Approximate Bayesian Computation

Outline

1. What is Approximate Bayesian Computation?
2. When do we use ABC instead of other methods would we use it?
3. How do we use it?
   a) Choices in the ABC-rejection algorithm
   b) Short introduction to more advanced ABC

1. What is Approximate Bayesian Computation?

Bayesian inference is based on the idea of updating belief with new evidence

- **Belief**: Prior distribution. Parameters are random variables instead of fixed quantities (they have there own distribution)
- **Evidence**: Likelihood function tells you the probability of the data given the parameters

Bayesian inference

θ : Mathematical model parameter, D : data

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

Bayesian inference

θ : Mathematical model parameter, D : Data

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

Bayesian inference

θ : Mathematical model parameter, D : Data

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

Probability of data given θ (likelihood)

EVIDENCE

Bayesian inference

θ : Mathematical model parameter, D : Data

Prior probability

BELIEF

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

Probability of data given θ (likelihood)

EVIDENCE

Bayesian inference

θ : Mathematical model parameter, D : Data

Prior probability
BELIEF

Posterior probability $\longrightarrow$ $P(\theta|D) \propto P(D|\theta)P(\theta)$

Probability of data given θ (likelihood)
EVIDENCE

Bayesian inference

θ : Mathematical model parameter, D : Data

Prior probability

BELIEF

Posterior probability

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

Probability of data given θ **(likelihood)**

EVIDENCE

What if we can't use a likelihood function?

ABC rejection algorithm

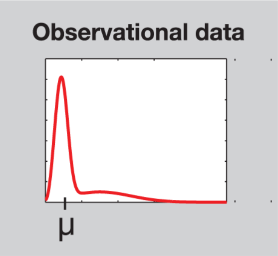

Figure from https://doi.org/10.1371/journal.pcbi.1002803

ABC rejection algorithm

1. Sample θ^* from the prior distribution $P(\theta)$

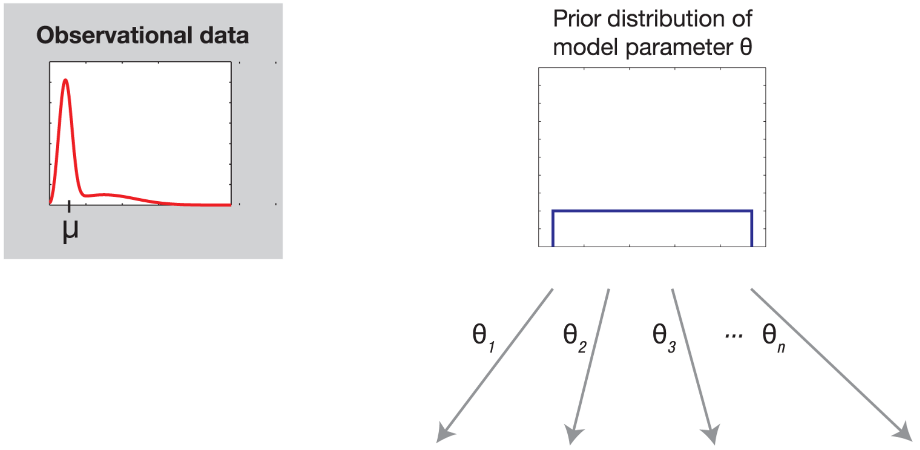

Figure from https://doi.org/10.1371/journal.pcbi.1002803

ABC rejection algorithm

1. Sample θ^* from the prior distribution $P(\theta)$
2. Simulate a dataset D^* from your model using θ^*

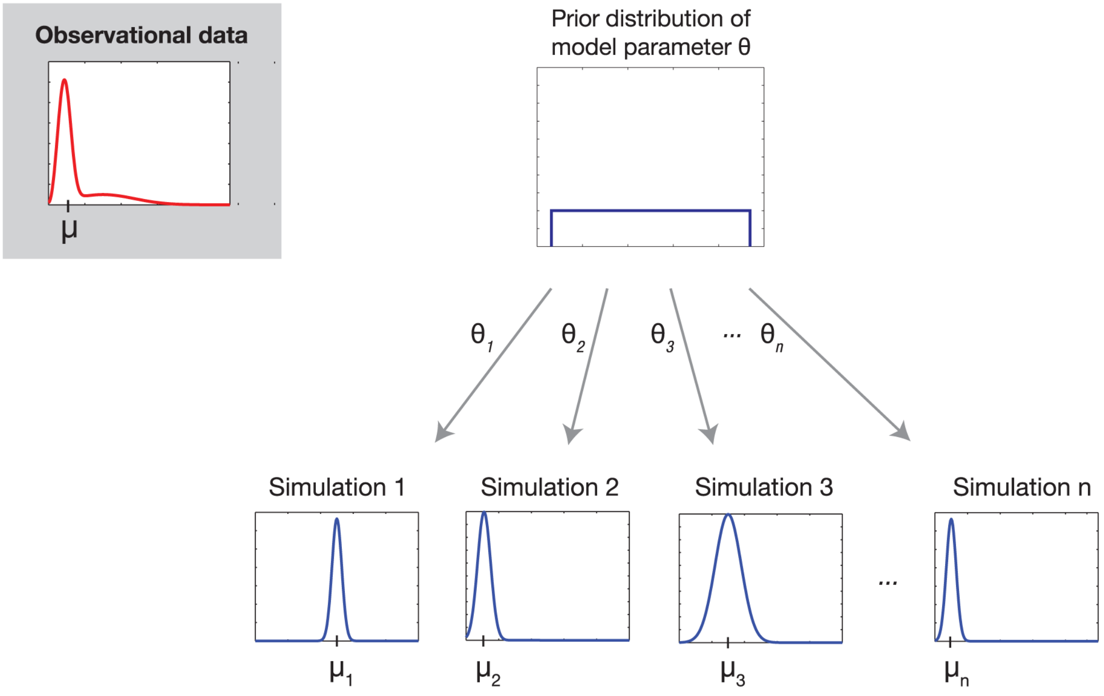

Figure from https://doi.org/10.1371/journal.pcbi.1002803

ABC rejection algorithm

1. Sample θ^* from the prior distribution $P(\theta)$
2. Simulate a dataset D^* from your model using θ^*
3. If $d(D, D^*) \leq \epsilon$ accept θ^*, otherwise reject

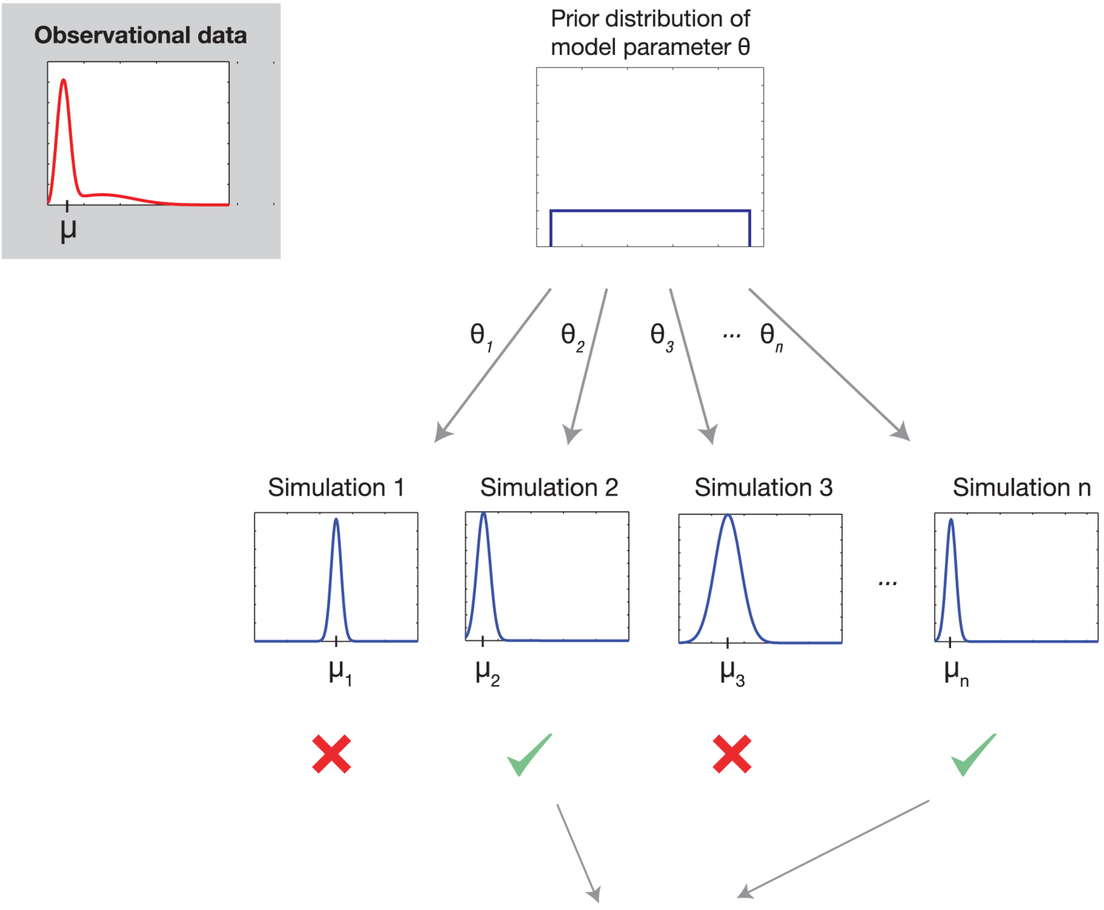

Figure from https://doi.org/10.1371/journal.pcbi.1002803

ABC rejection algorithm

1. Sample θ^* from the prior distribution $P(\theta)$
2. Simulate a dataset D^* from your model using θ^*
3. If $d(D, D^*) \leq \epsilon$ accept θ^*, otherwise reject
4. Repeat until you have accepted N accepted samples

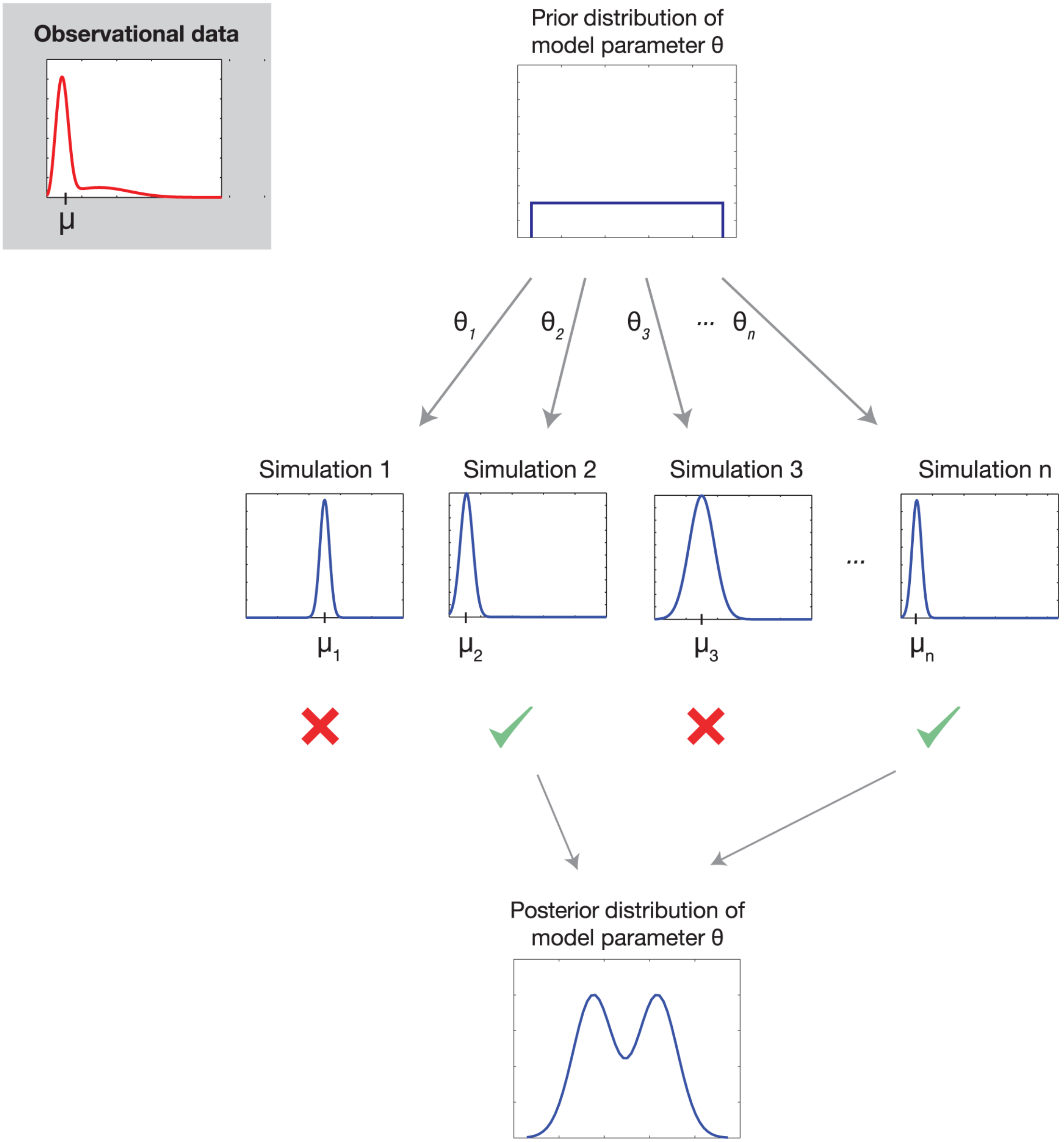

Figure from https://doi.org/10.1371/journal.pcbi.1002803

ABC rejection algorithm

1. Sample θ^* from $P(\theta)$
2. Simulate a dataset D^* from your model using θ^*
3. **Calculate the summary statistic for the observed data $\boldsymbol{\mu} = \boldsymbol{S}(\boldsymbol{D})$ and simulated data $\boldsymbol{\mu} = \boldsymbol{S}(\boldsymbol{D}^*)$**
4. **If $\boldsymbol{d}\big(\boldsymbol{S}(\boldsymbol{D}), \boldsymbol{S}(\boldsymbol{D}^*)\big) \leq \boldsymbol{\epsilon}$ accept $\boldsymbol{\theta}^*$, otherwise reject**
5. Repeat until you have accepted N accepted samples

ABC rejection algorithm

1. Sample θ^* from $P(\theta)$
2. Simulate a dataset D^* from your model using θ^*
3. **Calculate the summary statistic for the observed data $\boldsymbol{\mu} = \boldsymbol{S}(\boldsymbol{D})$ and simulated data $\boldsymbol{\mu} = \boldsymbol{S}(\boldsymbol{D}^*)$**
4. **If $\boldsymbol{d}\left(\boldsymbol{S}(\boldsymbol{D}), \boldsymbol{S}(\boldsymbol{D}^*)\right) \leq \boldsymbol{\epsilon}$ accept $\boldsymbol{\theta}^*$, otherwise reject**
5. Repeat until you have accepted N accepted samples

Summary statistic for model trajectory

Distance measure between summary statistic and data

1. What is Approximate Bayesian Computation?

A method to approximate the posterior distribution $P(\theta|D)$ without a likelihood function

$$\boxed{P(\theta|D) \approx P(\theta|d\big(S(D), S(D^*)\big) \leq \epsilon)}$$

2. When do we use ABC instead of other methods?

2. When do we use ABC instead of other methods?

- Data quality is poor, which means we have to aggregate it
- The likelihood function might be costly to evaluate (it takes a long time)
  - Large data sets
  - Complicated likelihood function

2. When do we use ABC instead of other methods?

- Data quality is poor, which means we have to aggregate it
- The likelihood function might be costly to evaluate (it takes a long time)
  - Large data sets
  - Complicated likelihood function

- Intuitive method of model fitting
  - Parameter -> model trajectory -> accept or reject

3. How do we use ABC?

a. Choices in the ABC- rejection algorithm

Choice of summary statistic(s) $\boldsymbol{S}(\boldsymbol{D})$

- This is how we choose whether to accept or reject parameter values
- Sufficient summary statistic will give the same result as the likelihood
- "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter"

Fisher, R.A. (1922). "On the mathematical foundations of theoretical statistics". *Philosophical Transactions of the Royal Society A*. **222**: 309–368. doi:10.1098/rsta.1922.0009. JFM 48.1280.02. JSTOR 91208.

Choice of summary statistic(s) $\boldsymbol{S(D)}$

- This is how we choose whether to accept or reject parameter values
- Sufficient summary statistic will give the same result as the likelihood
- "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter"

- If we haven't written down a likelihood then we can't know if our summary statistics are sufficient…

Fisher, R.A. (1922). "On the mathematical foundations of theoretical statistics". *Philosophical Transactions of the Royal Society A.* **222**: 309–368. doi:10.1098/rsta.1922.0009. JFM 48.1280.02. JSTOR 91208.

Choice of summary statistic(s) $\boldsymbol{S(D)}$

- This is how we choose whether to accept or reject parameter values
- Sufficient summary statistic will give the same result as the likelihood
- "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter"

- In practice
  - Look at published model fitting studies using ABC methods for ideas for sufficient statistics
  - **Check with simulated data**!

Fisher, R.A. (1922). "On the mathematical foundations of theoretical statistics". *Philosophical Transactions of the Royal Society A*. **222**: 309–368. doi:10.1098/rsta.1922.0009. JFM 48.1280.02. JSTOR 91208.

Number of particles (N)

- The more the better, but computation time must be taken into account

Tolerance value $\boldsymbol{\epsilon}$

- Determines whether you accept or reject parameter(s) based on how closely the model prediction matches you data
  - Too small and the algorithm will take a long time to run
  - Too big and the final distribution of particles will be too wide

Tolerance value ϵ

- Determines whether you accept or reject parameter(s) based on how closely the model prediction matches you data
  - Too small and the algorithm will take a long time to run
  - Too big and the final distribution of particles will be too wide

- The magnitude of the tolerance value ϵ will depend on your distance measure

For example, if the summary of the data $S(D)$ is the cumulative number of cases, we could have:

- $S(D) = 100\,000$ (from the data)
- $S(D^*) = 99\,900$ (model prediction)
- If the distance measure $d()$ is the sum of squared difference the,

$$d\big(S(D), S(D^*)\big) = (100\,000 - 99\,00)^2 = (100)^2 = 10\,000$$

The prediction was 100 people short of the data, distance measure is $10\,000$. Hence here a reasonable choice of tolerance might be $\boldsymbol{\epsilon} = 10\,000$.

3. How do we use ABC?

b. Short introduction to more advanced ABC

Improvements to ABC rejection algorithm: ABC-Sequential Monte Carlo (ABC-SMC)

- Instead of one tolerance ϵ, there is a vector of tolerances $\epsilon_1, \dots, \epsilon_T$

1. We perform ABC rejection with a very large tolerance ϵ_1 and store our N accepted parameter values as population 1.

Improvements to ABC rejection algorithm: ABC-Sequential Monte Carlo (ABC-SMC)

- Instead of one tolerance ϵ, there is a vector of tolerances $\epsilon_1, \ldots, \epsilon_T$

1. We perform ABC rejection with a very large tolerance ϵ_1 and store our N accepted parameter values as population 1.
2. Then we propose parameters by re-sampling parameters from population 1 and perturb the parameters by a small amount. Accept/reject according ϵ_2.
3. Add **weight** to each parameter value according to the prior distribution, how likely you were to obtain that value from perturbation and the previous weights.

Improvements to ABC rejection algorithm: ABC-Sequential Monte Carlo (ABC-SMC)

- Instead of one tolerance ϵ, there is a vector of tolerances $\epsilon_1, \ldots, \epsilon_T$

1. We perform ABC rejection with a very large tolerance ϵ_1 and store our N accepted parameter values as population 1.
2. Then we propose parameters by re-sampling parameters from population 1 and perturb the parameters by a small amount. Accept/reject according ϵ_2.
3. Add **weight** to each parameter value according to the prior distribution, how likely you were to obtain that value from perturbation and the previous weights.
   - Repeat steps 2-3 T times, sampling from the previous population. Each time decrease the tolerance value.

Practical

In summary: ABC

- Can be used when data quality is poor, likelihood is complex or unknown and is an intuitive model fitting technique
- ***But*** you have to specify a suitable summary statistic(s)
- ABC can be slow, there are many extensions: ABC-SMC, ABC-PMC etc.

Reading

General introductions

- McKinley, Trevelyan J.; Vernon, Ian; Andrianakis, Ioannis; McCreesh, Nicky; Oakley, Jeremy E.; Nsubuga, Rebecca N.; Goldstein, Michael; White, Richard G. Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models. Statist. Sci. 33 (2018), no. 1, 4--18. doi:10.1214/17-STS618. https://projecteuclid.org/euclid.ss/1517562021
- Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, et al. (2013) Approximate Bayesian Computation. PLOS Computational Biology 9(1): e1002803.https://doi.org/10.1371/journal.pcbi.1002803
- Hartig, F. , Calabrese, J. M., Reineking, B. , Wiegand, T. and Huth, A. (2011), Statistical inference for stochastic simulation models – theory and application. Ecology Letters, 14: 816-827. doi:10.1111/j.1461-0248.2011.01640.x
- **Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6 187-202; DOI: 10.1098/rsif.2008.0172.**

Reading

Examples of ABC

- Conlan, A.J., McKinley, T.J., Karolemeas, K., Pollock, E.B., Goodchild, A.V., Mitchell, A.P., Birch, C.P., Clifton-Hadley, R.S. and Wood, J.L., (2012). Estimating the hidden burden of bovine tuberculosis in Great Britain. *PLoS Computational Biology*, *8*(10), p.e1002730.
- McKinley, T., Cook, A. R. and Deardon, R. (2009). Inference in epidemic models without likelihoods. *Int. J. Biostat.* **5**.
- Beaumont MA, Zhang W, and Balding DJ. (2002) Approximate Bayesian Computation in Population Genetics. GENETICS. 162 (4) 2025-2035.