Approximate Bayesian Computation

Becky Griffiths   Supervised by Dr Richard Everitt and Ian Roberts

Motivation

- If we have some observed data, x_{obs}, the Bayesian approach to performing inference about a parameter, θ, is to use the posterior distribution of θ:
$$\pi(\theta|x_{obs}) = \frac{\pi(\theta)f(x_{obs}|\theta)}{m(x_{obs})}.$$
- The likelihood, $f(x|\theta)$, is the probability of observing data equal to x given a parameter $\theta \in \Theta$, where Θ is the set of all possible parameters.
- The prior distribution of θ, $\pi(\theta)$, expresses our initial knowledge about the parameter.
- If the normalising constant, $m(x_{obs}) = \int_\Theta f(x_{obs}|\theta)\pi(\theta)\,\mathrm{d}\theta$, is intractable, MCMC methods can be used to sample from the posterior.
- MCMC methods alone cannot help us if we have a model for which we cannot write down the likelihood function, or it is extremely expensive to evaluate.
- But if we can simulate realisations of the model, $x \sim f(\cdot\,|\theta)$, we can use **approximate Bayesian computation** to sample from a distribution 'close' to the true posterior.
- Examples of models where ABC is useful include stochastic differential equations and agent-based models, which are both widely used in ecology and epidemiology.

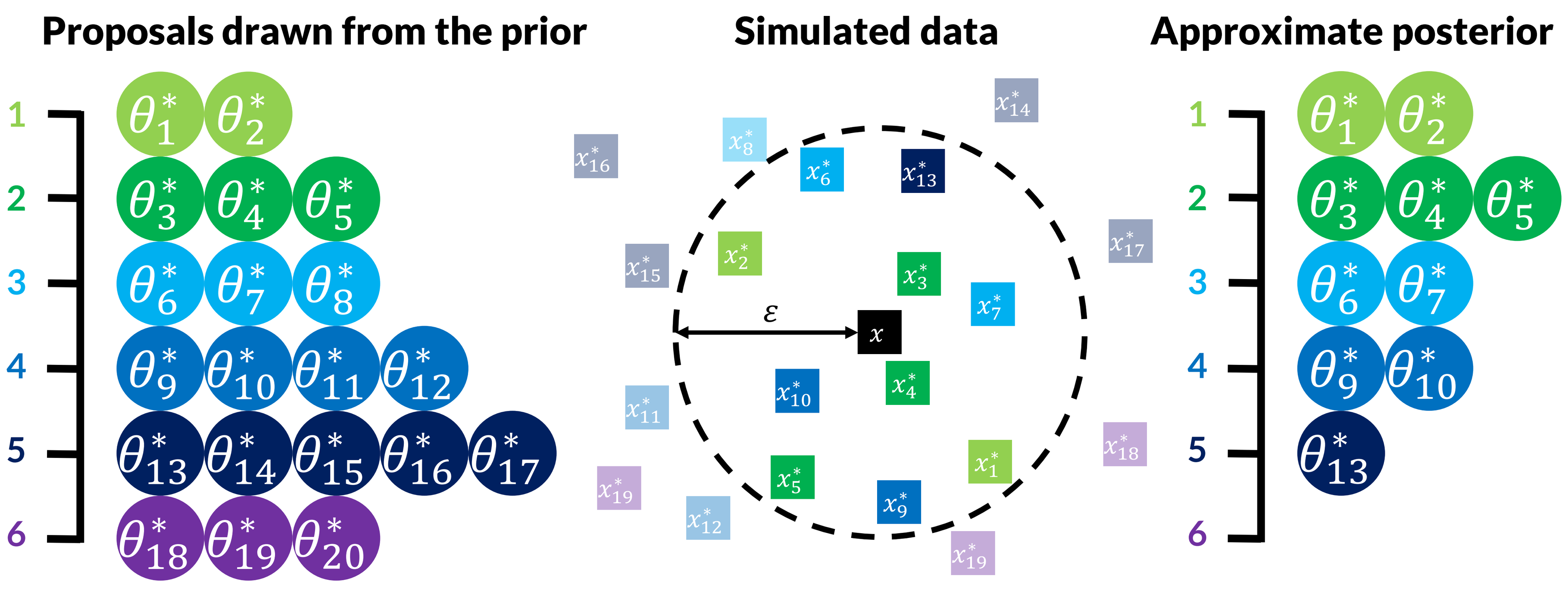


Figure 1: An example of an ABC rejection algorithm

Rejection ABC

- The rejection ABC algorithm follows these steps:
  1. Sample a set of proposal parameters $\theta_1^*, \dots \theta_N^*$ from the prior, $\pi(\theta)$
  2. For each θ_i^*, simulate a set of data, $x_i^* \sim f(x|\theta_i^*)$
  3. For $i = 1, \dots, N$:
     If x_i^* is within a certain tolerance level, ε, of the true data (according to some distance metric) keep θ_i^*, otherwise discard it.
- This procedure generates samples (θ, x) from the joint distribution $\pi(\theta, x|x_{obs}) \propto \pi(\theta)f(x|\theta)\mathbb{I}(|x - x_{obs}| \le \varepsilon)$.
- If we discard the simulated data values then we have samples from the marginal distribution $\pi_\varepsilon(\theta|x_{obs}) \propto \int_\Theta \pi(\theta)f(x|\theta)\mathbb{I}(|x - x_{obs}| \le \varepsilon)\,\mathrm{d}x$.
- As $\varepsilon \to 0$, $\pi_\varepsilon(\theta|x_{obs})$ converges to the true posterior [1].
- The choice of $\varepsilon > 0$ is the first source of approximation in ABC. The second is the use of **summary statistics**, which is motivated by the 'curse of dimensionality'.

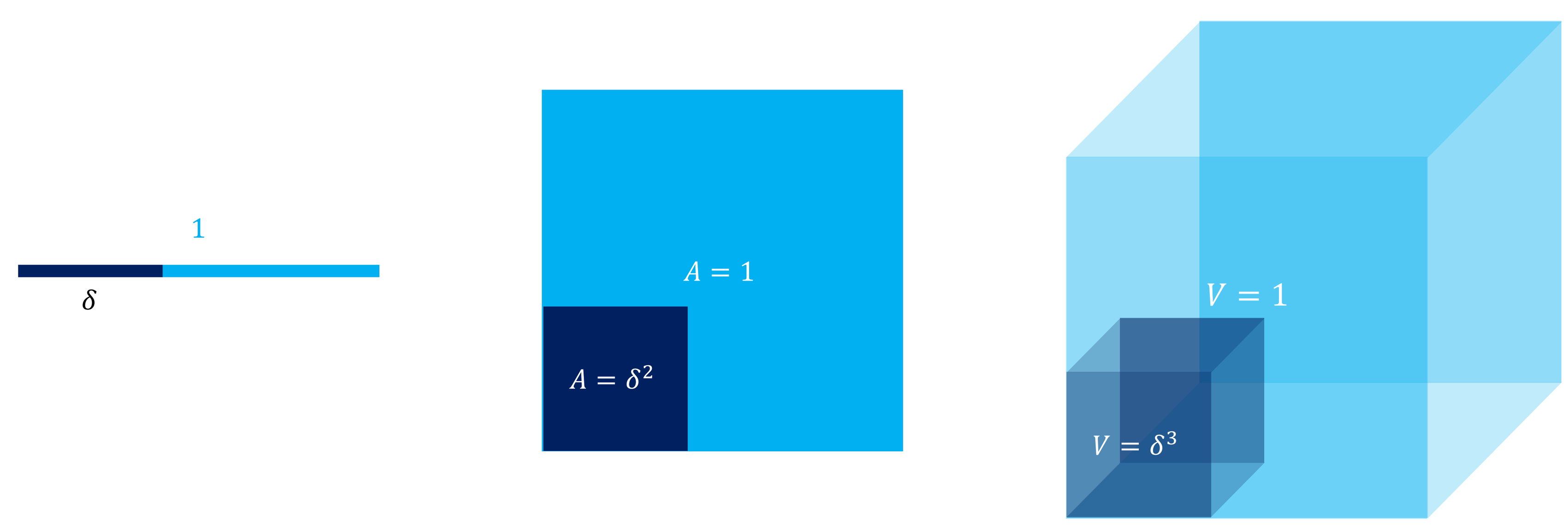
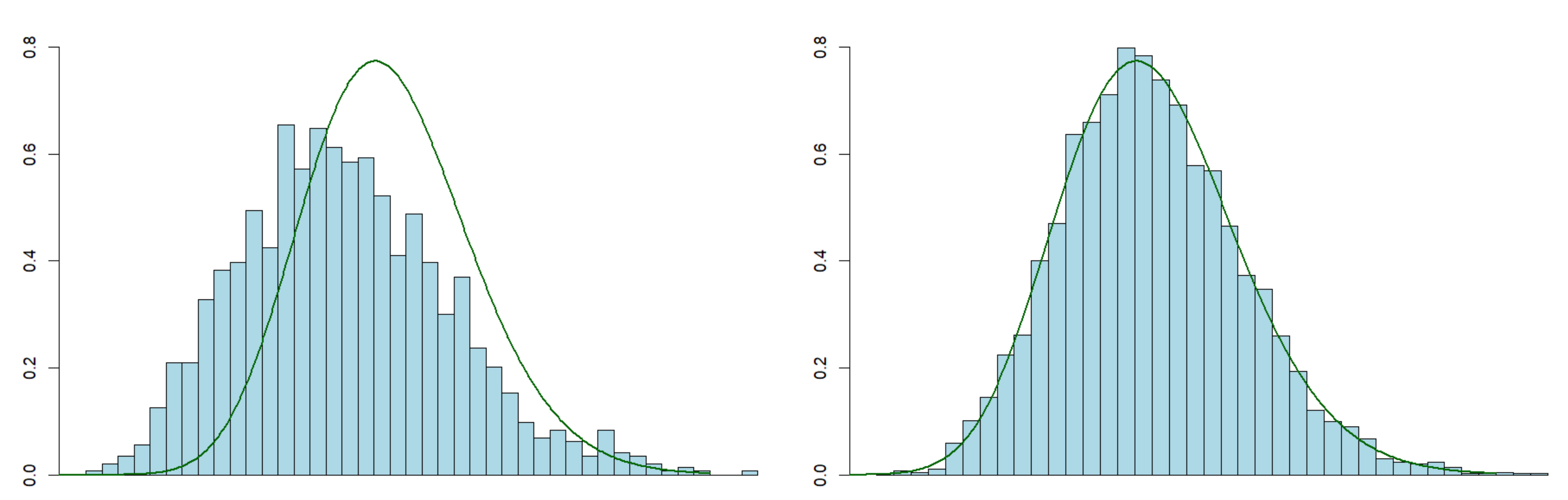


Figure 2: An illustration of the curse of dimensionality

Summary Statistics

- Simulating data close to the true data becomes increasingly unlikely as the dimension of the data grows, even using when using the true parameter value.
- Imagine that the data space is an n-dimensional hypercube with side length 1, and that the tolerance level is 0.5δ for some for some $\delta < 1$. When $n = 1$, the fraction of the data space that is acceptably close to the true data point is δ. For n-dimensions, the fraction is proportional to δ^n, which goes to zero as $n \to \infty$.
- To deal with this issue lower dimensional summaries, $s(x_{obs})$, are typically used in place of the full data.
- I performed rejection ABC on a Bayesian model with a Poisson likelihood and gamma prior. Figures 3a and 3b show histograms of the thousand best sample parameters out of a million based on the L^1 distance of the simulations they produced to the full true data, and the L^1 distance of the means of the simulations they produced to the mean of the true data, respectively. The analytical posterior is plotted in green.
- Selecting good summary statistics for more complex models can be difficult, and is an area of active research. Ideally we would want $s(x_{obs})$ to be both low dimensional and **sufficient** for the model we are considering, meaning that $\pi(\theta|x_{obs}) = \pi(\theta|s(x_{obs}))$ but this is often an impossible requirement [2].

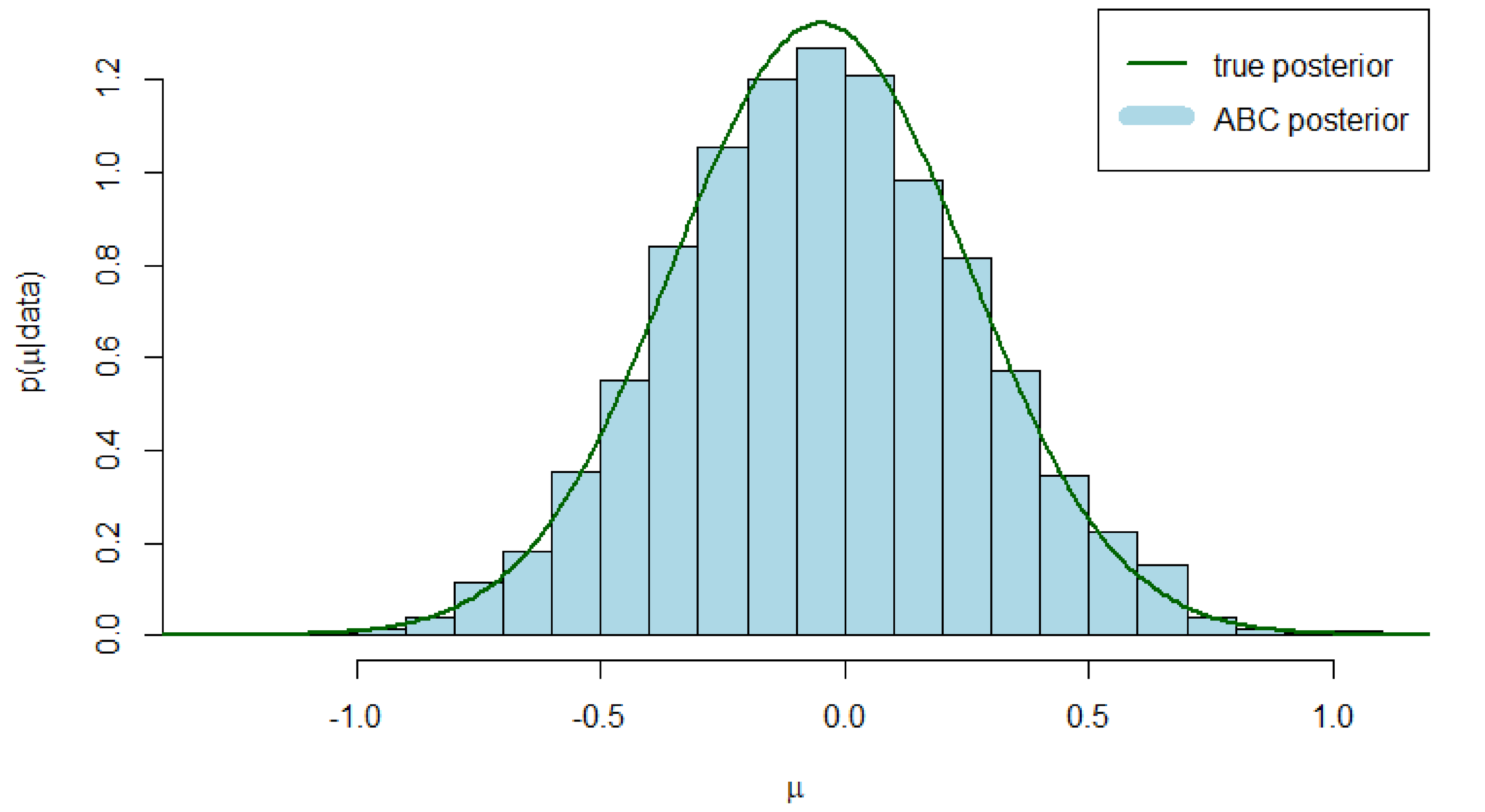


(a) $\pi(\lambda|x_{obs})$ (b) $\pi(\lambda|s(x_{obs}))$, $s(x) = \bar{x}$

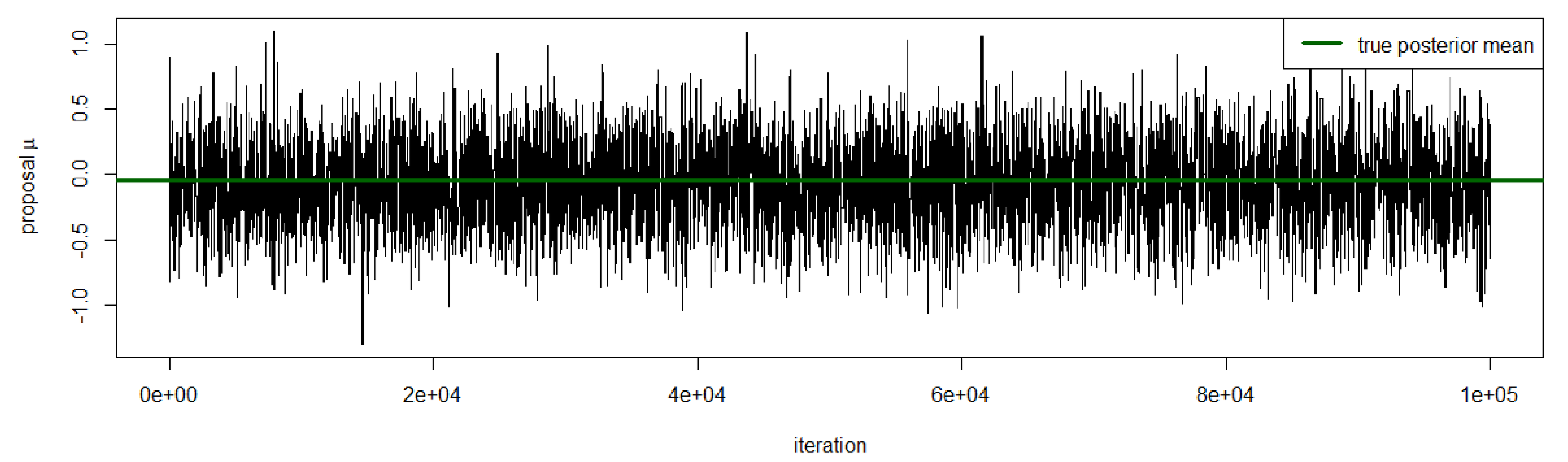
Figure 3: The effect of summary statistics on ABC approximations

ABC MCMC

- Rejection ABC has the downside that unless the prior is close to the ABC posterior, many of the simulated parameter values will be rejected.
- We can run it until we find our first accepted proposal, θ^*, and then explore the parameter space more efficiently using MCMC, starting the chain at θ^*.
- The Metropolis-Hastings algorithm with proposal density $q(\theta'|\theta)$ and acceptance probability $\alpha = \frac{\pi(\theta')\mathbb{I}(|x' - x_{obs}| \le \varepsilon)q(\theta|\theta')}{\pi(\theta)\mathbb{I}(|x - x_{obs}| \le \varepsilon)q(\theta'|\theta)}$, where x' is data simulated using θ', targets the joint ABC posterior proportional to $\pi(\theta)f(x|\theta)\mathbb{I}(|x - x_{obs}| \le \varepsilon)$ [3].
- Since $\mathbb{I}(|x' - x_{obs}| \le \varepsilon)$ is equal to either zero or one, if the acceptance ratio without this term leads to a rejection, we can reject without needing to simulate x' [4].
- I ran Metropolis-Hastings ABC with a normal prior for 100,000 iterations on observations from a normal likelihood with known variance, and the results are displayed in Figure 4.



(a) The analytical posterior, and the one produced by MCMC ABC



(b) Trace plot of the Markov chain produced by the algorithm

Figure 4: Inferring the mean of a normal distribution with ABC MCMC

ABC SMC

- We want our tolerance level to be as small as possible, to produce the best approximation to the true posterior.
- But setting ε very low can mean few proposals are accepted, leading to poor mixing if we use ABC MCMC, or a high computational cost to getting a good number of samples from $\pi_\varepsilon(\theta|x_{obs})$ if we use rejection.
- Sequential Monte Carlo methods are a class of Monte Carlo methods used to sample from sequences of distributions via importance sampling and resampling [5].
- We can use SMC to iteratively sample from a sequence of ABC posteriors, $\pi_{\varepsilon_0}(\theta|x_{obs})$, … $\pi_{\varepsilon_n}(\theta|x_{obs})$ with decreasing ε values starting with a high ε_0, so that a good acceptance rate is maintained at each iteration.

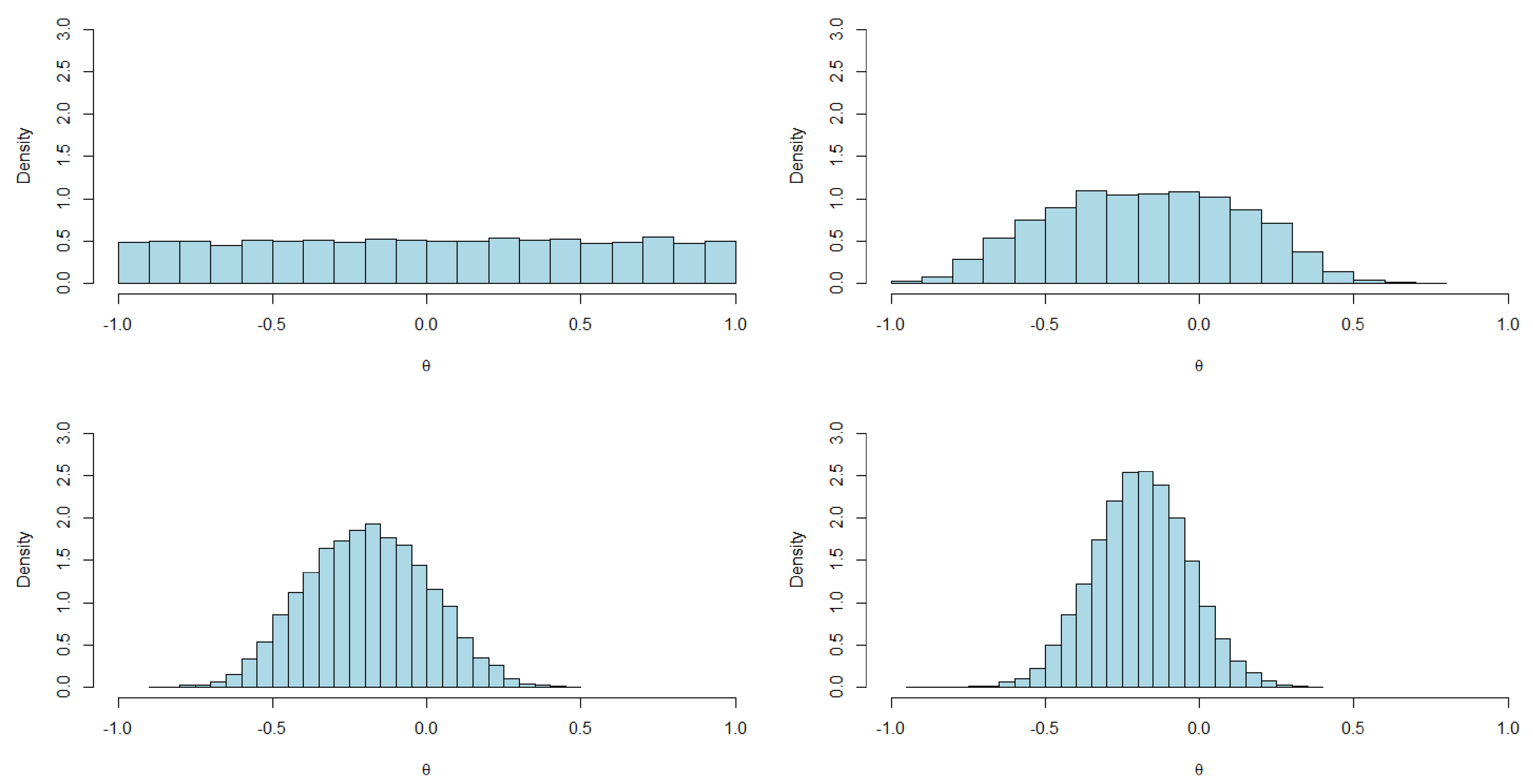


Figure 5: Histograms of samples from iterations of SMC ABC starting with a uniform prior

References

[1] Scott A. Sisson and Yanan Fan. "ABC Samplers". In: *Handbook of Approximate Bayesian Computation*. Chapman and Hall, 2019

[2] Dennis Prangle. "Summary Statistics in Approximate Bayesian Computation". In: *Handbook of Approximate Bayesian Computation*. Chapman and Hall, 2019

[3] Richard Wilkinson. "Approximate Bayesian computation (ABC) gives exact results under the assumption of model error". In: *Statistical Applications in Genetics and Molecular Biology* 12.2 (2013), pp. 129–141. DOI: 10.1515/sagmb-2013-0010.

[4] Umberto Picchini and Julie Lyng Forman. "Accelerating inference for diffusions observed with measurement error and large sample sizes using approximate Bayesian computation". In: *Journal of Statistical Computation and Simulation* 86.1 (2016), pp. 195–213. DOI: 10.1080/00949655.2014.1002101.

[5] Pierre del Moral, Arnaud Doucet, and Ajay Jasra. "An adaptive sequential Monte Carlo method for approximate Bayesian computation". In: *Statistics and Computing* 22 (2012), pp. 1009–1020 DOI: 10.1007/s11222-011-9271-y.