewhat comparable to the milestone that started the empirical approach (Galileo) and the deterministic mathematical approach to dynamics (Newton and Laplace)'*

Scientists are building simulators (intractable ones), and fitting them to data . There are not enough house-broken Bayesians to go around. Bayesian statistics will be done without the help of statisticians.

- There is a place for simple methods that can be credibly applied.
- Likelihood methods for complex simulators are complex.
- Modelling is something that can be done well by scientists not trained in complex statistical methods.

The need for simulation based methods

Rohrlich (1991): Computer simulation is

> *'a key milestone somewhat comparable to the milestone that started the empirical approach (Galileo) and the deterministic mathematical approach to dynamics (Newton and Laplace)'*

Scientists are building simulators (intractable ones), and fitting them to data . There are not enough house-broken Bayesians to go around. Bayesian statistics will be done without the help of statisticians.

- There is a place for simple methods that can be credibly applied.
- Likelihood methods for complex simulators are complex.
- Modelling is something that can be done well by scientists not trained in complex statistical methods.

ABC methods can be crude but they have an important role to play. We can improve the application of ABC algorithms by treating them as modelling approaches.

My "definition" of intractability

$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}$$

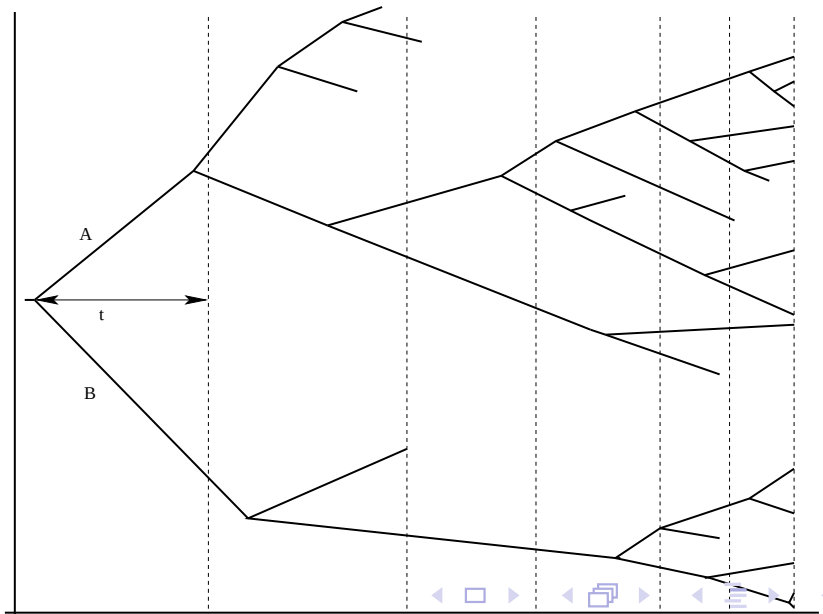
A Bayesian inference problem is intractable if

$$\pi(D|\theta)$$

is unknown and can't be evaluated (unknown is subjective). I.e., if the analytic distribution of the simulator, $f(\theta)$, run at θ is unknown. Note, this is worse than the usual normalising constant intractability

Example:

The density of the cumulative process of a branching process is unknown in general. We could probably impute everything, but this will be costly.



Uniform ABC algorithms

Uniform ABC

- Draw θ from $\pi(\theta)$
- Simulate $X \sim f(\theta)$
- Accept θ if $\rho(\mathcal{D}, X) \leq \epsilon$

For reasons that will become clear later, call this *Uniform ABC*.

Uniform ABC algorithms

Uniform ABC

- Draw θ from $\pi(\theta)$
- Simulate $X \sim f(\theta)$
- Accept θ if $\rho(\mathcal{D}, X) \leq \epsilon$

For reasons that will become clear later, call this *Uniform ABC*.

- As $\epsilon \to \infty$, we get observations from the prior, $\pi(\theta)$.
- If $\epsilon = 0$, we generate observations from $\pi(\theta \mid \mathcal{D},\ \text{PSH})$ (where PSH = perfect simulator hypothesis - no model or measurement error unless it is simulated).

ϵ reflects the tension between computability and accuracy.

The hope is that $\pi_{ABC}(\theta) \approx \pi(\theta|D,\ \text{PMH})$ for ϵ small.

Uniform ABC algorithms

Uniform ABC

- Draw θ from $\pi(\theta)$
- Simulate $X \sim f(\theta)$
- Accept θ if $\rho(\mathcal{D}, X) \leq \epsilon$

For reasons that will become clear later, call this *Uniform ABC*.

- As $\epsilon \to \infty$, we get observations from the prior, $\pi(\theta)$.
- If $\epsilon = 0$, we generate observations from $\pi(\theta \mid \mathcal{D},\ \text{PSH})$ (where PSH = perfect simulator hypothesis - no model or measurement error unless it is simulated).

ϵ reflects the tension between computability and accuracy.

The hope is that $\pi_{ABC}(\theta) \approx \pi(\theta|D,\ \text{PMH})$ for ϵ small.
There are uniform ABC-MCMC, ABC-SMC, ABC-EM, ABC-EP, ABC-MLE algorithms, etc.

Two ways of thinking

We think about linear regression in two ways

- Algorithmically: find the straight line that minimizes the sum of the squared errors.
- Probability model: we have a linear model with Gaussian errors, and we estimate the parameters using maximum-likelihood.

Two ways of thinking

We think about linear regression in two ways

- Algorithmically: find the straight line that minimizes the sum of the squared errors.
- Probability model: we have a linear model with Gaussian errors, and we estimate the parameters using maximum-likelihood.

We think about the Kalman filter in two ways:

- Algorithmically: linear quadratic estimation - find the best guess at the trajectory using linear dynamics and a quadratic penalty function
- Probability model: the (Bayesian) solution to the linear Gaussian filtering problem.

Two ways of thinking

We think about linear regression in two ways

- Algorithmically: find the straight line that minimizes the sum of the squared errors.
- Probability model: we have a linear model with Gaussian errors, and we estimate the parameters using maximum-likelihood.

We think about the Kalman filter in two ways:

- Algorithmically: linear quadratic estimation - find the best guess at the trajectory using linear dynamics and a quadratic penalty function
- Probability model: the (Bayesian) solution to the linear Gaussian filtering problem.

The same dichotomy exists for ABC.

Algorithmic view of ABC

I'd suggest that most of the early ABC developments have been in the algorithmic tradition.

1. Find a good metric, ρ - e.g., L_2 norm
2. Find a good ϵ - e.g., best 1% of simulations?
3. Find a good summary $S(D)$

The choices made are usually not motivated by modelling considerations.

Poor choices for any of these aspects can have unintended consequences.

Choice of metric ρ

Consider the following system

$$X_{t+1} = f(X_t) + N(0, \sigma^2) \tag{1}$$
$$Y_t = g(X_t) + N(0, \tau^2) \tag{2}$$

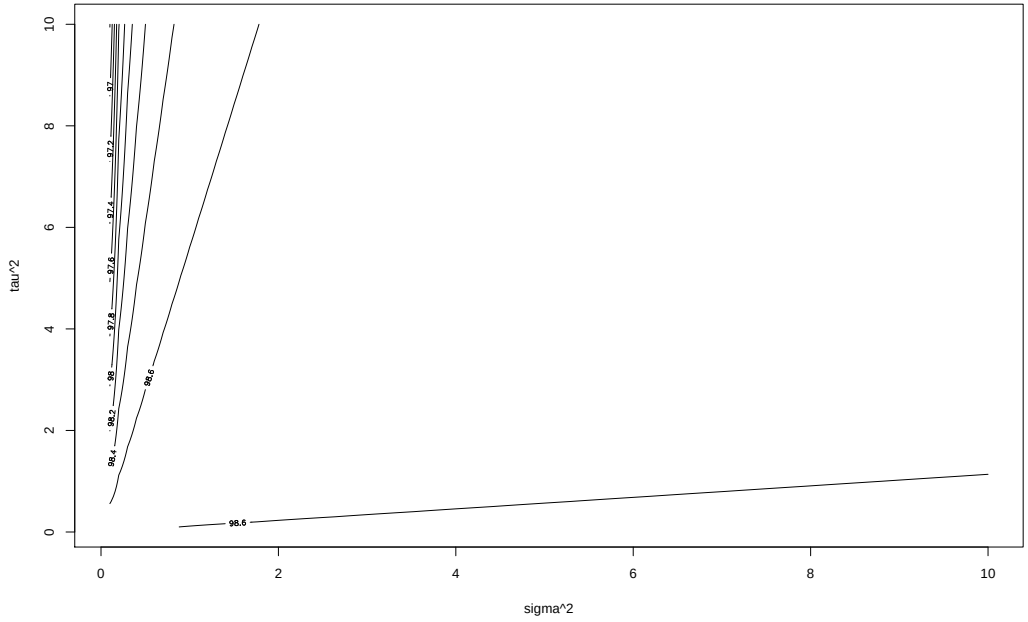
where we want to estimate measurement error τ and model error σ. Default choice of metric (or similar)

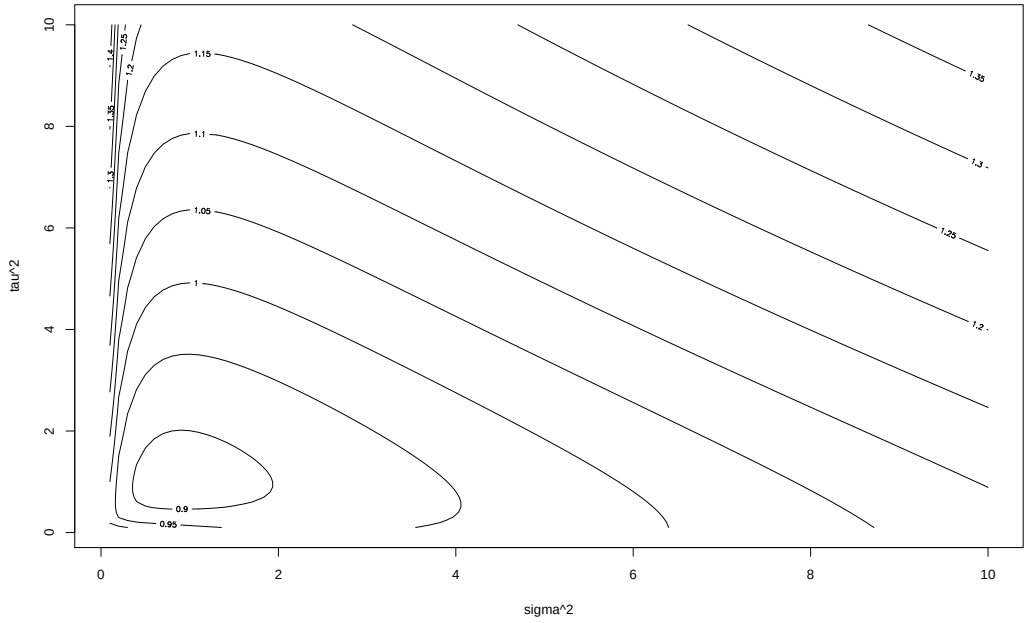
$$\rho(Y, y^{obs}) = \sum (y_t^{obs} - Y_t)^2$$

or CRPS (a proper scoring rule)

$$\rho(y^{obs}, F(\cdot)) = \sum crps(y_t^{obs}, F_t(\cdot)) = \sum_t \int (F_t(u) - \mathbb{I}_{y_t \leq u})^2 \mathrm{du}$$

where $F_t(\cdot)$ is the distribution function of $Y_t | y_{1:t-1}$.


tau^2
sigma^2
0
2
4
6
8
10
97
97.2
97.4
97.6
97.8
98
98.2
98.4
98.6
98.6


tau^2
sigma^2
0
2
4
6
8
10
0.9
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3
1.35

Choice of tolerance ϵ

Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, known variance σ^2, $\mu \sim U[a, b]$.
The mean of the data is sufficient for μ and so we can compare data sets by comparing means. Using $\rho(\mathbf{D}, \mathbf{X}) = |\bar{\mathbf{D}} - \bar{\mathbf{X}}|$ gives the following ABC:

- Pick μ from the prior distribution,
- Simulate $X_1, \ldots, X_n$ from $N(\mu, \sigma^2)$,
- Accept μ if $|\bar{X} - \bar{D}| \leq \epsilon$.

Choice of tolerance ϵ

Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, known variance σ^2, $\mu \sim U[a, b]$.
The mean of the data is sufficient for μ and so we can compare data sets by comparing means. Using $\rho(\mathbf{D}, \mathbf{X}) = |\bar{\mathbf{D}} - \bar{\mathbf{X}}|$ gives the following ABC:

- Pick μ from the prior distribution,
- Simulate $X_1, \ldots, X_n$ from $N(\mu, \sigma^2)$,
- Accept μ if $|\bar{X} - \bar{D}| \leq \epsilon$.

Calculation of $\pi_{ABC}(\mu)$ and $\pi(\mu|D)$ is possible, and we can show that

$$\mathbb{V}\mathrm{ar}_{ABC}(\mu) \approx \frac{\sigma^2}{n} + \frac{1}{3}\epsilon^2$$

$$d_{TV}(\pi_{ABC}(\mu), \pi(\mu|\mathcal{D})) \approx \frac{cn\epsilon^2}{\sigma^2} + o(\epsilon^2)$$

The tolerance required for a given accuracy depends on the size of the posterior variance σ^2/n. For small posterior variances we shall need to use a smaller value of ϵ, whereas for large variances we can use a larger ϵ.

Choice of tolerance, Jonty's palaeo example - Jake Carson

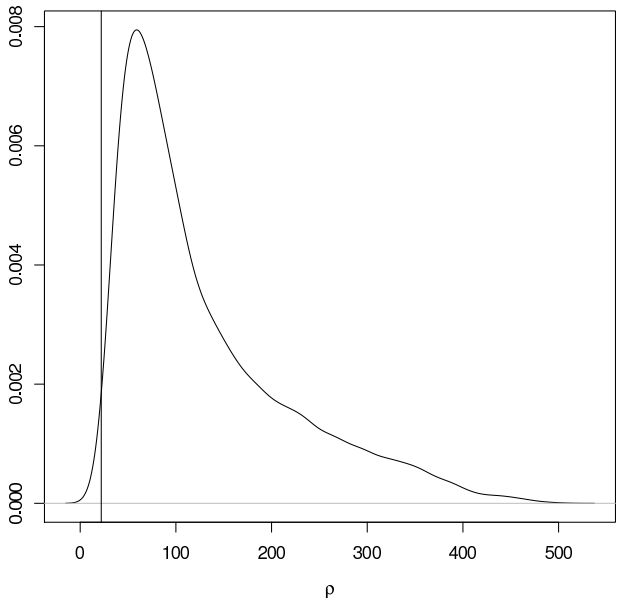
Depending on how the problem is set up, it is possible to use too-small a value for ϵ (e.g., when taking the best 1%).

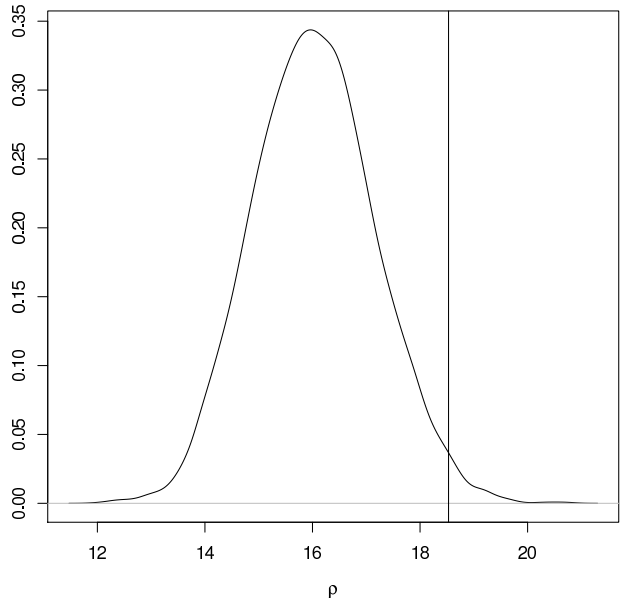
Recall the palaeoclimate example Jonty introduced yesterday. We were having difficulties with acceptance rates even in simulation studies:

$\theta \sim \pi(\theta)$; $X_{1:T} \sim SDE(\theta)$ (don't simulate measurement error)

Accept θ if $\rho(D_{1:T}, X_{1:T}) \leq \epsilon$

$\rho(D_{1:T}, X_{1:T})$ as $X_{1:T} \sim f(\hat{\theta})$ and as $D_{1:T} \sim \pi(D_{1:T}|\hat{X}_{1:T})$





Choice of summary S

ABC algorithms usually include the use of summary statistics, $S(\mathcal{D})$.

- Accept θ if $\rho(S(\mathcal{D}), S(X)) < \delta$

Considerable research effort has focused on automated methods to choose good summaries (sufficiency is not typically achievable) - great if X is some fairly homogenous field of output which we expect the model to reproduce well. Less useful if X is a large collection of different quantities.

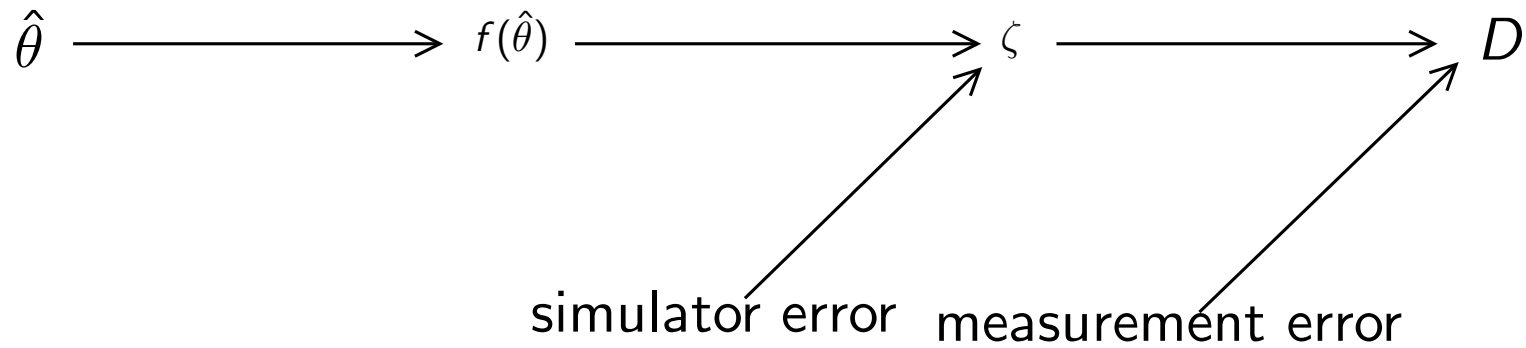
Instead ask, what aspects of the data do we expect our model to be able to reproduce? And with what degree of accuracy? $S(D)$ may be highly informative about θ, but if the model was not built to reproduce $S(D)$ then why should we calibrate to it? Cf. Simon Wood's use of phase sensitive summaries in dynamical models (although I consider the choice an inevitable modelling decision, rather than something to automate).

Calibration framework - cf Jeremy's talk

Lets now consider the probabilistic interpretation of ABC. Wilkinson 2008 showed that ABC is giving "exact" inference, but for a different model to that intended.
$\pi(\mathcal{D}|\theta)$ is not just the simulator likelihood function.

Common way of thinking:

- Relate the best-simulator run ($X = f(\hat{\theta})$) to reality ζ
- Relate reality ζ to the observations D.



See, for example, Kennedy and O'Hagan (2001) or Goldstein and Rougier (2009).

Calibration framework

Mathematically, we can write the likelihood as

$$\pi(D|\theta) = \int \pi(D|x)\pi(x|\theta)\mathrm{dx}$$

where

- $\pi(D|x)$ is a pdf relating the simulator output to reality - call it the *acceptance kernel*.
- $\pi(x|\theta)$ is the likelihood function of the simulator (ie not relating to reality)

The posterior is

$$\pi(\theta|D) = \frac{1}{Z}\int \pi(D|x)\pi(x|\theta)\mathrm{dx}.\ \pi(\theta)$$

where $Z = \int\int \pi(D|x)\pi(x|\theta)\mathrm{dx}\pi(\theta)\mathrm{d}\theta$

Calibration framework

Mathematically, we can write the likelihood as

$$\pi(D|\theta) = \int \pi(D|x)\pi(x|\theta)\mathrm{dx}$$

where

- $\pi(D|x)$ is a pdf relating the simulator output to reality - call it the *acceptance kernel*.
- $\pi(x|\theta)$ is the likelihood function of the simulator (ie not relating to reality)

The posterior is

$$\pi(\theta|D) = \frac{1}{Z}\int \pi(D|x)\pi(x|\theta)\mathrm{dx}.\ \pi(\theta)$$

where $Z = \int\int \pi(D|x)\pi(x|\theta)\mathrm{dx}\pi(\theta)\mathrm{d}\theta$

To simplify matters, we can work in joint (θ, x) space

$$\pi(\theta, x|D) = \frac{\pi(D|x)\pi(x|\theta)\pi(\theta)}{Z}$$

NB: we can allow $\pi(D|X)$ to depend on (part of) θ.

How does ABC relate to calibration?

Wilkinson 2008 and forthcoming

Consider how this relates to ABC:

$$\pi_{ABC}(\theta, x) := \pi(\theta, x|D) = \frac{\pi(D|x)\pi(x|\theta)\pi(\theta)}{Z}$$

Lets sample from this using the rejection algorithm with instrumental distribution

$$g(\theta, x) = \pi(x|\theta)\pi(\theta)$$

Note: $supp(\pi_{ABC}) \subseteq supp(g)$ and that there exists a constant $M = \frac{\max_x \pi(D|X)}{Z}$ such that

$$\pi_{ABC}(\theta, x) \leq Mg(\theta, x) \qquad \forall\, (\theta, x)$$

Generalized ABC (GABC)

Wilkinson 2008, Fearnhead and Prangle 2012

The rejection algorithm then becomes

Generalized rejection ABC (Rej-GABC)

1. $\theta \sim \pi(\theta)$ and $X \sim \pi(x|\theta)$ (ie $(\theta, X) \sim g(\cdot)$)
2. Accept (θ, X) if

$$U \sim U[0,1] \leq \frac{\pi_{ABC}(\theta, x)}{Mg(\theta, x)} = \frac{\pi(D|X)}{\max_x \pi(D|x)}$$

Generalized ABC (GABC)

Wilkinson 2008, Fearnhead and Prangle 2012

The rejection algorithm then becomes

Generalized rejection ABC (Rej-GABC)

1. $\theta \sim \pi(\theta)$ and $X \sim \pi(x|\theta)$ (ie $(\theta, X) \sim g(\cdot)$)
2. Accept (θ, X) if

$$U \sim U[0,1] \leq \frac{\pi_{ABC}(\theta, x)}{Mg(\theta, x)} = \frac{\pi(D|X)}{\max_x \pi(D|x)}$$

In uniform ABC we take

$$\pi(D|X) = \begin{cases} 1 & \text{if } \rho(D, X) \leq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

this reduces the algorithm to

2' Accept θ ifF $\rho(D, X) \leq \epsilon$

ie, we recover the *uniform* ABC algorithm.

Uniform ABC algorithm

This allows us to interpret uniform ABC. Suppose $X, D \in \mathcal{R}$

Proposition

Accepted θ from the uniform ABC algorithm (with $\rho(D, X) = |D - X|$) are samples from the posterior distribution of θ given D where we assume $D = f(\theta) + e$ and that

$$e \sim U[-\epsilon, \epsilon]$$

In general, uniform ABC assumes that

$$D|x \sim U\{d : \rho(d, x) \leq \epsilon\}$$

We can think of this as assuming a uniform error term when we relate the simulator to the observations.

Uniform ABC algorithm

This allows us to interpret uniform ABC. Suppose $X, D \in \mathcal{R}$

Proposition

Accepted θ from the uniform ABC algorithm (with $\rho(D, X) = |D - X|$) are samples from the posterior distribution of θ given D where we assume $D = f(\theta) + e$ and that

$$e \sim U[-\epsilon, \epsilon]$$

In general, uniform ABC assumes that

$$D|x \sim U\{d : \rho(d, x) \leq \epsilon\}$$

We can think of this as assuming a uniform error term when we relate the simulator to the observations.

ABC gives 'exact' inference under a different model!

Importance sampling GABC

In uniform ABC, importance sampling simply reduces to the rejection algorithm with a fixed budget for the number of simulator runs.

But for GABC it opens new algorithms:

GABC - Importance sampling

1. $\theta_i \sim \pi(\theta)$ and $X_i \sim \pi(x|\theta_i)$.
2. Give (θ_i, x_i) weight $w_i = \pi(D|x_i)$.

Importance sampling GABC

In uniform ABC, importance sampling simply reduces to the rejection algorithm with a fixed budget for the number of simulator runs.

But for GABC it opens new algorithms:

GABC - Importance sampling

1. $\theta_i \sim \pi(\theta)$ and $X_i \sim \pi(x|\theta_i)$.
2. Give (θ_i, x_i) weight $w_i = \pi(D|x_i)$.

IS-GABC has a larger effective sample size than Rej-GABC, or equivalently

$$\mathbb{V}\mathrm{ar}_{\mathrm{Rej}}(w) \geq \mathbb{V}\mathrm{ar}_{\mathrm{IS}}(w)$$

which is a Rao-Blackwell type result.
This allows us to see part of the Beaumont *et al.* 2002 as using weighted averages with a partial rejection control to estimate posterior integrals

Importance sampling GABC

In uniform ABC, importance sampling simply reduces to the rejection algorithm with a fixed budget for the number of simulator runs.

But for GABC it opens new algorithms:

GABC - Importance sampling

1. $\theta_i \sim \pi(\theta)$ and $X_i \sim \pi(x|\theta_i)$.
2. Give (θ_i, x_i) weight $w_i = \pi(D|x_i)$.

IS-GABC has a larger effective sample size than Rej-GABC, or equivalently

$$\mathbb{V}\mathrm{ar}_{\mathrm{Rej}}(w) \geq \mathbb{V}\mathrm{ar}_{\mathrm{IS}}(w)$$

which is a Rao-Blackwell type result.
This allows us to see part of the Beaumont *et al.* 2002 as using weighted averages with a partial rejection control to estimate posterior integrals
Generalisations of the MCMC-ABC and SMC-ABC algorithms to this framework are available.

Acceptance kernel $\pi(D|X)$ as an extension of modelling

Using ABC is equivalent to adding additional variability into the model.

- $\exists$ many interesting papers saying how to make this variability small
- Instead ask, given that we are stuck with this additional variability, can we use it in a useful manner, or if not, how can we make sure it does little harm?

Acceptance kernel $\pi(D|X)$ as an extension of modelling

Using ABC is equivalent to adding additional variability into the model.

- $\exists$ many interesting papers saying how to make this variability small
- Instead ask, given that we are stuck with this additional variability, can we use it in a useful manner, or if not, how can we make sure it does little harm?

How do we relate the simulator to the observations $\pi(D|S)$

- Measurement/sampling error on D
  - Measurement error may be built into the simulator. Could we remove it and use the ABC to do this?

Acceptance kernel $\pi(D|X)$ as an extension of modelling

Using ABC is equivalent to adding additional variability into the model.

- $\exists$ many interesting papers saying how to make this variability small
- Instead ask, given that we are stuck with this additional variability, can we use it in a useful manner, or if not, how can we make sure it does little harm?

How do we relate the simulator to the observations $\pi(D|S)$

- Measurement/sampling error on D
  - Measurement error may be built into the simulator. Could we remove it and use the ABC to do this?
- Discrepancy between the simulator and reality
  - In a deterministic model setting, Goldstein and Rougier 2008, and Kennedy and O'Hagan 2001 (amongst others), have offered advice for thinking about model discrepancies.
  - For statistical models, simulator error is a less clear concept.

Comments

- Prior specification of the distribution of the model error is hard!

Comments

- Prior specification of the distribution of the model error is hard!
- For many models, the variance of the model and measurement error may not be large enough to make the ABC acceptance rate high enough to generate any successes. This is telling us something has been mis-specified and we need to rethink the choices we have made.

Comments

- Prior specification of the distribution of the model error is hard!
- For many models, the variance of the model and measurement error may not be large enough to make the ABC acceptance rate high enough to generate any successes. This is telling us something has been mis-specified and we need to rethink the choices we have made.
- It can help us interpret our results, and highlight where we may being doing something undesirable with the modeling or inference.

Comments

- Prior specification of the distribution of the model error is hard!
- For many models, the variance of the model and measurement error may not be large enough to make the ABC acceptance rate high enough to generate any successes. This is telling us something has been mis-specified and we need to rethink the choices we have made.
- It can help us interpret our results, and highlight where we may being doing something undesirable with the modeling or inference.
- It can help guide our choice of metric. We don't expect any model to perfectly fit the data, and ABC can be viewed as adding in enough variability to allow a fit to be found. Given that we are adding in variability, we can control where it is placed.

Advantages of GABC

GABC

- allows us to make the inference we want to make
  - - makes explicit the assumptions about the relationship between simulator and observations.
  - - moves ABC from being an algorithm, to an extension of modelling.
- allows for new ABC algorithms, as (non-trivial) importance sampling algorithms are now possible.
- allows us to interpret the results of ABC
- allows for the possibility of more efficient ABC algorithms
  - - the 0-1 uniform cut-off is less flexible and forgiving than using generalised kernels for $\pi(D|X)$
  - - another way to view ABC is as a smoothing of the simulator likelihood

    $$\pi(D|\theta) = \int \pi(D|X)\pi(X|\theta)\mathrm{d}\theta$$

    This flattens out the simulator likelihood (cf Simon Wood's talk).

ABC as likelihood thresholding/history-matching

Consider a deterministic SSM, for example an ode, with observation error

$$x_{t+1} = f(x_t, \theta) \tag{3}$$

$$y_t = x_t + e_t \tag{4}$$

If we use uniform-ABC with

$$\rho(y, X) = \sum (y_t - x_t)^2$$

(as done for example in Toni *et al.* 2009 for the deterministic Lotka-Volterra model) then this is equivalent to assuming a Gaussian distribution for e_t, but thresholding the likelihood and only accepting θ if $\pi(y|x) > h(\epsilon)$.

If the prior for θ is uniform over some interval, then the uniform-ABC posterior will also be uniform, over the range of values which pass through the likelihood thresholding.

If we use a GABC algorithm, with $\pi(y|X) = \exp(-\rho(y, X)/\sigma^2)$ then we are assuming $e_t \sim N(0, \sigma^2)$ and we get the true posterior under this assumption (not the thresholded version).

History-matching

The history matching version would go something like

0. Build an emulator of the simulator $f(\theta)$
1. Relate the simulator to the system

$$y = f(\theta) + \epsilon$$

   where ϵ is our simulator discrepancy
2. Relate the system to the data (e represents measurement error)

$$\mathcal{D} = y + e$$

3. Declare θ implausible if

$$\| \mathcal{D} - \mathbb{E} f(\theta) \| > 3\sigma$$

   where σ^2 is the combined variance implied by the emulator, discrepancy and measurement error.

History-matching

The history matching version would go something like

0. Build an emulator of the simulator $f(\theta)$
1. Relate the simulator to the system

$$y = f(\theta) + \epsilon$$

   where ϵ is our simulator discrepancy
2. Relate the system to the data (e represents measurement error)

$$\mathcal{D} = y + e$$

3. Declare θ implausible if

$$\| \mathcal{D} - \mathbb{E}f(\theta) \| > 3\sigma$$

   where σ^2 is the combined variance implied by the emulator, discrepancy and measurement error.

If θ is not implausible we don't discard it. The result is a region of space that we can't rule out at this stage of the history-match. Usually we'd go through several stages of history matching,

Relationship between ABC and history-matching

- Using uniform priors with a deterministic simulator, and using uniform-ABC is similar to doing history matching (ignoring the use of emulators in history-matching and a few other tweeks).
- The different waves of history-matching is analogous to choosing a decreasing tolerance scheme ϵ_t in sequential-ABC methods.

The result of a history-match may be that there is no not-implausible region of parameter space (unlike the result of a likelihood-based MCMC calculation) - note the very different interpretation to ABC.

Relationship between ABC and history-matching

- Using uniform priors with a deterministic simulator, and using uniform-ABC is similar to doing history matching (ignoring the use of emulators in history-matching and a few other tweeks).
- The different waves of history-matching is analogous to choosing a decreasing tolerance scheme ϵ_t in sequential-ABC methods.

The result of a history-match may be that there is no not-implausible region of parameter space (unlike the result of a likelihood-based MCMC calculation) - note the very different interpretation to ABC.

We might find that after quantifying all the uncertainties, we can't find any parameter values that cause the model to fit the data. Similarly with GABC, we might find that no matter how careful we are in specifing the errors, we can't find any region of parameter space that fits.

- suggests we've made a mistake somewhere in the specification of the problem.
- do we want methods that fit the model regardless of the quality of the fit, or that find there are no good fits, making us think harder?

Why/when/should we include simulator discrepancy??

The importance of including a discrepancy seems to depend on whether our simulator is

- Statistical
- Physical

and also on whether our aim is

- explanation
- prediction.

and on the importance of the situation.

When should we be happy saying that our simulator is imperfect, but we're going to fit it anyway and ignore the problem?

Relationship between ABC and Simon Wood's approach

One way to view Wood 2010 is as an ABC algorithm, but using μ_θ and Σ_θ as the summary of $f(\theta)$, and assuming

$$\pi(D|S) = \exp(-1/2(D - \mu_\theta)^T \Sigma_\theta^{-1}(D - \mu_\theta))$$

A crude IS-GABC algorithm version of Wood 2010 would be

- Pick $\theta \sim \pi(\theta)$
- Simulate $s_1, \ldots, s_n \sim f(\theta)$, calculate μ_θ and Σ_θ.
- Give θ weight $w = \phi((D - \mu)^2/\Sigma)$

Wood 2010 could be seen as a MCMC-GABC algorithm.

Relationship between ABC and Simon Wood's approach

One way to view Wood 2010 is as an ABC algorithm, but using μ_θ and Σ_θ as the summary of $f(\theta)$, and assuming

$$\pi(D|S) = \exp(-1/2(D-\mu_\theta)^T \Sigma_\theta^{-1}(D-\mu_\theta))$$

A crude IS-GABC algorithm version of Wood 2010 would be

- Pick $\theta \sim \pi(\theta)$
- Simulate $s_1, \ldots, s_n \sim f(\theta)$, calculate μ_θ and Σ_θ.
- Give θ weight $w = \phi((D-\mu)^2/\Sigma)$

Wood 2010 could be seen as a MCMC-GABC algorithm.

This can be seen as accounting for the variability of the model run repeatedly at the same input, and then assuming the distribution is Gaussian. If so, running diagnostics such as QQ-plots does make sense.
Alternatively, we could see it as a way of smoothing the simulator likelihood making inference more tractable.
It is also analogous to building an emulator of a deterministic function in history matching, with the difference that in history matching the uncertainty represents lack of knowledge of the simulator output.

Noisy-ABC

Fearnhead and Prangle (2012) recently suggested a noisy-ABC algorithm:

Noisy-ABC

Initialise: Let $D' = D + e$ where $e \sim K(e)$ from some kernel $K(\cdot)$.

1. $\theta_i \sim \pi(\theta)$ and $X_i \sim \pi(x|\theta_i)$.
2. Give (θ_i, x_i) weight $w_i = K(X_i - D')$.

Noisy-ABC

Fearnhead and Prangle (2012) recently suggested a noisy-ABC algorithm:

Noisy-ABC

Initialise: Let $D' = D + e$ where $e \sim K(e)$ from some kernel $K(\cdot)$.

1. $\theta_i \sim \pi(\theta)$ and $X_i \sim \pi(x|\theta_i)$.
2. Give (θ_i, x_i) weight $w_i = K(X_i - D')$.

In my notation, this replaces the obsevered data D, with D' drawn from the acceptance kernel - $D' \sim \pi(D'|D)$

If we believe $\pi(D|X)$ relates the simulator to reality, then noisy-ABC is equivalent to adding another dose of measurement/model error to your data and using that in the inference.

The main argument in favour of noisy-ABC is that it is calibrated, unlike standard ABC.

Calibration

Calibration is a way of assessing probability statements against some idea of *truth*, a base measure $\mathbb{P}$.

- Truth is usually taken to be reality
- We are well-calibrated if $p\%$ of all predictions reported at probability p are true. Idea goes back at least to Dawid (1984).

Calibration

Calibration is a way of assessing probability statements against some idea of *truth*, a base measure $\mathbb{P}$.

- Truth is usually taken to be reality
- We are well-calibrated if $p\%$ of all predictions reported at probability p are true. Idea goes back at least to Dawid (1984).
- Calibration is a difficult idea for Bayesians (subjectivists), even when $\mathbb{P}$ is reality's measure.
  - Seidenfeld (1985) wrote

> *'Calibration in the long run is otiose, and in the short run is an inducement to hedging'.*

Calibration, when $\mathbb{P}$ is reality, is a desirable frequency property to possess, but Bayesian's should beware of consciously aiming for it.

Calibration

The definition of calibration used in FP is superficially similar:

- $\mathbb{P}_{ABC}$ is well calibrated if

$$\mathbb{P}(\theta \in A | E_q(A)) = q$$

where $E_q(A)$ is the event that the ABC posterior assigns probability q to event A

i.e., given that A is an event assigned probability q by $\mathbb{P}_{ABC}$, then we are calibrated if A occurs with probability q according to base measure $\mathbb{P}$.

Calibration

The definition of calibration used in FP is superficially similar:

- $\mathbb{P}_{ABC}$ is well calibrated if

$$\mathbb{P}(\theta \in A | E_q(A)) = q$$

where $E_q(A)$ is the event that the ABC posterior assigns probability q to event A

i.e., given that A is an event assigned probability q by $\mathbb{P}_{ABC}$, then we are calibrated if A occurs with probability q according to base measure $\mathbb{P}$.

The difference with the standard definition, is the definition of the *base-measure*, $\mathbb{P}$. In FP's definition:

- $\mathbb{P}$ does not represent reality.
- It is defined by the prior, simulator, and summary used.
  - i.e., this definition of calibration ensures you are calibrated against your own beliefs.
  - The prior is calibrated under this definition.

Further, noisy-ABC is calibrated only if we repeated the analysis with multiple noisy datasets.

Conclusions

Approximate Bayesian Computation gives exact inference for the wrong model.

- We can view ABC as implicitly defining a probability model for the relationship between the data and the simulator.
  - ABC algorithms can be considered as adding additional variability on to the model outputs.
  - Thinking in this way can help us understand what ABC is doing
- We can generalise ABC algorithms to move beyond the use of uniform error structures and use the added variation to include information about the error on the data and in the model.
  - We could (should?) account for model error.
  - Relating simulators to reality is hard, even with expert knowledge. However, most modellers have beliefs about where their simulator is accurate, and where it is not.
- If done wisely, hopefully ABC can be viewed not as an approximate form of Bayesian inference, but instead as coming closer to the inference we want to do.

Conclusions

Approximate Bayesian Computation gives exact inference for the wrong model.

- We can view ABC as implicitly defining a probability model for the relationship between the data and the simulator.
  - ABC algorithms can be considered as adding additional variability on to the model outputs.
  - Thinking in this way can help us understand what ABC is doing
- We can generalise ABC algorithms to move beyond the use of uniform error structures and use the added variation to include information about the error on the data and in the model.
  - We could (should?) account for model error.
  - Relating simulators to reality is hard, even with expert knowledge. However, most modellers have beliefs about where their simulator is accurate, and where it is not.
- If done wisely, hopefully ABC can be viewed not as an approximate form of Bayesian inference, but instead as coming closer to the inference we want to do.

Thank you for listening!