Approximate Bayesian Computation MCMC

Hadi Hammoud Nils Fleischmann

Abstract

In this project, we explore the application of Rejection Sampling and Markov Chain Monte Carlo (MCMC) algorithms in the context of Approximate Bayesian Computations (ABC). Initially, the algorithms were implemented on an academic example with known likelihood and posterior functions, where the ABC MCMC algorithm outperformed the ABC Rejection algorithm in terms of acceptance rate and the approximation of the posterior, as measured by the Kullback-Leibler (KL) divergence. Subsequently, the project delves into the practical application of ABC for modeling the pharmacokinetics of Theophylline. The methodology involves employing MCMC for parameter inference and designing an efficient Monte Carlo approach to estimate the concentration of Theophylline in blood after 12 hours of injection. The report provides a comprehensive assessment of the effectiveness and efficiency of the proposed methods in enhancing both the accuracy of parameter estimation and the computational efficiency of Bayesian inference within complex pharmacokinetic models.

1. Introduction

In the realm of statistical inference, estimating a set of parameters from observed data is a fundamental task. Bayesian statistics provides a powerful framework for parameter estimation by incorporating prior knowledge into the analysis. At the core of Bayesian inference lies Bayes' rule, a mathematical formula that allows us to update our understanding of the parameters based on both prior beliefs and observed data. In this context, one assumes the parameters follow a prior distribution $\pi(\theta)$, representing our initial beliefs or assumptions regarding the values that the parameters could take. After having observed some data D, this prior belief is updated by means of the likelihood function $P(D|\theta)$, which expresses how well the chosen parameters θ explain the observed data. The posterior distribution of interest, $f(\theta|D)$, is then determined by Bayes' rule:

$$f(\theta|D) = \frac{P(D|\theta)\pi(\theta)}{P(D)} \tag{1}$$

where $P(D) = \int P(D|\theta)\pi(\theta)d\theta$, called the evidence, represents the normalizing constant.

The power of Bayes' rule lies in its ability to iteratively update our beliefs as more data becomes available, providing a refined estimate of the parameters. In the context of generating observations from the posterior distribution $f(\theta|D)$ through stochastic simulation methods, a critical dependency often lies in possessing explicit knowledge of the likelihood function $P(D|\theta)$, potentially up to a multiplicative constant. However, when dealing with intricate probabilistic models, the assessment of such likelihoods is frequently either unfeasible or computationally challenging. In such scenarios, resorting to likelihood-free methods, prominently Approximate Bayesian Computation (ABC), becomes necessary.

ABC methods have significantly mitigated computational complexities, enabling the handling of numerous real-world problems with increased tractability. They have been successfully applied in a variety of fields, including molecular genetics (Marjoram & Tavaré, 2006), ecology (Sisson et al., 2007), epidemiology (Tanaka et al., 2006), and evolutionary biology (Leman et al., 2005). In the scope of this project, we focus specifically on the application of ABC to model the pharmacokinetics of Theophylline.

2. Methods

2.1. ABC-Rejection

The simplest ABC rejection algorithm scheme can be found in Algorithm 1 below. The algorithm operates by sampling potential parameters θ^* from the prior distribution $\pi(\theta)$. Subsequently, a data sample D^* is generated based on these candidate parameters θ^*. The algorithm then assesses the similarity between the generated data and the observed data D using a predefined discrepancy metric $\rho(\cdot, \cdot)$ and a tolerance value ε. During the acceptance/rejection step, candidate parameters are deemed samples from the posterior distribution if the simulated data closely aligns with the observed data according to the chosen discrepancy metric $\rho(\cdot, \cdot)$ and tolerance ε. If not, a new candidate is drawn from the prior distribution. This process is iterated until a sufficiently large sample size N, approximately following the posterior distribution, is obtained.

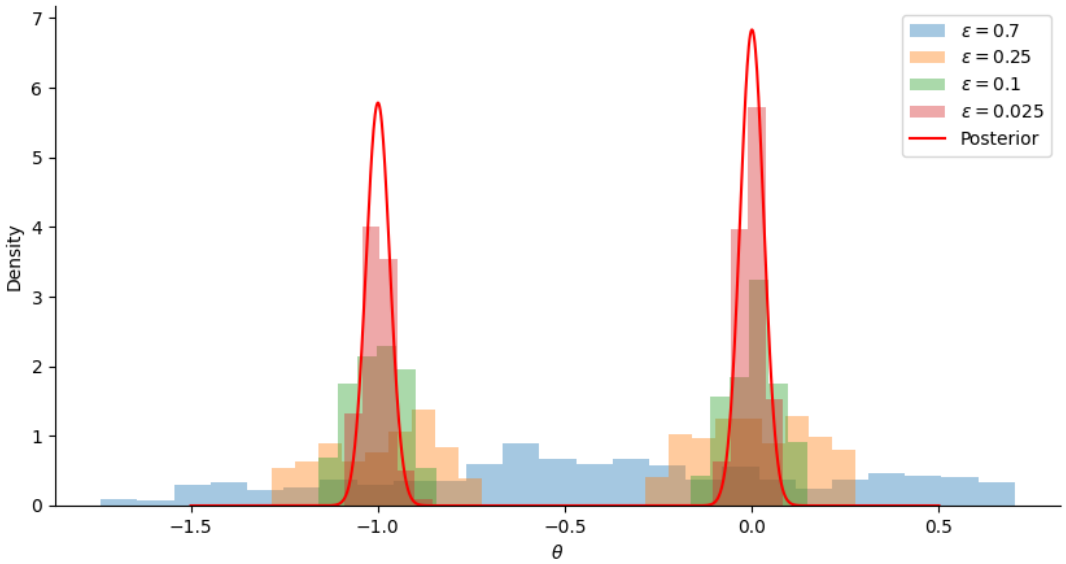

Figure 1. Overlay of histograms representing the distribution of $N = 500$ samples generated by Algorithm 1 for the tolerance levels $\varepsilon = \{0.7, 0.25, 0.1, 0.025\}$. The true posterior distribution, given by (2), is shown by the red line.

Algorithm 1 Basic ABC rejection algorithm

for $i = 1, \ldots, N$ **do**
  Sample candidate parameters from prior $\theta^* \sim \pi(\cdot)$
  Generate data from model given θ^*, $D^* \sim P(\cdot|\theta^*)$
  if $\rho(D^*, D) < \varepsilon$ **then**
    Set $\theta_i \leftarrow \theta^*$
  else
    Go back to Step 2.
  end
end

Note that in fact, Algorithm 1 yields samples from the approximate posterior distribution.

Theorem 1. Algorithm 1 generates samples distributed as $f(\theta \mid \rho(D^*, D) < \varepsilon)$

Proof. The proof is a straightforward implication of Bayes rule with the following observation: let X be a sample generated from Algorithm 1 and θ^* be the sample from the prior, then $X \sim \theta^* | \rho(D^*, D) < \varepsilon$, or equivalently

$$\begin{aligned} X &\sim \pi(\theta^* | \rho(D^*, D) < \varepsilon) \\ &= \frac{\mathbb{P}(\rho(D^*, D) < \varepsilon \mid \theta^*)\pi(\theta^*)}{\mathbb{P}(\rho(D^*, D) < \varepsilon)} \\ &= f(\theta^* \mid \rho(D^*, D) < \varepsilon) \end{aligned}$$

□

As $\varepsilon \to \infty$ the samples generated will be distributed according to the prior. If, on the contrary, $\varepsilon = 0$, and the observation D^* is accepted only if $D^* = D$, then the algorithm generates observations from the true posterior density $f(\theta|D)$.The selection ε thus represents a trade-off between the method's computational feasibility and accuracy.

2.2. ABC-MCMC

The sampling strategy outlined in Algorithm 1 might not be the best choice in the context of ABC as acceptance rates can be very low. When dealing with complex models or high-dimensional parameter spaces, the basic rejection algorithm may generate a significant number of proposed samples from $\pi(\theta)$ that are rejected during the acceptance/rejection step. A low acceptance rate means that a large portion of the computational effort is spent generating samples that do not contribute to the final posterior approximation.

In these scenarios, sophisticated approaches have been developed, such as integrating the ABC scheme with the Metropolis-Hastings framework, as illustrated in Algorithm 2.

Notice that the candidate parameter vector is created using an arbitrary proposal transition density $q(\cdot, \cdot)$ and is accepted based on a Metropolis–Hastings type acceptance probability. In this probability calculation, the (intractable) likelihood ratio $P(\cdot \mid \theta^*)/P(\cdot \mid \theta_i)$ is approximated by 1

Algorithm 2 ABC-MCMC algorithm

Initialize θ_0
for $i = 0, \ldots, N$ **do**
  Sample candidate parameters $\theta^* \sim q(\theta_i, \cdot)$
  Generate data from model given θ^*, $D^* \sim P(\cdot|\theta^*)$
  if $\rho(D^*, D) < \varepsilon$ **then**
    let $\alpha = \min\left(1, \frac{\pi(\theta^*)q(\theta^*,\theta_i)}{\pi(\theta_i)q(\theta_i,\theta^*)}\right)$
    Set $\theta_{i+1} \leftarrow \theta^*$ with $p = \alpha$ and $\Theta_{i+1} \leftarrow \Theta_i$ otherwise
  else
    Set $\Theta_{i+1} \leftarrow \Theta_i$
  end
end

if the simulated and observed data are deemed sufficiently close according to the chosen metric $\rho(\cdot, \cdot)$, and by 0 otherwise. This approximation will be helpful in proving the next theorem.

Theorem 2. $f(\theta \mid \rho(D^*, D) < \varepsilon)$ is the stationary distribution of the chain generated by Algorithm 2

Proof. We will show that $f(\theta \mid \rho(D^*, D) < \varepsilon)$ satisfies the detailed balance condition, implying that it is indeed the stationary distribution of the chain. Denote the transition step of the chain by $r(\theta \to \theta')$, and (without loss of generality) choose $\theta' \neq \theta$ satisfying

$$\frac{\pi(\theta')\, q(\theta' \to \theta)}{\pi(\theta) q(\theta \to \theta')} \leq 1$$

and let $\delta := \rho(D^*, D) < \varepsilon$, then

$$\begin{aligned}
&f(\theta \mid \delta) r(\theta \to \theta') \\
&= f(\theta \mid \delta) \cdot \left\{ q(\theta \to \theta') \cdot \frac{\pi(\theta') q(\theta' \to \theta)}{\pi(\theta) q(\theta \to \theta')} \right\} \\
&= \frac{\mathbb{P}(\delta \mid \theta)\pi(\theta)}{\mathbb{P}(\delta)} \cdot \left\{ q(\theta \to \theta') \cdot \frac{\pi(\theta') q(\theta' \to \theta)}{\pi(\theta) q(\theta \to \theta')} \right\} \\
&= \frac{\mathbb{P}(\delta \mid \theta')\pi(\theta')}{\mathbb{P}(\delta)} \cdot q(\theta' \to \theta) \\
&= f(\theta' \mid \delta) q(\theta' \to \theta) \\
&= f(\theta' \mid \delta) r(\theta' \to \theta)
\end{aligned}$$

□

In the outlined proof, the approximation of the likelihood ratio was used in the third equation and the assumption that the acceptance ratio being ≤ 1 was used in the last equation. Note that the case where the acceptance ratio is > 1 is analogous.

3. Experiments

3.1. Academic Example

We first start by an example with known likelihood posterior distribution functions, allowing for a direct comparison.

In this example, the observed data $D = \{x_i\}_{i=1}^M \subset \mathbb{R}$ is an i.i.d sample drawn with probability $1/2$ from $\mathcal{N}(\theta, \sigma_1^2)$ and with probability $1/2$ from $\mathcal{N}(\theta + a, \sigma_1^2)$. The prior is taken to be $\pi = \mathcal{N}(0, \sigma^2)$. Then, the posterior distribution is a Gaussian mixture given by

$$\begin{aligned}
f(\theta|D) = \alpha\mathcal{N}\left(\frac{\sigma^2}{\sigma^2 + \sigma_1^2/M}\bar{x}, \frac{\sigma_1^2}{M + \sigma_1^2/\sigma^2}\right) \\
+ (1-\alpha)\mathcal{N}\left(\frac{\sigma^2}{\sigma^2 + \sigma_1^2/M}(\bar{x} - a), \frac{\sigma_1^2}{M + \sigma_1^2/\sigma^2}\right)
\end{aligned} \tag{2}$$

with

$$\alpha = \frac{1}{1 + \exp\left\{a\left(\bar{x} - \frac{a}{2}\right)\frac{M}{M\sigma_2^2 + \sigma_1^2}\right\}} \tag{3}$$

where $\bar{x} = \frac{1}{M}\sum_{i=1}^M x_i$ denotes the sample mean of the data and $\mathcal{N}(\mu, \sigma^2)$ denotes the density of a Gaussian random variable with mean μ and variance σ^2. We consider the following parameters: $M = 100$, $\sigma_1^2 = 0.1$, $\sigma^2 = 3$, $a = 1$ and we assume that the sample mean of the data is exactly $\bar{x} = 0$.

ABC-Rejection. Our first approach is to approximate the posterior with Algorithm 1. We consider tolerances $\varepsilon = \{0.7, 0.25, 0.1, 0.025\}$ to draw until $N = 500$ samples are accepted, with discrepancy metric defined as

$$\rho(S(D^\star), S(D)) = |\bar{x}^\star - \bar{x}| \tag{4}$$

where $\bar{x}^* = \frac{1}{M}\sum_{i=1}^M x_i^*$ is the sample mean of the generated data $D^* = \{x_i^*\}_{i=1}^M$ following the distribution of the observed data.

The distribution of $\rho(S(D^\star), S(D))$ is displayed in figure 2. As expected, smaller tolerances reduce the likelihood of obtaining a sample D^* that meets the acceptance criterion.

The parameter ε plays a crucial role as it dictates the quality of the approximation. Figure 1 displays the histograms of the samples generated by Algorithm 1 for different tolerance levels. Evidently, For $\varepsilon = 0.7$, the histogram does not appear to align with the true posterior and seems to be influenced by the prior distribution. However, as the tolerance decreases, the histograms of the sampled data exhibit a convergence towards the true posterior. This observation indicates that Algorithm 1, when applied with a sufficiently small tolerance, effectively approximates the posterior for the given academic example.

Table 1 further illustrates this trade-off. For each value of ε, we present the acceptance rate and measure the quality of

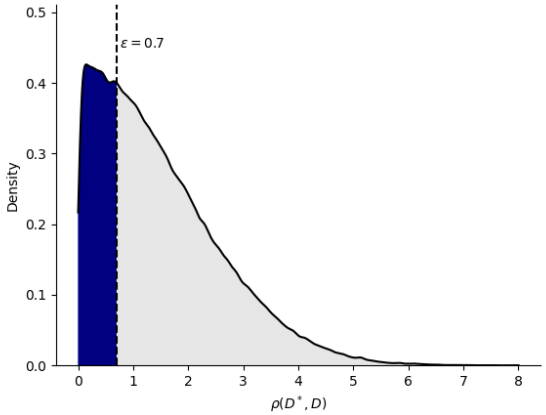

Figure 2. The distributions of proposed distance values by Algorithm 1. The blue region highlights the portion of the distribution that satisfies the acceptance threshold for a tolerance level of $\varepsilon = 0.7$.

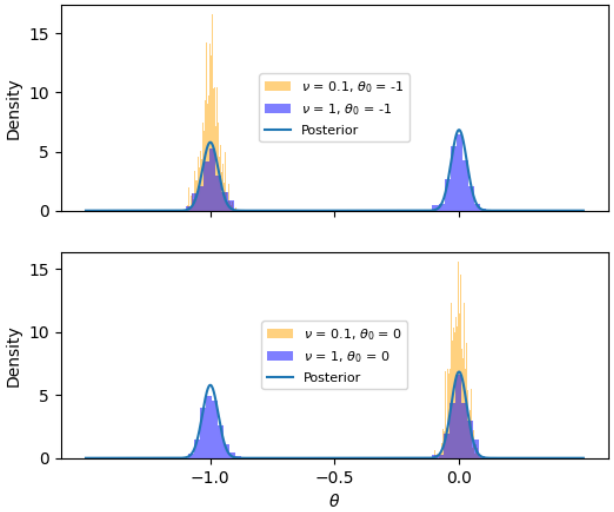

Figure 3. Histograms of samples generated from Algorithm 2 with $\nu = \{0.1, 1\}$ and two different initial states $\theta_0 = \{0, -1\}$

approximation by calculating the KL divergence between a kernel density estimation of the generated samples and the true posterior. At $\varepsilon = 0.7$, the acceptance rate stands at a reasonable 0.29; however, the KL divergence is nearly 40, signifying a poor approximation of the posterior. In contrast, for $\varepsilon = 0.025$, the KL divergence approaches zero, albeit with a lower acceptance rate of 0.01. This implies the necessity to simulate approximately 50,000 datasets D^* to achieve a sample size of 500. While this is feasible for the simple academic model in question, it may be impractical for applications involving more complex models due to the increased computational demand.

ABC-MCMC To mitigate the high rejection rates observed with Algorithm 1, we now apply ABC-MCMC to the given academic problem. We use random walk proposal $q(\theta, \cdot) \sim \mathcal{N}(\theta, \nu^2)$ and initial state $\theta_0 = 0$. Experiments were performed with different $\nu's$ and run long enough to have an effective sample size $N_{eff} \approx 500$.

While investigating the impact of ν, we observed its crucial role in shaping the 'perceptive scope' of the chain—determining which proposed regions it encompasses. While this is the natural role of the standard deviation, in the context of our posterior distribution, it implies that small $\nu's$ lead to high sensitivity to the initial state . To illustrate, this sensitivity is pronounced due to the true posterior being a mixture of two gaussians, resembling two distinct populations. Consequently, a chain with small ν (relative to the distance between the centers of the gaussians) tends to concentrate or become stuck around the population closer to its initial state, limiting its ability to reach the other. Figure 3 depicts our conclusion where we experimented the samples generated from Algorithm 2 with $\nu = \{0.1, 1\}$, while varying the initial state $\theta_0 = \{0, -1\}$ and fixing $\varepsilon = 0.025$. It is evident that chains with $\nu = 0.1$ primarily focused around the nearby population and didn't get the opportunity to explore the other population. On the contrary, chains with $\nu = 1$, although initially centered on one population, was proposed samples from the other population, resulting in coverage of the entire distribution.

We then compared the performance of Algorithms 1 and 2 (with $\nu = 1$) in terms of acceptance rates and approximation of the posterior distribution. Table 1 displays the numerical comparison. The results illustrates the success of MCMC in achieving its objective. It consistently achieved higher acceptance rates compared to the rejection algorithm, without compromising on the accuracy of posterior approximation. Across all chosen thresholds, Algorithm 2 exhibited acceptance rates that were strictly higher than those of Algorithm 1, while also better approximating the posterior in 3 of the 4 tests.

3.2. A model in pharmacokinetics

In the previous subsection, we explored a simple academic example to demonstrate the principles and effectiveness of the ABC-MCMC algorithm. To delve deeper into this algorithm, we now consider a real-world problem. Specifically, our focus is on a dynamical model of the pharmacokinetics of Theophylline, a drug used in the treatment of asthma and chronic obstructive pulmonary disease. This model represents the concentration of Theophylline in the bloodstream, X_t, at any given time t, using the following stochastic differential equation:

Tolerance	ABC Rejection		ABC MCMC	
	Acceptance Rate	KL-Divergence	Acceptance Rate	KL-Divergence
$\varepsilon = 0.7$	0.291 ± 0.022	39.857	0.402 ± 0.014	38.501
$\varepsilon = 0.25$	0.098 ± 0.008	8.779	0.157 ± 0.007	8.991
$\varepsilon = 0.1$	0.042 ± 0.003	1.154	0.064 ± 0.003	1.054
$\varepsilon = 0.025$	0.011 ± 0.001	0.061	0.015 ± 0.001	0.036

Table 1. Acceptance rates (with 95 % CI interval) and Kullback-Leibler (KL) Divergences for Algorithms 1 and 2. The KL Divergences $D_{\mathrm{KL}}(\cdot \mid\mid f(\theta \mid D))$ were computed by approximating the sampled distribution with kernel density estimation, utilizing a bandwidth of 0.03.

$$dX_t = \left(\frac{DK_a e^{-K_a t}}{Cl} - K_e X_t \right) dt + \sigma dW_t \qquad (5)$$

where D is the known oral dose of the drug, K_e is the elimination rate constant, K_a the absorption rate constant, Cl denotes the drug clearance, and σ is the intensity of intrinsic stochastic noise, driven by Brownian motion W_t.

We will consider an experimental design for a single subject involving nine blood samples taken at times $\{t_1, \ldots, t_9\}$ = $\{0.25, 0.5, 1, 2, 3.5, 5, 7, 9, 12\}$ hours post-dosing. The oral dose is set at $D = 4$ mg, administered from $t_0 = 0_+$. The initial drug concentration in the blood is $X_0 = 0$.

For approximating the stochastic differential equation, we always employ the Euler-Maruyama on a fine time grid. Following (Donnet & Samson, 2008), we discretize each interval of consecutive times $[t_i, t_{i+1}]$ into 20 steps of equal size. This method is expected to provide a close approximation to the true model.

We utilize the parameters $\theta = (K_e, K_a, Cl, \sigma) = (0.08, 1.5, 0.04, 0.2)$ and generate synthetic measurement data $D = \{x_i\}_{i=1}^9$ using the Euler-Maruyama discretization. Figure 4 displays 1000 samples generated with these parameters, highlighting in blue the sample D used in subsequent analysis. It is observed that Theophylline concentration initially increases in the first 4 hours post-intake and then gradually decreases.

Again, our aim is to approximate the posterior $f(\theta|D)$ after observing the nine blood samples D. We consider the following log-normal distributions as priors for the parameters of interest θ: $\log K_e \sim \mathcal{N}(-2.7, 0.62)$, $\log K_a \sim \mathcal{N}(0.14, 0.42)$, $\log Cl \sim \mathcal{N}(-3, 0.82)$, $\log \sigma \sim \mathcal{N}(-1.1, 0.32)$. As the initial state for the ABC-MCMC algorithm (Algorithm 2) we choose $\theta_0 = (0.07, 1.15, 0.05, 0.33)$.

Given that θ comprises four distinct parameters, selecting an appropriate proposal distribution in the multivariate context becomes challenging. This is due to the differing scales and potential correlations between the parameters. To address this, we have chosen to employ an adaptive Metropolis algorithm, where the proposal distribution is iteratively updated

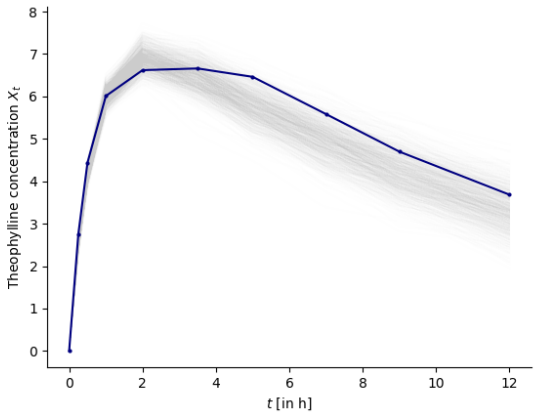

Figure 4. 1000 samples from the pharmacokinetic model generated for the parameters $(K_e, K_a, Cl, \sigma) = (0.08, 1.5, 0.04, 0.2)$. The blue line represents the sample D based on which the posterior distribution was based on.

using information from previous iterations, akin to the approach described in (Haario et al., 2001). As with the prior, we opt for a log-gaussian distribution for the proposal:

$$q\left(\theta_t, \log(\cdot)\right) \sim \mathcal{N}\left(\left(\log K_e, \log K_a, \log Cl, \log \sigma\right), C_t\right)$$

with the mean at the current point θ_t and the covariance matrix $C_t = C_t(\theta_0, \ldots, \theta_{t-1})$. In line with our prior knowledge, we set the initial covariance matrix as: $C_0 = diag\,(0.62, 0.42, 0.82, 0.32)$, where the variances correspond to those of the prior. We choose an initial period, denoted as $t_0 = 500$, and define the covariance matrix update for $t > t_0$ as:

$$C_{t+1} = \frac{t-1}{t} C_t + \frac{s_4}{t} \left(t\bar{\theta}_{t-1}\bar{\theta}_{t-1}^T - (t+1)\,\bar{\theta}_t\bar{\theta}_t^T + \theta_t\theta_t^T + \varepsilon I \right)$$

where $\bar{\theta}_{t-1}$ is the mean of the samples up to time $t-1$, and εI is a diagonal matrix for numerical stability. As the

scaling parameter, we adopt the value $s_4 = (2.4)^2/4$ from (Gelman et al., 1996).

For the discrepancy metric $\rho(D^*, D)$, we consider the conditional expectation of θ given D, denoted $E(\theta|D)$. To approximate this expectation, we employ a linear regression model

$$S(D) = E(\theta|D) = \beta_0 + \beta_1 x_1 + \ldots + \beta_9 x_9 + \xi$$

We sample different values for $\theta^{(1)}, \ldots, \theta^{(p)}$ using the prior distribution, then generate corresponding data $D^{(1)}, \ldots, D^{(p)}$ for $p = 9000$. Then the linear regression model is fitted on the pairs $(\theta^{(i)}, D^{(i)})$, assuming that the noise $\xi = (\xi_1, \ldots, \xi_4)$ is a random vector with zero mean, independent components, and constant variance. Finally, the discrepancy is computed as a weighted Euclidean norm between the summary statistics: $\rho(S(D^*), S(D)) = \|S(D^*) - S(D)\|$, where

$$\|\theta\|_2 = \sum \frac{\theta_i^2}{\theta_0^2}$$

is a weighted Euclidean norm. To gain a clearer understanding of this discrepancy, we obtained 10,000 samples $\theta^{(i)}$ from the prior distribution, generated corresponding data $D^{(i)}$, and computed the discrepancy $\rho(S(D^{(i)}), S(D))$. The distribution of these discrepancies is depicted in Figure 5. Surprisingly, a significant portion of the probability mass is concentrated between 0 and 1. Consequently, we might anticipate relatively low rejection rates if we do not choose our tolerance sufficiently small.

To estimate the posterior distribution, we ran the ABC-MCMC algorithm for 10,000 iterations using tolerances $\varepsilon = \{0.25, 0.7, 1\}$. Table 2 presents the results of these algorithm runs. Notably, a lower tolerance ε leads to a reduced acceptance rate. For instance, at $\varepsilon = 1$, 76.4% of proposals are accepted based on the discrepancy, whereas at $\varepsilon = 0.25$, only 37.9% are accepted. This trend is also evident in the effective sample sizes for each parameter, where lower ε values yield fewer effective samples from the 10,000 iterations. Examining the estimated posterior means, we observe noticeable deviations from the true underlying parameter values used to generate data D.

Further investigation is conducted through the Kernel Density Estimations (KDE) for the marginal distributions of the estimated posteriors, as shown in figure 8. The estimated posteriors appear to strongly resemble the prior, particularly for $\varepsilon = 0.7$ and $\varepsilon = 1$. This alignment is expected given the high acceptance rates; allowing for data with higher discrepancy levels to be accepted may reduce the influence of D on the resulting samples, akin to drawing from the prior.

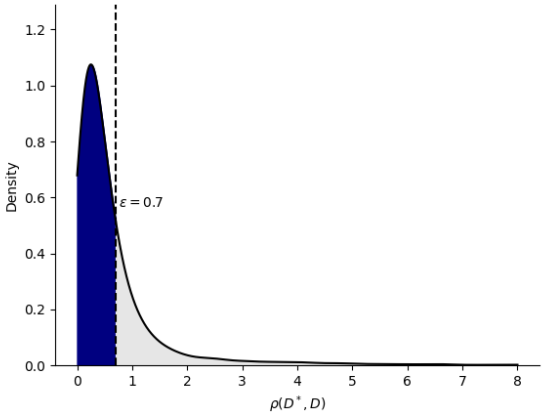

Figure 5. The distributions of discrepancy values for data sampled according to the prior distribution. The blue region highlights the portion of the distribution that satisfies the acceptance threshold for a tolerance level of $\varepsilon = 0.7$.

Tolerance	$\varepsilon = 1$	$\varepsilon = 0.7$	$\varepsilon = 0.25$
Acceptance Rate	0.764	0.663	0.379
$\bar{K_e}$	0.064	0.061	0.056
$\bar{K_a}$	1.153	1.139	1.187
$\bar{Cl}$	0.056	0.052	0.045
$\bar{\sigma}$	0.326	0.331	0.334
$N_{\textbf{eff}}(K_e)$	676	647	465
$N_{\textbf{eff}}(K_a)$	598	597	319
$N_{\textbf{eff}}(Cl)$	774	592	425
$N_{\textbf{eff}}(\sigma)$	714	642	312

Table 2. Summary of ABC-MCMC results for $N = 10,000$ iterations: Acceptance rates, posterior mean parameter estimates, and effective sample sizes for each parameter.

For $\varepsilon = 0.25$, however, deviations from the prior are evident for parameters K_a and Cl, ideally moving towards the true posterior, albeit with a lower acceptance rate. Figure 6 illustrates the two-dimensional kernel density estimations from the run with $\varepsilon = 0.25$. While many pairs show a KDE resembling a normal distribution with independent components, some pairs, like Cl and K_e or Cl and K_a, deviate significantly.

These plots include not only the true underlying θ value but also the predicted value for $E(\theta|D)$, calculated using a linear regression model. Surprisingly, these values differ considerably for parameters σ and Cl, suggesting that the simulated data D may be atypical for the underlying θ or pointing to potential limitations of the linear regression model, especially for values outside the prior distribution used for training. This observation warrants further investigation.

Finally, to assess our choice of the proposal distribution, we examined the trace plots for each parameter, as depicted in Figure 7. Based on these plots, the correlation produced by the MCMC algorithm appears acceptably low, indicating sufficient mixing of the chain.

3.3. Monte Carlo Estimation

Now, let us consider the mean of the approximated posterior distribution for the pharmacokinetic model, which we obtained using the ABC-MCMC method with a tolerance of $\varepsilon = 0.25$. The estimated parameters are $\theta = (0.06, 1.28, 0.05, 0.35)$. Using these parameters, our aim is to estimate $E[X_9]$, the expected concentration of Theophylline in the bloodstream after 12 hours.

In order to get samples for $Z = X_9$, we again utilize the Euler-Maruyama method to approximate the stochastic differential equation. However, since each sample involves simulating the stochastic ODE with 9×20 discretization steps, employing a simple Monte Carlo estimator—which converges to the true mean at a rate of $1/\sqrt{N}$ where N is the number of generated samples—might be impractical. Therefore, our goal is to use a more efficient Monte Carlo estimator that requires fewer samples for a given confidence level by making better use of the available samples.

One approach to achieve this is to reduce the variance of the samples. Many variance reduction techniques rely on variables correlated with the variable of interest, which, in our case, is X_9. Recall the Euler-Maruyama update formula given by:

$$X^{(t+\Delta t)} = X^{(t)} + \left(\frac{DK_a e^{-K_a t}}{Cl} - K_e X^{(t)} \right) \Delta t + W_t,$$

where W_t is a normal random variable with zero mean and variance $\Delta t \sigma^2$. A logical choice for a correlated variable would be to use the realizations of the W_t to construct such a variable, particularly since many attributes of its underlying distribution are known.

Note that for a given discretization, we can compute the true underlying mean of samples generated by the Euler-Maruyama method by omitting the stochastic term W_t from the update formula.

Antithetic variable. One method to achieve variance reduction is by creating an antithetic variable Z_a for Z. This variable should have the same distribution as Z but be negatively correlated with it. Assuming we know how to generate such an antithetic variable we could generate $N/2$ i.i.d. pairs $(Z^{(i)}, Z'^{(i)}) \sim (Z, Z_a)$, for $i = 1, \ldots, \frac{N}{2}$, and estimate the mean using the term

$$\hat{\mu} = \frac{1}{N/2} \sum_{i=1}^{N/2} \hat{Z}^{(i)} = \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\frac{Z^{(i)} + Z'^{(i)}}{2} \right).$$

Following this approach, the variance $\mathrm{Var}[\hat{\mu}]$ is given by $(\mathrm{Var}[Z] + \mathrm{Cov}(Z, Z_a))/N$, which is less than $\mathrm{Var}[\hat{\mu}_{\mathrm{CMC}}]$.

Ensuring Z_a has the same distribution as Z is a strong restriction. One possible method to generate such a variable is by employing the same Euler-Maruyama scheme we use to generate Z. However, at each step, instead of drawing W_t from the normal distribution, we use the realization of W_t for Z and flip its sign. Since W_t is symmetric around zero, Z_a, generated using this method, should follow the same distribution as Z. Additionally, the direct influence of W_t on X_t means that flipping the sign should result in a variable that is negatively correlated with Z.

In fact, for this specific stochastic differential equation, the antithetic variables generated in this manner are perfectly negatively correlated. Furthermore, the mean $(Z + Z_a)/2$ always equals the true mean, so convergence occurs after just one step. This phenomenon is due to the Euler-Maruyama update term only linearly depending on X_t. Figure 9 illustrates this by plotting the concentration curve for one sample and its antithetic variable, alongside the approximated solution of the differential equation without the stochastic term. It is evident that at each time step t, Z and Z_a are symmetrically aligned around the approximated solution of the differential equation.

Control variates. Another method for variance reduction involves identifying a control variate, Y, for Z. This variate should possess a known expectation and be highly correlated with Z. In our example, we can generate such a variable by summing the realizations of W_t used in the Euler-Maruyama update. Given their significant influence on the approximation of the stochastic differential equation, we expect these sums to correlate strongly with Z. Indeed, in our specific case, the correlation coefficient between Y and Z is 0.95, indicating a strong correlation. Additionally, the expectation of the sum of zero-centered random variables is zero. A further advantage is that these variables are inherently obtained during the simulation of Z.

We utilize the control variate Y to compute $Z_\alpha = Z - \alpha(Y - E[Y])$ and then apply Monte Carlo methods to Z_α instead of Z. This results in $\mathrm{Var}[Z_\alpha] = \mathrm{Var}[Z] + \alpha^2 \mathrm{Var}[Y] - 2\alpha \mathrm{Cov}(Z, Y)$. Ideally, α is chosen as $\mathrm{Cov}(Z, Y)/Var(Z)$ to minimize variance. However, the actual covariance value is unknown, so we use a pilot run with 100 pairs of (Z, Y) to estimate the covariance.

To assess this approach's efficiency, we compared it to the standard Monte Carlo method. We conducted 100 independent runs, each drawing 1000 samples for both approaches.

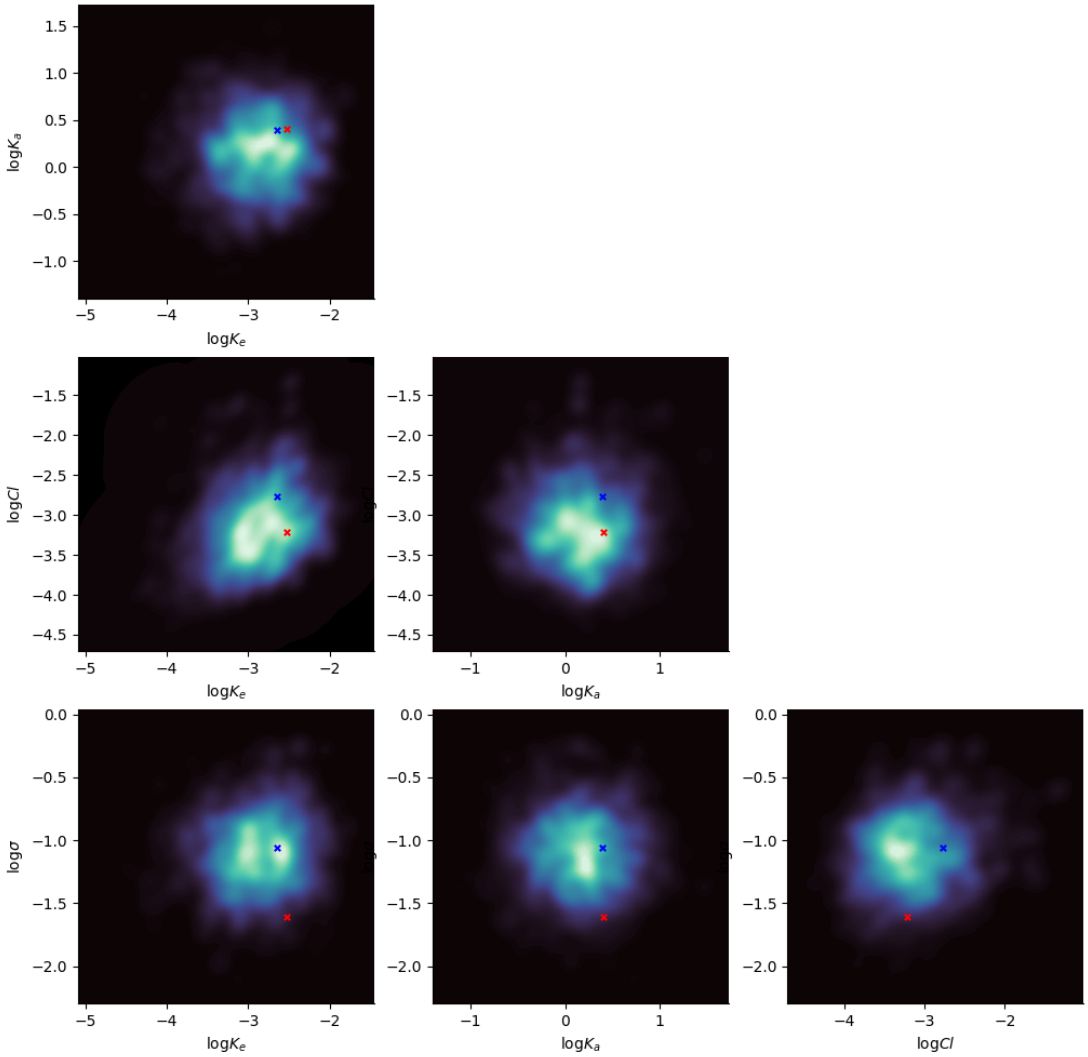

Figure 6. Pair plots for all parameter combinations, displaying two-dimensional kernel density estimations based on the 10,000 iteration samples from the ABC-MCMC algorithm at a tolerance of $\varepsilon = 0.25$. Each plot contrasts the actual value of θ (marked in red) with the predicted value of $E\left(\theta|D\right)$ (marked in blue).

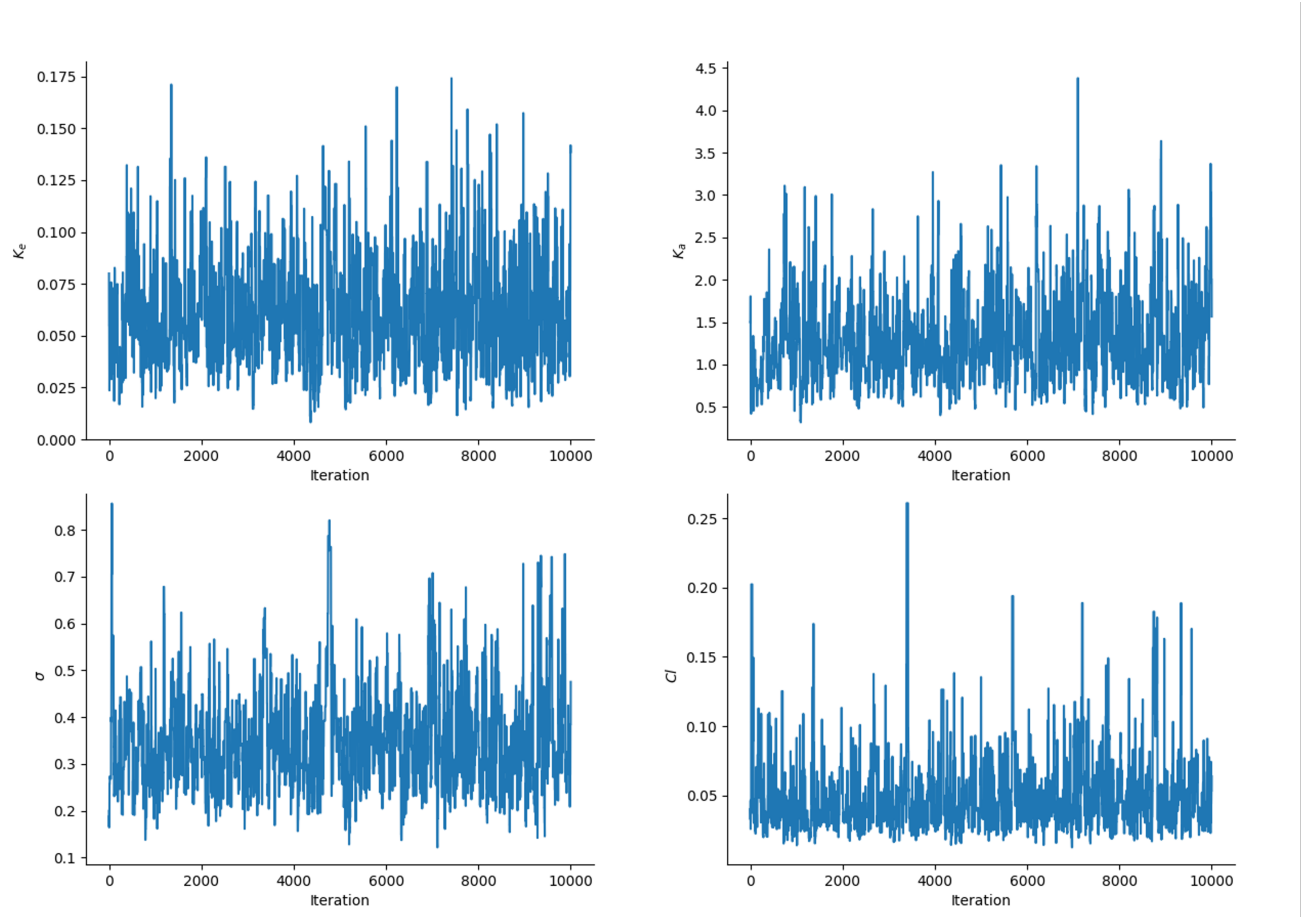

Figure 7. Trace plot displaying the progression of all components of θ over 10,000 iterations of the ABC-MCMC algorithm, executed with a tolerance level of $\varepsilon = 0.25$.

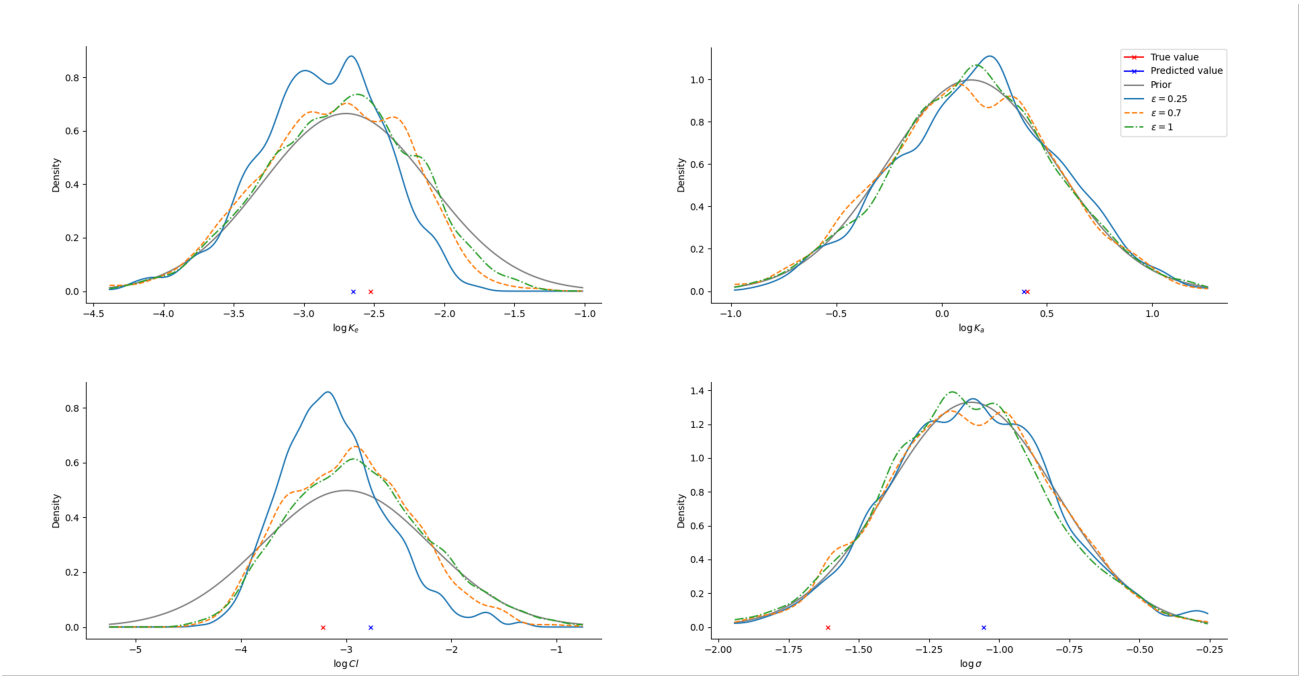

Figure 8. Kernel density estimators for the marginal distributions of estimated posteriors for the tolerance levels $\varepsilon = \{1, 0.7, 0.25, 0.025\}$. The true prior for each marginal is shown in grey. In addition the true underlying value for each parameter as well as the predicted value for the respective parameter based on the data D is shown.

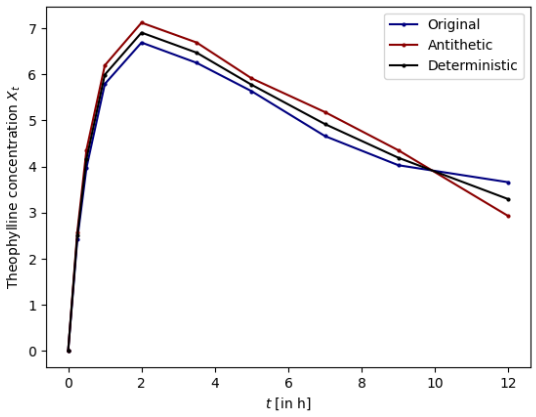

Figure 9. Concentration curve for one sample and its antithetic variable, alongside the approximated solution of the differential equation without the stochastic term.

At each sampling step, we calculated the current estimate and its absolute difference from the true mean. Figure 10 displays the experiment results. The plot reveals that both methods converge roughly at a rate of $1/\sqrt{N}$. However, the error curve of the control variate method starts lower than that of the Monte Carlo method. In fact, the Monte Carlo method requires 1000 samples to achieve an error as low as the control variate method with only 100 samples (plus 100 for the pilot run).

4. Conclusions

This project has undertaken a thorough investigation into the utilization of Rejection Sampling and Markov Chain Monte Carlo algorithms within the framework of Approximate Bayesian Computations (ABC). The examination of these methods on an academic example has underscored the notable acceleration achieved through the integration of MCMC algorithms into the ABC scheme. Notably, the ABC MCMC algorithm surpassed the performance of the ABC rejection algorithm, exhibiting higher acceptance rates while maintaining a comparable level of posterior approximation.

Building on the leanings from the academic example, we applied the ABC-MCMC algorithm to a real-world application. To deal with a multidimensional θ we employed an adaptive proposal distribution that uses past samples to estimate the

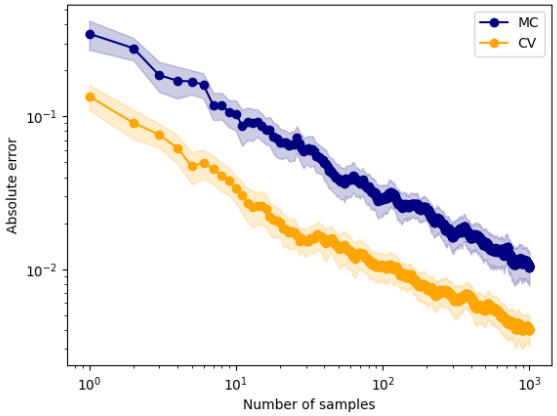

Figure 10. Comparison of Monte Carlo estimation with control variates against the standard Monte Carlo method using 100 independent runs consisting of 1000 samples for each method. For each step of both method the error of the mean to the true value was computed. The respective mean errors for each step as well as their 95% confidence intervals are reported in the plot.

covariance of the parameters in order to increase the algorithms acceptance rate. Our findings showed that when the tolerance levels were too high, the ABC-MCMC method tended to align closely with the prior. Further we identified the proposed discrepancy metric as potential bottleneck.

Finally, we developed two new Monte Carlo estimators to approximate the expected concentration of Theophylline in the blood after 12 hours. Both of them make use of the realizations of the normal distribution in the Euler-Maruyama update to reduce the variance of the samples and thus enable a more efficient estimation of the expectation.

References

Donnet, S. and Samson, A. Parametric inference for mixed models defined by stochastic differential equations. *ESAIM: Probability and Statistics*, 12:196–218, 2008. doi: 10.1051/ps:2007045. URL https://doi.org/10.1051/ps:2007045. Published online: 2008-01-23.

Gelman, A., Meng, X.-L., and Stern, H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–807, 1996.

Haario, H., Saksman, E., and Tamminen, J. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, Apr 2001.

Leman, S. C., Chen, Y., Stajich, J. E., Noor, M. A. F., and Uyenoyama, M. K. Likelihoods from summary statistics: Recent divergence between species. *Genetics*, 171(3):1419–1436, 11 2005. doi: 10.1534/genetics.104.040402. URL https://doi.org/10.1534/genetics.104.040402.

Marjoram, P. and Tavaré, S. Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics*, 7:759–770, 2006. doi: 10.1038/nrg1961.

Sisson, S. A., Fan, Y., and Tanaka, M. M. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007. doi: 10.1073/pnas.0607208104. URL https://doi.org/10.1073/pnas.0607208104.

Tanaka, M. M., Francis, A. R., Luciani, F., and Sisson, S. A. Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics*, 173(3):1511–1520, 7 2006. doi: 10.1534/genetics.106.055574. URL https://doi.org/10.1534/genetics.106.055574.