

Approximate Bayesian computation (ABC)

Reporter

Anirban Sarker

Lab of Molecular Systematics and Ecology

- **Population genetics** is a subfield of genetics that deals with genetic differences within and between populations.



- ABC is a group of methods, that is suitable to solve the problems, arise in population genetics along with other fields.




- **ABC** constitutes a class of computational methods rooted in Bayesian statistics.
- **Bayesian statistics** is a theory in the field of statistics, in which the evidence about the true state of the world is expressed in terms of *degrees of belief*, as known as Bayesian probabilities.
- Instead of frequency or propensity of some phenomenon

History

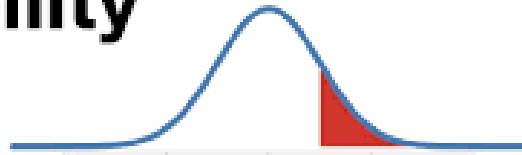
- The term *Bayesian* derives from the name of Thomas Bayes.
- He was a great mathematician, philosopher and theologian of 18 th century.
- The term ABC was established by Mark Beaumont *et al.*
- Bayesian phylogenetic methods were introduced in the 1990s.



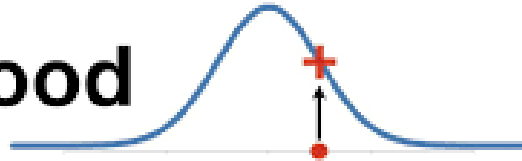
- 
- The estimation of demographic parameters from genetic data often requires the computation of likelihoods.
 - But the situation changed recently with the advent of ABC algorithms allowing one to obtain parameter posterior distributions based on simulations not requiring likelihood computations.

Probability

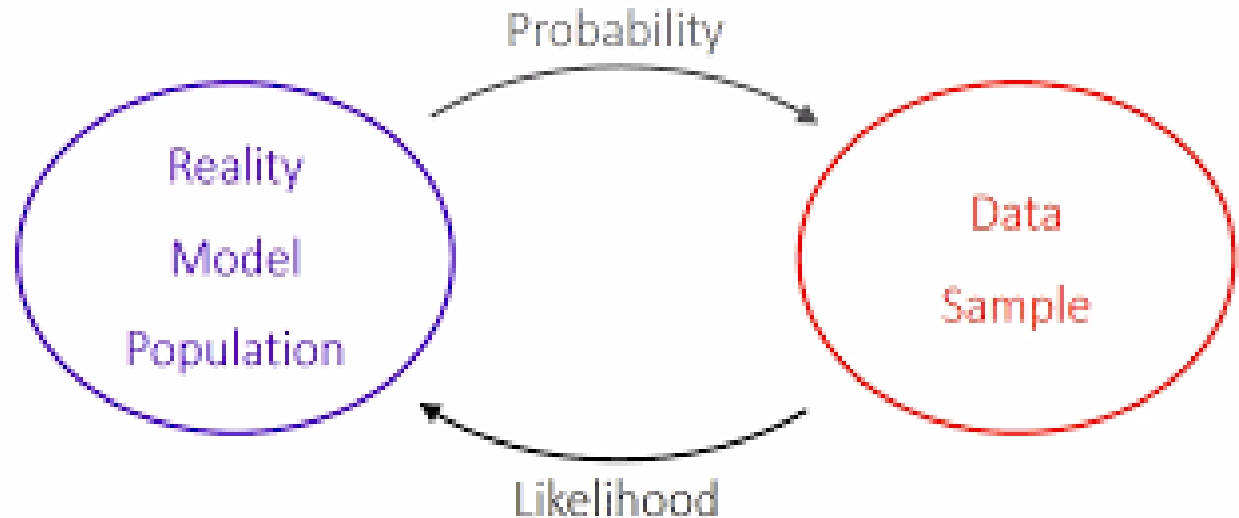
Vs



Likelihood



- *Probability* describes the plausibility of observed data assumed to be described by a statistical model, without reference to any observed data.
- whereas *likelihood* describes the plausibility of a parameter given the observed data.

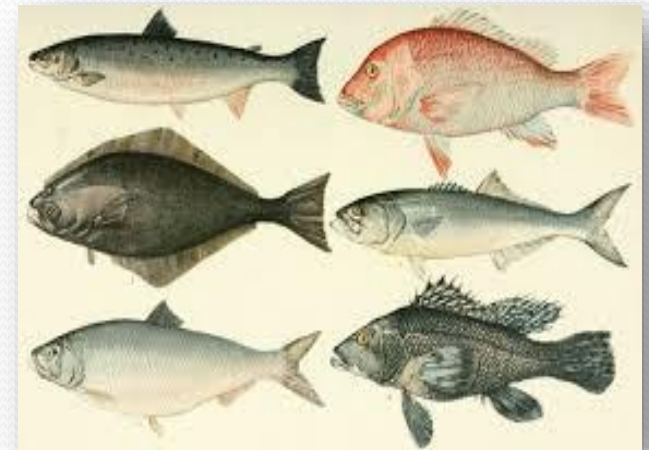


Applications

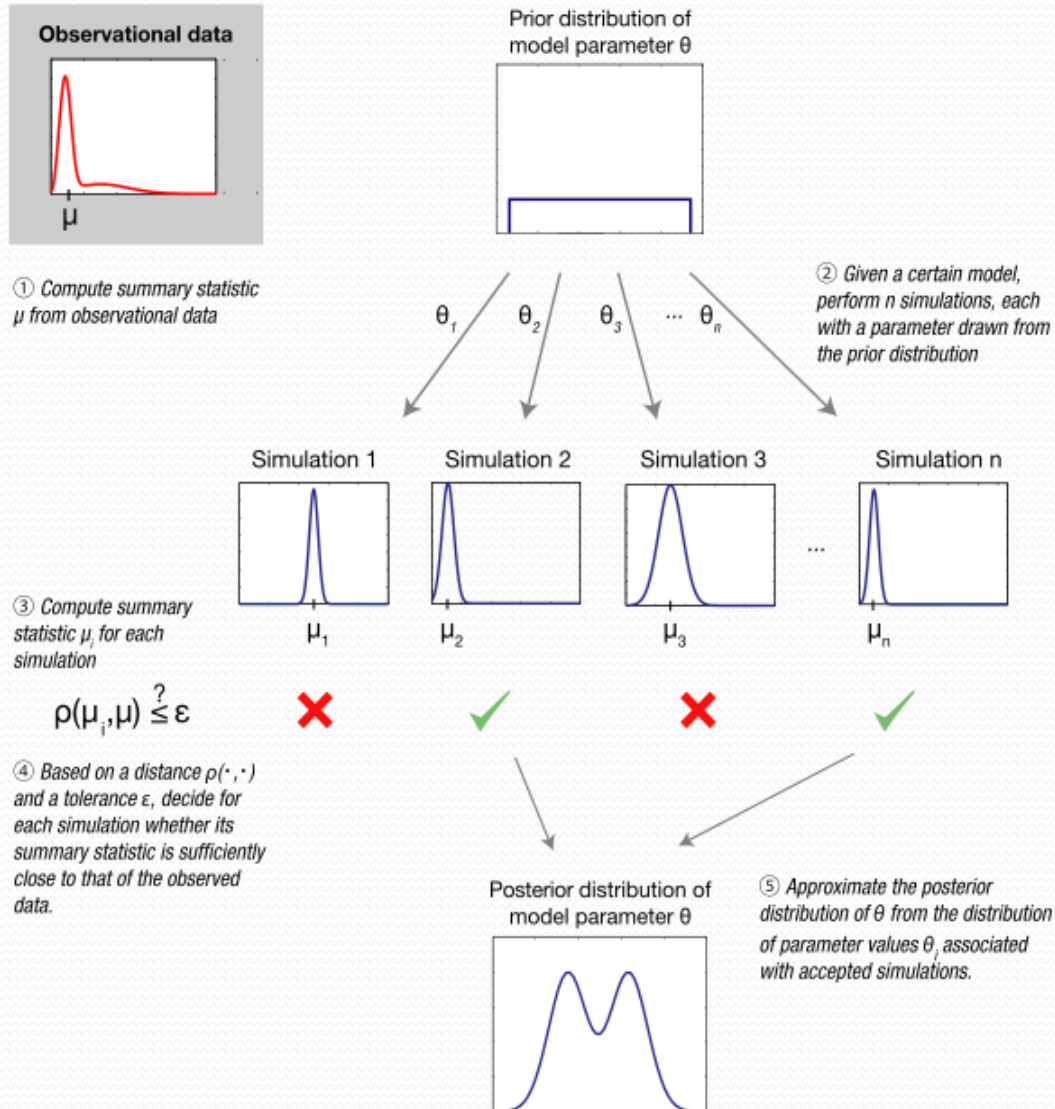
- Analysis of ecology or systems biology
- Phylogeographic analysis of virus spread in humans
- Inference of phylogeographic history and migration between species
- Analysis of species diversification rates
- Inference of phylogenetic relationships among species or populations.
- Analysis of complex problems arising in biological sciences, e.g. in population genetics and systems biology.

What type of data we can use

- DNA and amino acid sequence alignments
- Morphological characters




Conceptual overview

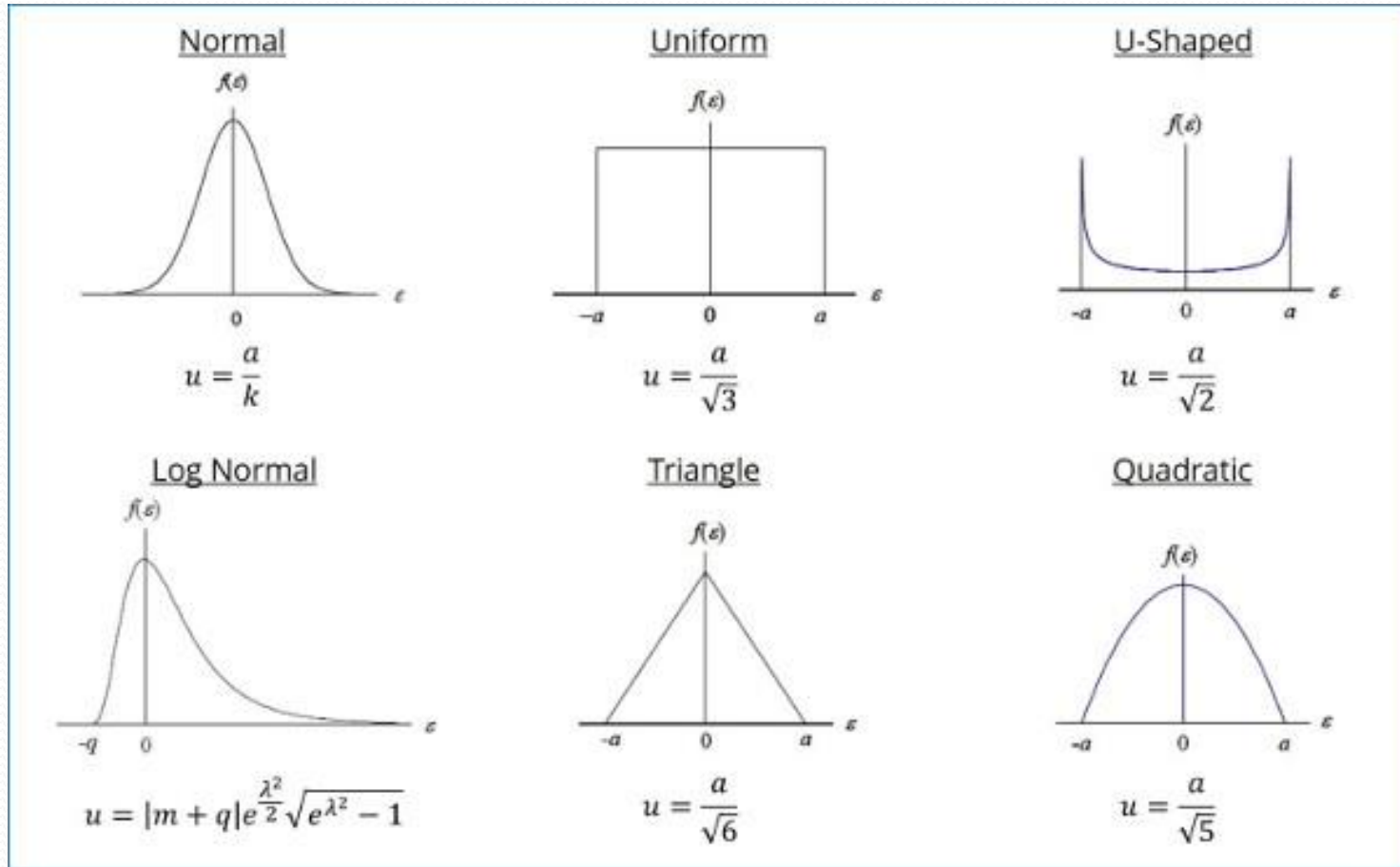


Summary statistics

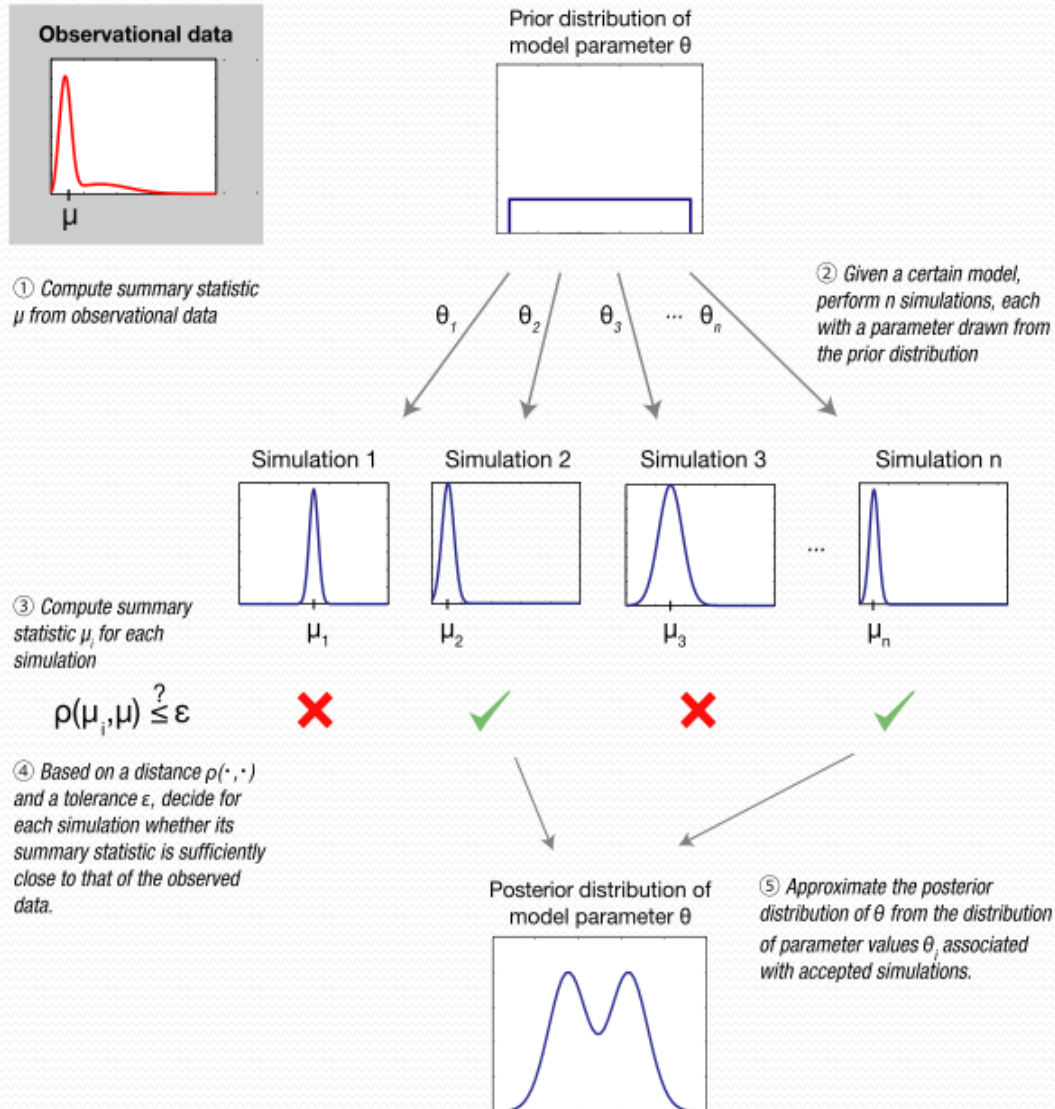
- In order to evaluate the distance between the simulated data and the observed data, the data needs to be summarized in a quantitative form.
- The choice of summary statistics, however, is a tough one.
- Too few summary statistics may not capture the characteristics of the model, which leads to overall low power for parameter estimation.
- Too many summary statistics, on the other hand, may introduce random noise.

- 
- Observed data: AAB^{red}AAA^{blue}AB^{green}BB^{blue}BBB^{red}BBB^{red}BB^{blue}A
 - Summary statistic value: 6

Prior probability distribution



Conceptual overview



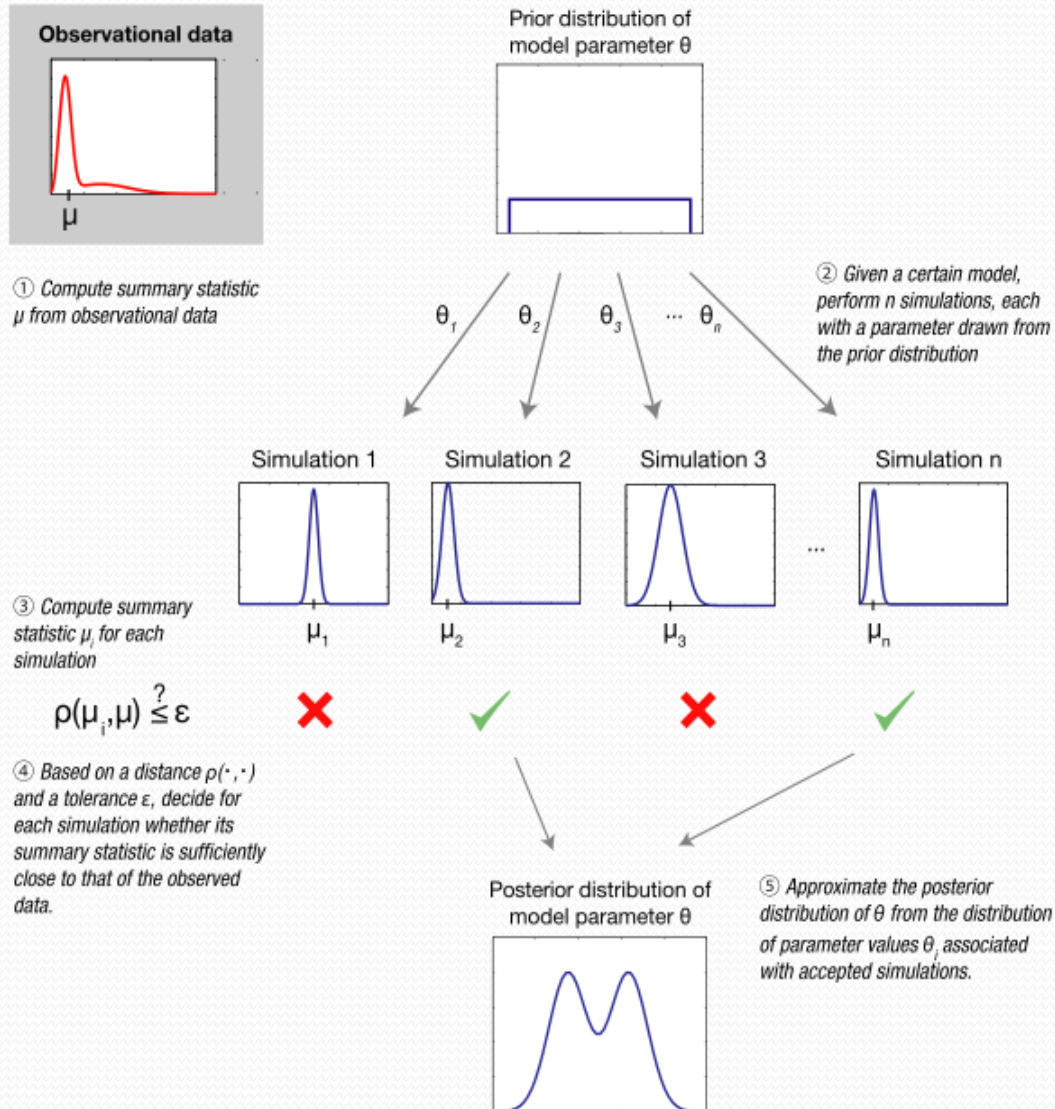
Example of ABC rejection algorithm

i		Simulated datasets (step 2)	Summary statistic (step 3)	Distance (step 4)	Outcome (step 4)
1	0.08	AABBAABABAB AABBAABABAB	8	2	accepted
2	0.68	AABBABABAAAB BABABBAB			
3	0.87	BBBABBABBBA BABBBBBA			
4	0.43	AABAAAAABBA BBBBBBBBA			
5	0.53	ABBBBBAABBAB BABAABBB			

Example of ABC rejection algorithm

i		Simulated datasets (step 2)	Summary statistic (step 3)	Distance (step 4)	Outcome (step 4)
1	0.08	AABAAAABAAB AAABAAAAA	8	2	accepted
2	0.68	AABBABABAAAB BABABBAB	13	7	rejected
3	0.87	BBBABBABBBA BABBBBBA	9	3	rejected
4	0.43	AABAAAAABBA BBBBBBBBA	6	0	accepted
5	0.53	ABBBBBAABBAB BABAABBB	9	3	rejected

Conceptual overview

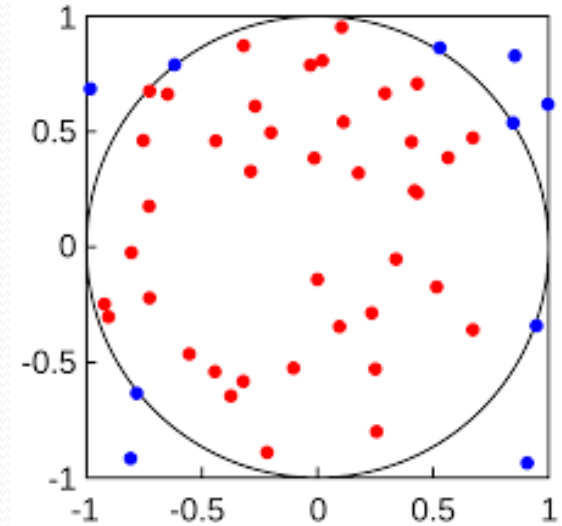


Choice of threshold level

- At the time of choice, we should make good balance between bias and Monte Carlo error.
- When value is 0, this is identical to algorithm.
- When this is 3, is better than previous one upto 100,000 simulations.
- But at the time of 5, bias dominates.

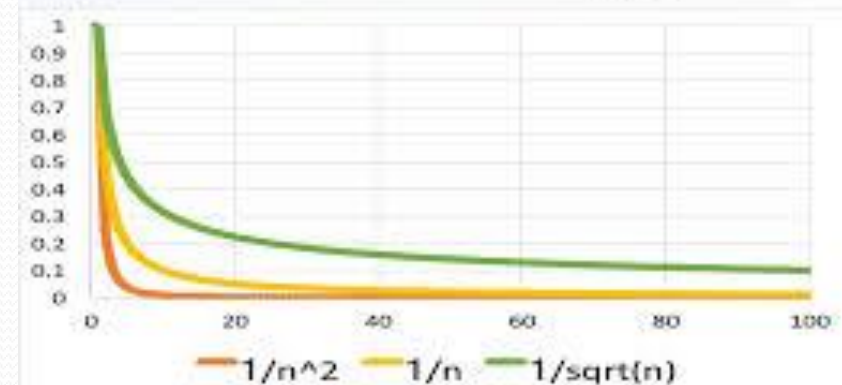
Monte Carlo integration & error

- In mathematics, **Monte Carlo integration** is a technique for numerical integration using random numbers.
- Points inside the circle is 40
- Points inside the square is 50
- Ratio: $40/50=0.8$
- Circle area: $4 \times 0.8= 3.2$



MONTE CARLO ERROR

$$Error \propto \frac{1}{\sqrt{n}}$$



Why we use ABC

- The availability of user-friendly software for running sophisticated models of evolution.
- The development of powerful models of data analysis.

Disadvantages

- Bayesian phylogenetic models are complex.
- Analyses are often conducted using default priors, which may not be appropriate and may lead to biased or incorrect results.
- The priors and model assumptions made in those programs are not always clear to the user.

Challenges

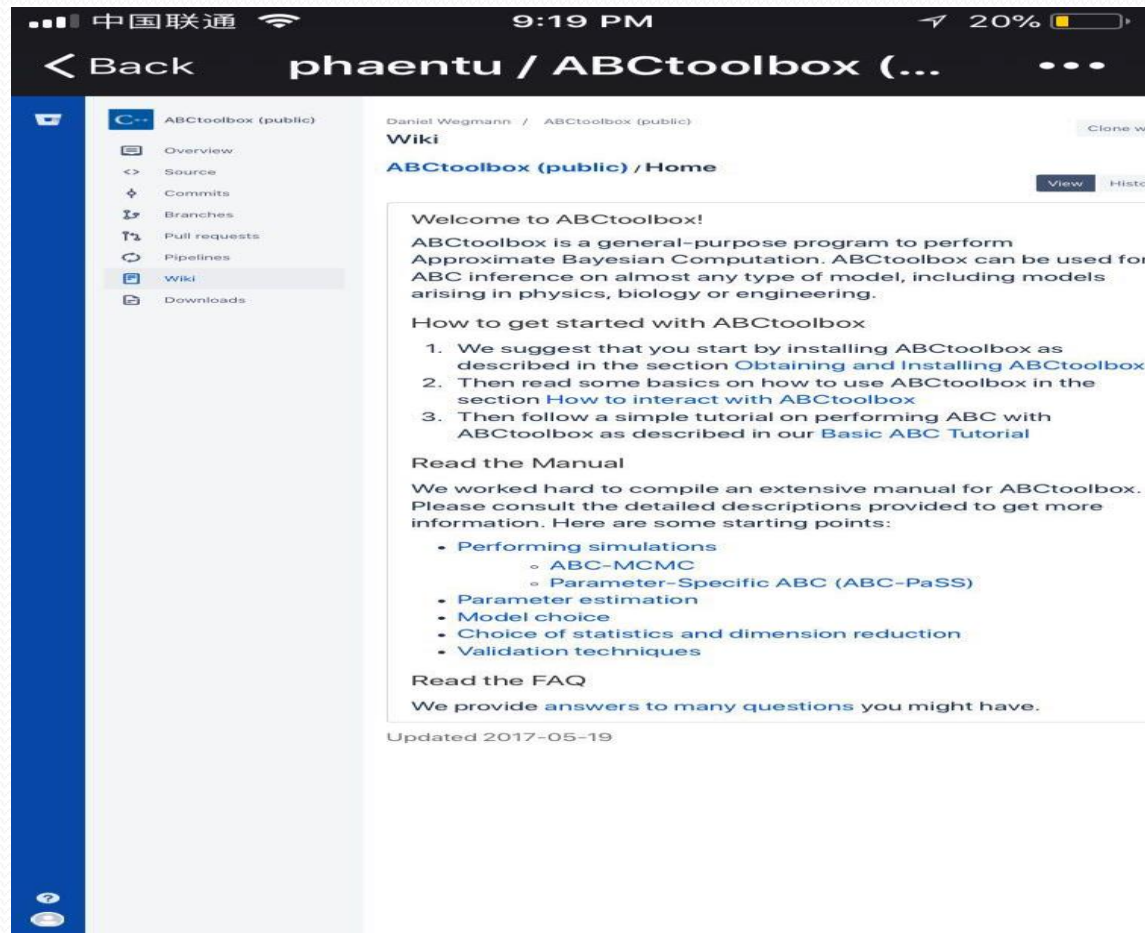
- ❖ First, data simulation, which is in the core of any ABC analysis, is specific to the model. Thus, many existing ABC software are specific to a particular class of models.
- ❖ Second, an ABC analysis often follows a trial–error approach, where users experiment with different models, ABC algorithms or summary statistics.
- ❖ Third, ABC is subject to intensive research, and many new algorithms have been published in the past few years .

ABC-softwares

- DIY-ABC
- abc R package
- Easy ABC R package
- ABC-SysBio
- ABCtoolbox
- msBayes
- PopABC
- ONeSAMP
- ABC4F
- 2BAD
- ELFI

Web-address of ABCtoolbox

<https://bitbucket.org/phaentu/abctoolbox-public/wiki/Home>



The screenshot shows a mobile browser interface displaying the Bitbucket wiki for the ABCtoolbox project. The page title is "ABCtoolbox (public) / Home". The main content area contains a welcome message and a list of starting points for users. The left sidebar shows navigation options like Overview, Source, Commits, Branches, Pull requests, Pipelines, Wiki, and Downloads. The top status bar shows the time as 9:19 PM and 20% battery.

China 中国联通 9:19 PM 20%

< Back phaentu / ABCtoolbox (...)

ABCtoolbox (public)

Overview
Source
Commits
Branches
Pull requests
Pipelines
Wiki
Downloads

Daniel Wegmann / ABCtoolbox (public)

Wiki

ABCtoolbox (public) / Home

Welcome to ABCtoolbox!

ABCtoolbox is a general-purpose program to perform Approximate Bayesian Computation. ABCtoolbox can be used for ABC inference on almost any type of model, including models arising in physics, biology or engineering.

How to get started with ABCtoolbox

1. We suggest that you start by installing ABCtoolbox as described in the section [Obtaining and Installing ABCtoolbox](#)
2. Then read some basics on how to use ABCtoolbox in the section [How to interact with ABCtoolbox](#)
3. Then follow a simple tutorial on performing ABC with ABCtoolbox as described in our [Basic ABC Tutorial](#)

Read the Manual

We worked hard to compile an extensive manual for ABCtoolbox. Please consult the detailed descriptions provided to get more information. Here are some starting points:

- [Performing simulations](#)
 - [ABC-MCMC](#)
 - [Parameter-Specific ABC \(ABC-PaSS\)](#)
- [Parameter estimation](#)
- [Model choice](#)
- [Choice of statistics and dimension reduction](#)
- [Validation techniques](#)

Read the FAQ

We provide [answers to many questions you might have](#).

Updated 2017-05-19

Tenualosa BD 2019

Project name :

Data file :

Directory :

Data file info :

Thanks to all

