Approximate Bayesian Computation

Introduction to Bayesian Computation

- In all model-based statistical inference, the likelihood function is of central importance.
- For simple models, an analytical formula for the likelihood function can typically be derived.
- For complex models, an analytical formula might be computationally very costly to evaluate.
- ABC methods bypass the evaluation of the likelihood function.
- These methods can inevitably make assumptions and approximations whose impact still needs to be carefully assessed.

History

- ABC related ideas date back to the 1980s
- Initially, systematic simulation schemes were used to approximate the likelihood.
- Simon Tavaré was first to propose an ABC algorithm for posterior inference.
- The term Approximate Bayesian Computation was established by Mark Beaumont in his ABC-approach for problems in population genetics.

Bayes theorem

- Baye's Theorem
  - Conditional probability of a particular parameter value θ given data D to the probability of D given θ.

Likelihood

Prior

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

Posterior

The Evidence

Bayes theorem

- The Evidence
  - A collection of observations.
  - It is what is being observed and measured.

Likelihood

Prior

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

Posterior

The Evidence

Bayes theorem

- The Posterior
  - The probability of an event after taking in consideration the evidence.
  - Can be calculated be approximated using ABC.

Likelihood

Prior

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

Posterior

The Evidence

Bayes theorem

- The Prior
  - Represents beliefs about θ before D is available
  - Often specified by choosing a tractable distribution such that random generation of values of θ are straight forward.

Likelihood

Prior

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

Posterior

The Evidence

Bayes theorem

- The Likelihood
  - Refers to the process of determine the best data distribution given a specific situation in data.
  - Is used to generally maximize the chance of a particular situation to occur.
  - It is computationally expensive or sometimes completely infeasible to evaluate.

Likelihood

Prior

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

Posterior

The Evidence

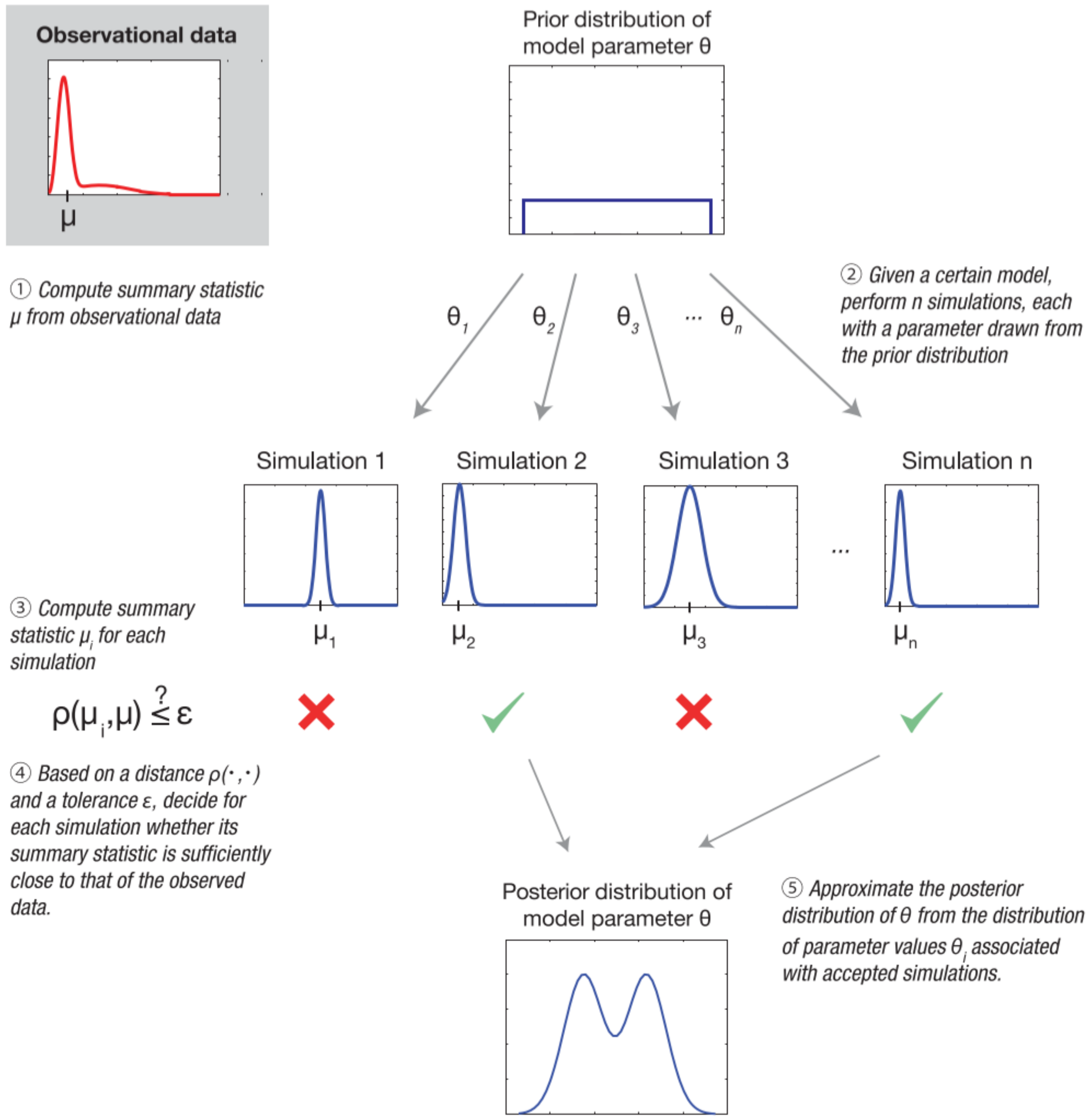
The ABC Rejection Algorithm

- Given a sampled parameter point θ, dataset $\widehat{D}$ is then simulated under the statistical model M specified by θ.
- If the generated $\widehat{D}$ is too different from the observed data D, the sampled parameter value is discarded.
- $\widehat{D}$ is accepted with tolerance $\varepsilon \geq 0$ if: $\mathrm{p}(\widehat{D}, D) \leq \varepsilon$
- The outcome of which is a sample of parameter values approximately distributed according to the desired posterior distribution without needing to evaluate the likelihood function.
- -High dimensional data without using a summary statistics the likelihood will be low.

What is a Summary Statistics?

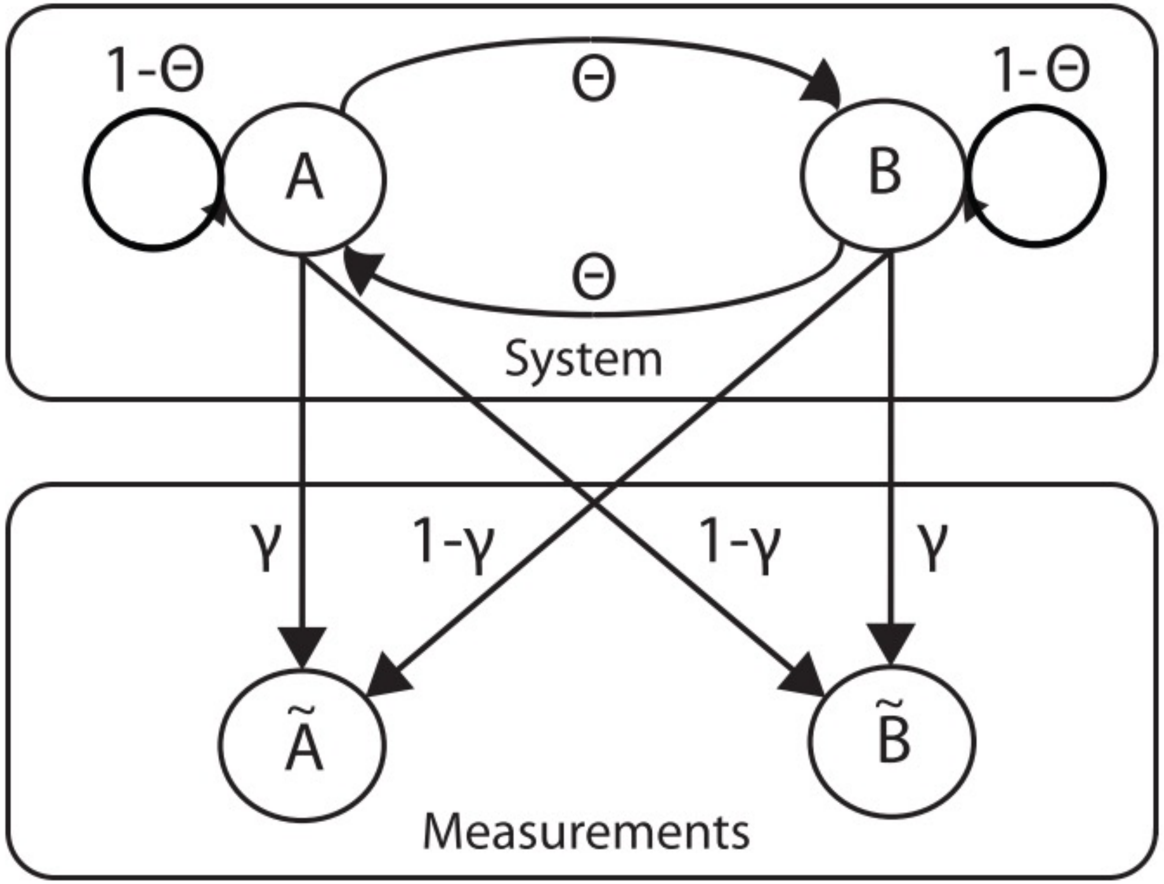
- A Summary statistic is a basic statistic
- As an example
  - Suppose we wanted to observe the heights of everyone in the university.
  - Instead of getting a list of every different height in the university we can use the mean of several close related heights to reduce the amount of dimensions we are using.
- What are other examples of summary statistics?

Conceptual Overview



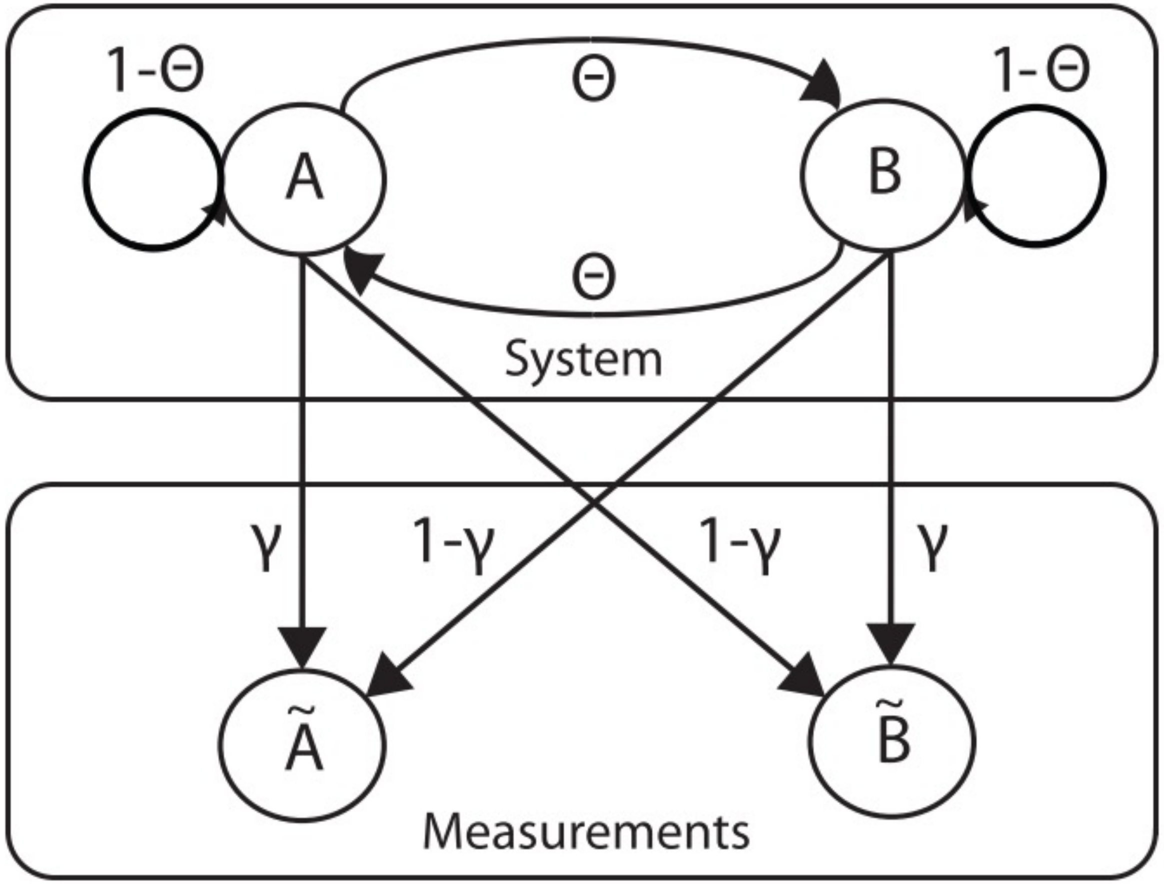
Example

- The probability to transition to states A-B, B-A is θ
- The probability to remain in each state is 1-θ
- The probability to measure the state correctly is γ.
- The probability to measure the state incorrectly is 1- γ



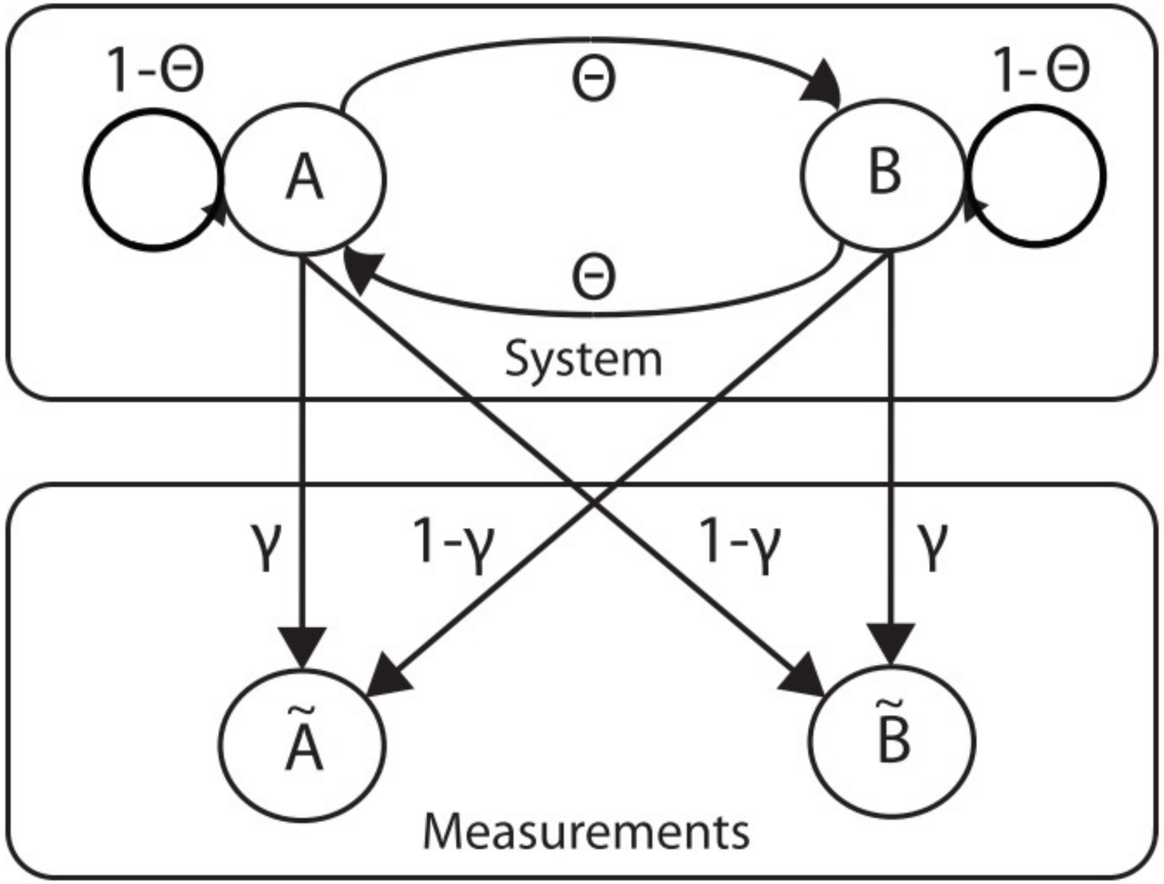
Example

- There are conditional dependencies between each state.
- Computational issue of basic ABC is due to the large dimensionality of the data.
- Can be reduced by Summary Statistic S



Example

- There are conditional dependencies between each state.
- Computational issue of basic ABC is due to the large dimensionality of the data.
- Can be reduced by Summary Statistic S



Approximating the Posterior

- Step 1: Assume that the observed data are the state sequence AAAABAABBAAAAAAABAAAA generated with θ=0.25 with a summary statistic (Number of switches between states) $\omega_E = 6$
- Step 2: Assume nothing is known about θ, a uniform prior in the interval [0,1]
  - n parameter points are drawn from the prior and the model is simulated for each of the parameter points θ_i , $i = 1, \ldots, n$, . In this example $n = 5$.
- Step 3: The summary statistic is being computed for each sequence for each sequence of simulated data, $\omega_{s,i}$ $i = 1, \ldots., n$

Approximating the Posterior from the table

- Step 4: Calculate the distance between the observed and simulated transition frequencies for all parameter points.
  - ρ($\omega_{S,i}, \omega_E$)=$|\omega_{s,i} - \omega_E|$.
  - Parameter points for which the distance is smaller than or equal to ε are accepted as approximate samples from the posterior.
- Step 5: The posterior distribution is approximated with the accepted parameter points.
  - The posterior distribution should have a non-negligible probability for parameter values in a region around the true value of θ in the system.
  - In this example, the posterior probability mass is evenly split between the values of .08 and 0.43.

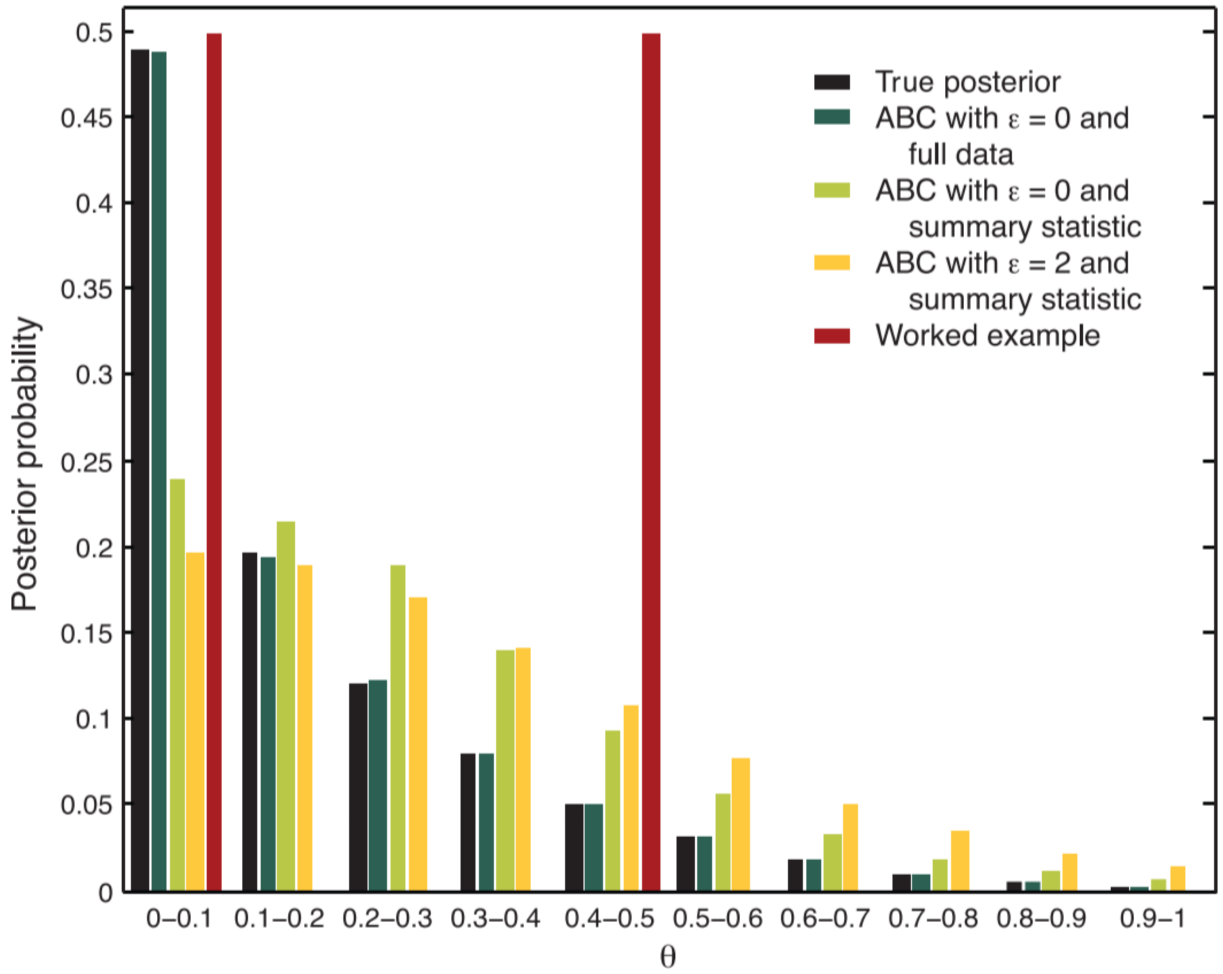
Example table of ABC rejection Algorithm

Table 1. Example of ABC rejection algorithm.

i	θ_i	Simulated Datasets (Step 2)	Summary Statistic $\omega_{S,i}$ (Step 3)	Distance ρ ($\omega_{S,i}$,ω_E) (Step 4)	Outcome (Step 4)
1	0.08	AABAAAABAABAAABAAAAA	8	2	accepted
2	0.68	AABBABABAAABBABABBAB	13	7	rejected
3	0.87	BBBABBABBBBABABBBBBA	9	3	rejected
4	0.43	AABAAAAABBABBBBBBBBA	6	0	accepted
5	0.53	ABBBBBAABBABBABAABBB	9	3	rejected

Posterior probabilities using large n

- Observe how the worked example doesn’t match the true posterior. Why do you think that is?
- Observe ε=0 and using full data that it matches closely to the true posterior.
- Why does the summary statistic with ε=0 have a closer posterior probability than when ε=2?



Model Comparison Question

- The ABC-framework can be used to compute the posterior probabilities of different candidate models.
- First a model is sampled from the prior distribution for the models
- Then given the model sampled, the model parameters are sampled from the prior distribution assigned to that model.
- Then do a simulation as before.
- The relative frequencies for the different models now approximate the posterior distribution for these models.

Model Comparison Question

- Once the posterior probabilities of models have been estimated, Bayesian model comparison can be performed.
- Compare the relative plausibility's of two models
  - We can computer their posterior ratio
  - Which is related to the Bays factor $B_{1,2}$:
    - $\frac{p(M_1|D)}{p(M_2|D)} = \frac{p(D|M_1)}{p(D|M_2)} * \frac{p(D|M_1)}{p(D|M_2)} = B_{1,2} * \frac{\mathrm{p}(M_1)}{p(M_2)}$
  - If the model priors are equal $\left(p(M1) = p(M2)\right)$ the Bayes factor equals the posterior ratio.
- Why do you think we should be cautious of this model comparison?

Model Comparison Question

- Once the posterior probabilities of models have been estimated, Bayesian model comparison can be performed.
- Compare the relative plausibilities of two models
  - We can computer their posterior ratio
  - Which is related to the Bays factor $B_{1,2}$:
    - $\frac{p(M_1|D)}{p(M_2|D)} = \frac{p(D|M_1)}{p(D|M_2)} * \frac{p(D|M_1)}{p(D|M_2)} = B_{1,2} * \frac{\mathrm{p}(M_1)}{p(M_2)}$
  - If the model priors are equal $\left(p(M1) = p(M2)\right)$ the Bayes factor equals the posterior ratio.
- Why do you think we should be cautious of this model comparison?
  - These measures can be highly sensitive to the choice of parameter prior distributions and summary statistics.

Potential Risks and remedies

Table 2. Potential risks and remedies in ABC-based statistical inference.

Error Source	Potential Issue	Solution	Subsection
Nonzero tolerance ε	The inexactness introduces a bias in the computed posterior distribution.	Theoretical/practical studies of the sensitivity of the posterior distribution to the tolerance. Noisy ABC.	Approximation of the posterior
Nonsufficient summary statistics	The information loss causes inflated credible intervals.	Automatic selection/semi-automatic identification of sufficient statistics. Model validation checks (e.g., Templeton 2009 [19]).	Choice and sufficiency of summary statistics
Small number of models/mis-specified models	The investigated models are not representative/lack predictive power.	Careful selection of models. Evaluation of the predictive power.	Small number of models
Priors and parameter ranges	Conclusions may be sensitive to the choice of priors. Model choice may be meaningless.	Check sensitivity of Bayes factors to the choice of priors. Some theoretical results regarding choice of priors are available. Use alternative methods for model validation.	Prior distribution and parameter ranges
Curse-of-dimensionality	Low parameter acceptance rates. Model errors cannot be distinguished from an insufficient exploration of the parameter space. Risk of overfitting.	Methods for model reduction if applicable. Methods to speed up the parameter exploration. Quality controls to detect overfitting.	Curse-of-dimensionality
Model ranking with summary statistics	The computation of Bayes factors on summary statistics may not be related to the Bayes factors on the original data, which may therefore render the results meaningless.	Only use summary statistics that fulfill the necessary and sufficient conditions to produce a consistent Bayesian model choice. Use alternative methods for model validation.	Bayes factor with ABC and summary statistics
Implementation	Low protection to common assumptions in the simulation and the inference process.	Sanity checks of results. Standardization of software.	Indispensable quality controls

Approximation of the Posterior

- As an attempt to correct some of the error due to a non-zero ε, the usage of local linear weight regression with ABC to reduce the variance of the posterior
- The method assigns weights to the parameters according to how well the simulated summaries adhere to the observed ones and performs linear regression between the summaries and the weighted parameters in the vincinity of the observed summaries.
- The regression coeffiecients are used to correct sampled parameters in the direction of observed summaries.

Approximation of the Posterior

- Statistical inference using ABC with a non-zero tolerance ε is not inherently flawed:
- The optimal ε can sometimes be not zero.
- The bias caused by a non-zero tolerance can be characterized and compensated by introducing a specific form of noise to the summary statistics.
- Asymptotic consistency for "noisy ABC" has been established and together with formulas for the asymptotic variance of the parameter estimates for a fixed tolerance.

Choosing Summary Statistics

- Poorly chosen summary statistics will often lead to inflated credible intervals due to the implied loss of information.
- One approach to capture most of the information present in data would be to use many statistics
  - Accuracy and stability of ABC appears to decrease rapidly with increasing numbers of summary statistics.
- Another approach is to focus on relevant statistics only
  - Relevancy depending on the whole inference problem, on the model used and on the data at hand.

Choosing Summary Statistics

- There's an algorithm that iteratively assess whether an additional statistic introduces a meaningful modification of the posterior.
  - The challenge is that a large ABC approximation error may heavily influence the conclusions about the usefulness of a statistic at any stage of the procedure.
- Another method uses 2 main steps
  - First, a reference approximation of the posterior is constructed by minimizing the entropy.
  - Secondly, sets of candidate summaries are then evaluated by comparing the ABC approximated posteriors with the reference posterior.

Choosing Summary Statistics

- Recently a method for
- Another method uses 2 main steps
  - First, a reference approximation of the posterior is constructed by minimizing the entropy.
  - Secondly, sets of candidate summaries are then evaluated by comparing the ABC approximated posteriors with the reference posterior.
- A subset of statistics is selected from a large set of candidate statistics
- Methods for identifying summary statistics that could also simultaneously assess the influence on the approximation of the posterior would be valuable.

Choosing Summary Statistics

- Why would these methods be valuable?

Choosing Summary Statistics

- Why would these methods be valuable?
  - The paper suggests that the choice of summary statistics and the choice of tolerance constitute two sources of error in the resulting posterior distribution.
  - These errors may lead to incorrect model predictions.

Bayes Factor with ABC and Summary Statistics

- A combination of Insufficient summary statistics and ABC for model selection can be problematic.
- Let S(D) be denoted by$B^S_{1,2}$then the relation between $B_{1,2}$ and $B^S_{1,2}$ takes the form:

$$B_{1,2} = \frac{p(D|M_1)}{p(D|M_2)} = \frac{p(D|S(D),M_1)}{p(D|S(D),M_2)}\frac{p(S(D)|M_1)}{p(S(D)|M_2)} = \frac{p(D|S(D),M_1)}{p(D|S(D),M_2)} B^s_{1,2}$$

Bayes Factor with ABC and Summary Statistics

- Thus a summary statistic S(D) is sufficient for comparing two models M_1 and M_2 if and only if:

$$p(D|S(D), M_1) = p(D|S(D), M_2),$$

- Which results in that $B_{1,2} = B^S_{1,2}$
- From the equation above if the condition is not satisfied there might be a huge difference between $B_{1,2}$ and $B^S_{1,2}$.
- M_1 and M_2 alone or for both models does not guarantee sufficiency for ranking the models.

Bayes Factor with ABC and Summary Statistics

- Any Sufficient summary statistic for a model M in which both M_1 and M_2 are nested is valid for ranking the nested models.
- The computation of Bayes factors on S(D) may be misleading for model selection purposes, unless the ratio between the Bayes factors on D and S(D) would be available or approximated reasonably well.
- This issue is only relevant for model selection when the dimension of the data has been reduced

Indispensable Quality Controls

- A number of heuristic approaches to the quality control of ABC
- Such as the quantification of the fraction of parameter variance
  - A common class of methods aims at assessing whether or not the inference yields valid results, regardless of the actually observed data.
- Another class of methods assesses whether the inference was successful in light of the given observed data.
  - For example by comparing the posterior predictive distribution of summary statistics to the summary statistics observed.

Indispensable Quality Controls

- Cross-validation techniques and predictive checks represent promising future strategies to evaluate the stability and out-of-sample predictive validity of ABC inferences.
- Out-of-sample predictive checks can reveal potential systematic biases within a model and provide clues on to how to improve its structure or parametrization.

Indispensable Quality Controls

- Another quality control based method for model selection employs ABC to approximate the effective number of model parameters and the deviance of the posterior predictive distributions of summaries and parameters.
- The Deviance information criterion is then used as measure of model fit.
- Models preferred based on this criterion can conflict with those supported by Bayes factors.
- For this reason it is useful to combine different methods for model selection to obtain correct conclusions.

General Risks in Statistical Inference

- There are risks that are not specific to ABC that the paper discusses.
- These risks include:
  - Prior distribution and parameter ranges
  - Small number of models
  - Large Datasets
  - Curse of dimensionality

Prior Distribution and parameter ranges

- The specification of the range and the prior distribution of parameters strongly benefits from previous knowledge.
- It is typically necessary to constrain the parameter ranges.
- Educated guesses are used at times for the parameter ranges.
- Bayes factors highly sensitive to the prior distribution of parameters
- Conclusions on model choice based on Bayes factor can be misleading unless the sensitivity of conclusions to the choice of priors is carefully considered.

Small number of models

- Model-based methods have been criticized for not exhaustively covering the hypothesis space.
- There’s a high computational cost to evaluate a single model in some instances.
- There isn’t strategy that works for all model development.

Large datasets

- Large datasets may constitute a computational bottleneck for model-based methods.
- It was pointed out that in some ABC-based analyses, part of the data needed to be omitted.

Large datasets

- A few proposed ideas to reduce this problem:
  - Using Metropolis-Hastings algorithm with ABC
    - Pro: Increased acceptance rate than plain ABC.
    - Con: Inherits the general burdens of Markov Chain Monte Carlo (MCMC) such as difficulty to assess convergence correlation among the samples from the posterior and relatively poor parallelizability
  - Sequential Monte Carlo(SMC) and population Monte Carlo
    - Takes an iterative approach to the posterior from the prior through a sequence of target distributions.
    - The advantage is the samples from the resulting posterior are independent.
    - Tolerance levels are not specified prior to analysis and must be adjusted adaptively.
    - It is straightforward to parallelize a number of steps using rejection sampling and SMC

Curse of dimensionality

- High dimensional datasets and high-dimensional parameter spaces can require an extremely large number of parameter points to get a reasonable level of accuracy for the posterior inferences.
- Computational cost is severely increased
- Computational analysis can become intractable.
- The probability of accepting the simulated values under a given tolerance with ABC rejection algorithm decreases EXPONENTIALLY with increasing dimensionality.

Software for Application of ABC

Table 3. Software incorporating ABC.

Software	Keywords and Features	Reference
DIY-ABC	Software for fit of genetic data to complex situations. Comparison of competing models. Parameter estimation. Computation of bias and precision measures for a given model and known parameters values.	[53]
ABC R package	Several ABC algorithms for performing parameter estimation and model selection. Nonlinear heteroscedastic regression methods for ABC. Cross-validation tool.	[54]
ABC-SysBio	Python package. Parameter inference and model selection for dynamical systems. Combines ABC rejection sampler, ABC SMC for parameter inference, and ABC SMC for model selection. Compatible with models written in Systems Biology Markup Language (SBML). Deterministic and stochastic models.	[55]
ABCtoolbox	Open source programs for various ABC algorithms including rejection sampling, MCMC without likelihood, a particle-based sampler, and ABC-GLM. Compatibility with most simulation and summary statistics computation programs.	[56]
msBayes	Open source software package consisting of several C and R programs that are run with a Perl "front-end." Hierarchical coalescent models. Population genetic data from multiple co-distributed species.	[57]
PopABC	Software package for inference of the pattern of demographic divergence. Coalescent simulation. Bayesian model choice.	[58]
ONeSAMP	Web-based program to estimate the effective population size from a sample of microsatellite genotypes. Estimates of effective population size, together with 95% credible limits.	[59]
ABC4F	Software for estimation of F-statistics for dominant data.	[60]
2BAD	Two-event Bayesian ADmixture. Software allowing up to two independent admixture events with up to three parental populations. Estimation of several parameters (admixture, effective sizes, etc.). Comparison of pairs of admixture models.	[61]

doi:10.1371/journal.pcbi.1002803.t003

Summary and Discussion

- For numerous applications calculating the likelihood and the prior is computationally expensive.
- ABC is a useful tool to approximate the posterior distribution.
- ABC toolkit is best suited for inference about parameters or predictive inferences about observables in the presence of a 1 or few candiate models
- Large sets of models or high dimensional target parameter spaces are major issues with ABC.