

Introduction to Approximate Bayesian Computation with Sequential Monte Carlo Methods

Kerry Liu

MT5731 Advanced Project

November 21, 2023

- Introduction to Approximate Bayesian Computation (ABC)
- ABC - Sequential Monte Carlo (SMC)
- Application of ABC - SMC

Bayes' Theorem

In Bayesian statistics, given parameters θ and sets of observed data $\mathbf{y}_1, \dots, \mathbf{y}_n$, we define the following:

- 1 **Prior Distribution:** $p(\theta)$ represents the *prior distribution*
- 2 **Posterior Distribution:** $\pi(\theta|y)$, the *posterior distribution*
- 3 **Likelihood** $f(y|\theta)$, which represents the probability of observing the data given the parameters θ .

Bayes' Theorem updates these beliefs and is expressed mathematically as:

$$\pi(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)}, \quad (1)$$

where $f(y)$, the *marginal likelihood*, acts as a normalizing constant. Since $f(y)$ is independent of θ , Equation (1) simplifies to:

$$\pi(\theta|y) \propto f(y|\theta)p(\theta). \quad (2)$$

What is Approximate Bayesian Computation (ABC)?

ABC methods provide a framework for analyzing datasets where the likelihood function $f(y|\theta)$ is intractable or unknown. This can be common in cases like:

- 1 The data generation process is a complex stochastic process or state-space model where $f(y|\theta)$ unknown analytically [4].
- 2 The dataset is large or high-dimensional, complicating direct likelihood computation.

What is Approximate Bayesian Computation (ABC)?

It approximates posterior through a two-step process:

- 1 Allow for the direct simulation of observations y^* from the prior predictive distribution $y^* \sim p(y|\theta^{(i)})$.
- 2 Enable ABC to use these simulations to approximate the posterior $\pi(\theta|y)$ without explicit likelihood calculations.

ABC Rejection Sampling Method

The ABC rejection sampling method involves three main steps:

- 1 **Initialize Parameters:** Set up the prior distribution, observed data, distance measure, tolerance threshold, and an integer $N > 0$ for the number of simulations.
- 2 **Sampling Process:** For each i from 1 to N , sample a parameter vector, generate simulated data, and accept the parameter vector if the distance between the observed and simulated data is within the tolerance.
- 3 **Output:** The set of accepted parameter vectors approximates the posterior distribution within the given tolerance.

The ABC rejection algorithm result approximates the true posterior.

$$\pi(\theta|y_{obs}) \approx \hat{\pi}(\theta|d(y_{obs}, y^*) \leq \varepsilon). \quad (3)$$

ABC Rejection Sampling Method

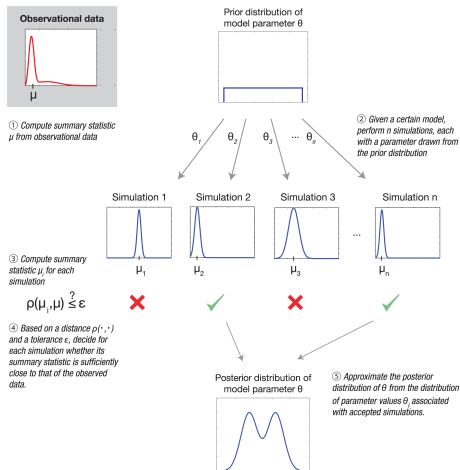


Figure: Illustration of the ABC rejection algorithm.^[3]

ABC Rejection Sampling Method

Key things for effective ABC:

- The distance function $d(\cdot, \cdot)$
- The tolerance ε
- Summary Statistics $S(D)$, such as the mean μ

ABC Rejection Sampling Example

- Observed data simulated from an $\mathcal{N}(5, 1^2)$ distribution.
- Uniform $U[-10, 10]$ prior for parameter θ and then simulate y through $\text{rnorm}(\theta)$.
- 1000 particles accepted with tolerance $\varepsilon = 0.1$. (See Result)

Need for ABC-SMC

ABC- Sequential Monte Carlo (SMC) addresses the limitations of ABC Rejection by refining parameter samples through a sequence of distributions with decreasing tolerance.

- Efficient for large datasets.
- Improves acceptance rate (See the **Sampling Comparison Result**).

ABC-SMC Algorithm

- $\text{SMC} \approx \text{Sequential importance sampling} + \text{Resampling and Perturbation}$.
- Sequence of distributions of interest (time-series problem).
- Approach a single distribution via a sequence that ends at the target.
(See Example followed)

ABC-SMC Algorithm

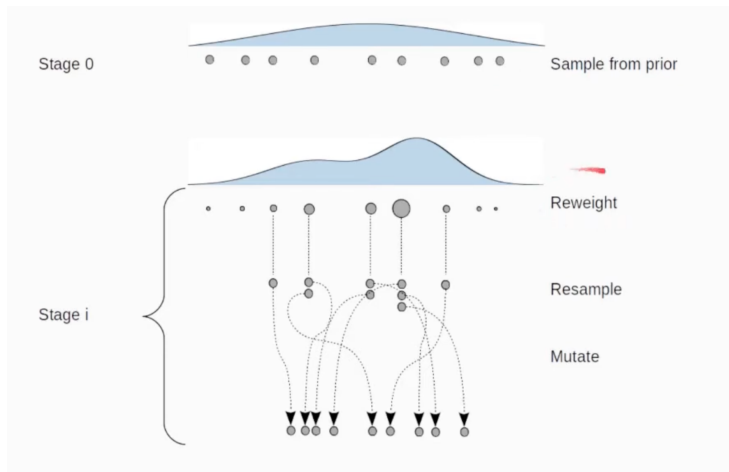


Figure: Illustration of the SMC algorithm resampling and reweighing process.^[7]

ABC-SMC Algorithm

The ABC SMC process consists of iterative steps to approximate the posterior:

- ➊ **Initialization:** Define total generations G , particles N , and a sequence of tolerances $\{\varepsilon_1, \dots, \varepsilon_G\}$.
- ➋ **Sampling and Perturbation:** For each generation, rejuvenate the sample:
 - ➊ Resample parameters from the prior (or from the previous generation).
 - ➋ Move the resampled particles to the next generation using a transition kernel.
- ➌ **Synthetic Data Generation:** Generate synthetic datasets using sampled parameters and compute the distance to the observed data.
- ➍ **Weighting and Acceptance:** Assign weights to samples based on their proximity to observed data and normalize weights across all particles.
- ➎ **Iteration:** Proceed to the next generation with updated tolerances.

ABC-SMC Algorithm

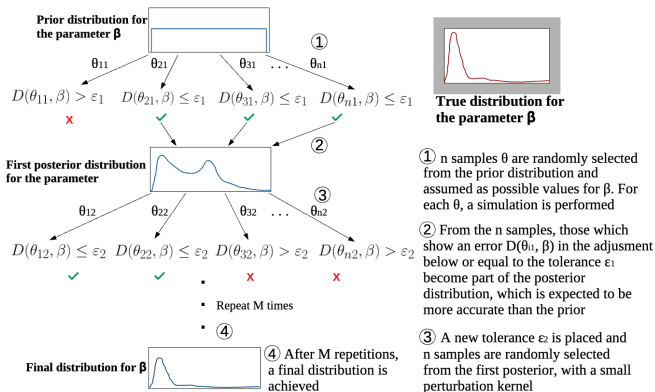


Figure: Illustration of the ABC-SMC algorithm. [3]

ABC-SMC Example

- Observations assumed to be $\mathcal{N}(5, 1^2)$.
- 5 sequential tolerance levels.
- $200 \times 5 = 1000$ particles, aligning with Rejection Sampling.

Sampling Comparison Result

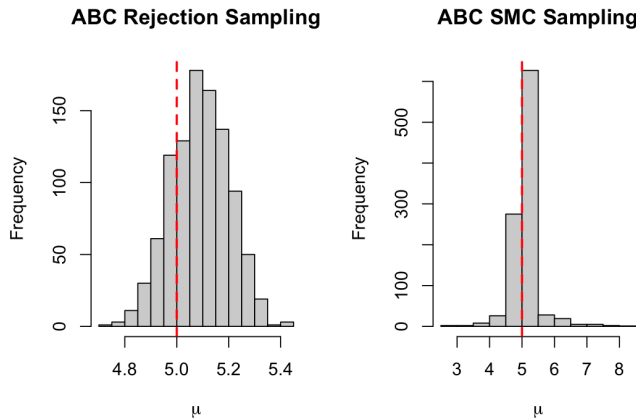


Figure: Posterior distributions of the θ obtained from ABC Rejection and ABC SMC Sampling, the red dashed line indicates the true parameter values

- ABC-SMC in R to model the population of England and Wales from 1801 to 2021 using logistic growth model $y^* \sim p(y|\theta^{(i)})$.^[5]

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right), \quad (4)$$

with the intrinsic growth rate (r) and the carrying capacity (K).

Real-world Application - Population Dynamics

- The algorithm iteratively refines the parameters' estimation through simulated data comparison.
- Each iteration involves:
 - Generating new parameter samples (perturbing the previous generation's samples using a kernel).
 - Simulating population trajectories using these parameters.
 - Comparing these simulations with data using a distance metric.
 - Updating particle weights based on their fit with the data.
- Resulting weighted particles provide an estimation of the posterior distribution for model parameters.

Real-world Application - Population Dynamics

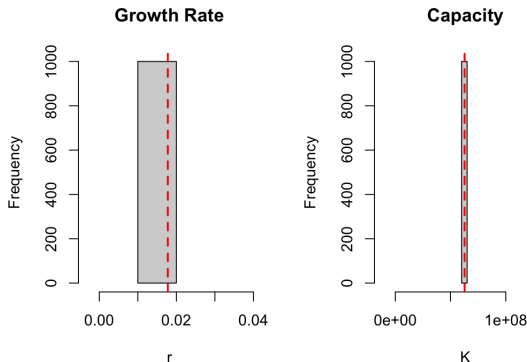


Figure: Posterior distributions of the growth rate r and carrying capacity K parameters, obtained after $g = 10$ generations with $N = 1000$ particles using a numerical logistic growth model. The red dashed line indicates the solution from non-linear least squares fitting of the growth equation for reference.

- Efficacy of ABC-SMC in concentrating posterior distributions.
- Challenges in 'particle degeneracy' as observed above^[1].
- Trade-off between acceptance rate and the number of effective (representative) samples, known as Effective Sample Size (ESS).
- Potential of tailored kernels for balancing the algorithm's steps ^[1].
- Weight calculation using kernel density estimation ^[1].



Sisson, S. A., Fan, Y., & Beaumont, M. (Eds.). (2018). *Handbook of Approximate Bayesian Computation*. 1st edition. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315117195>.



Minter, A., & Retkute, R. (2019). *Approximate Bayesian Computation for infectious disease modelling*. *Epidemics*, 29, 100368. <https://doi.org/10.1016/j.epidem.2019.100368>.



Sunnåker, M., Busetto, A. G., 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>.



Martin, G. M., McCabe, B. P. M., Maneesoonthorn, W., & Robert, C. P. (2014). *Approximate Bayesian Computation in State Space Models*. arXiv:1409.8363 [math.ST].



Office for National Statistics. (2021). *Census 2021*.



Lipkin, L., & Smith, D. (2001). *Mathematical Association of America - Logistic Growth Model*. Retrieved November 16, 2023, from <https://maa.org/book/export/html/115630>.



Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). *Probabilistic programming in Python using PyMC3*. PeerJ Computer Science, 2, e55. <https://doi.org/10.7717/peerj-cs.55>.