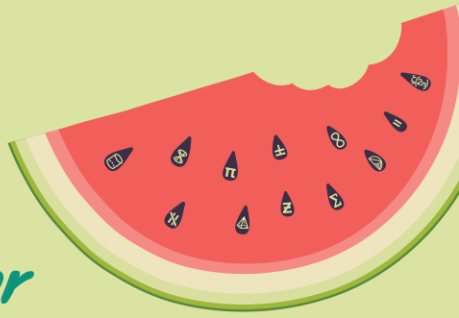
AMSI SUMMER RESEARCH SCHOLARSHIPS 2024–25

Get a taste for Research this Summer



Flexible Recalibration of Approximate Bayesian Models

Jack Fewtrell

Supervised by Professor Chris Drovandi & Mr Adam Bretherton

Queensland University of Technology

Jane Street®

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Abstract

Through the use of Bayesian inference, parameters of complex systems can be determined. Such systems may require computationally intractable likelihood functions, necessitating specialised techniques to perform inference, such as approximate Bayesian inference or approximate modelling. Approximate Bayesian inference techniques can be used to replace likelihood evaluations with comparisons between the observed and costly simulated data that is generated from the data generating process. Alternatively, approximate modelling refers to methods that reduce the complexity of models by approximating the likelihood so that it is tractable. However, approximate modelling can introduce approximation errors. Therefore, we may turn to techniques such as Bayesian score calibration, in an attempt to correct the approximate posterior resulting from the use of a surrogate likelihood. Unfortunately, Bayesian score calibration utilises a transformation that may be inflexible, particularly in complex systems where the misspecification in the approximate model changes across the parameter space. We propose sequential Bayesian score calibration to address the inflexible transformation used by Bayesian score calibration.

1 Introduction

To understand realistic systems, increasingly complex mathematical models are required. This complicates the Bayesian inference process, as complex models may have intractable likelihoods that requires specialised techniques to evaluate. Such specialised techniques may be computationally infeasible, affecting our understanding of those systems. Two key approaches a practitioner may use to evaluate these complex models are approximate Bayesian inference or approximate modelling.

Approximate Bayesian inference includes likelihood-free methods, such as simulation-based inference. Simulation-based inference avoids directly evaluating the intractable likelihood by comparing the observed and simulated data by using a suitable discrepancy function (Sisson, Fan, and Beaumont 2018). However, simulation-based inference techniques require up to millions of model simulations (Sisson, Fan, and Beaumont 2018; Bon et al. 2023). In contrast, approximate modelling uses a tractable approximation of the likelihood to trade misspecification for computational efficiency (Bon et al. 2023). Unfortunately, the introduction of such misspecification can affect the quality of inference. To correct this, we may turn to recalibration methods such as Bayesian score calibration, which corrects the approximate model by using a small number of simulations from the data generating process (Bon et al. 2023). To correct the approximate posterior, Bayesian score calibration uses a location-scale transforma-

tion. To learn this transformation, Bayesian score calibration firsts evaluates the approximate posterior, taking a small subsample of posterior samples as calibration samples. Next, each calibration sample generates a calibration dataset using the data generating process. Then, each calibration dataset is evaluated with the approximate model to generate a set of approximate samples. Finally, the parameters for the location-scale transform are learnt by optimising the energy score between the calibration samples and the transformed approximate samples from the associated generated calibration dataset. The location-scale transform assumes that the bias introduced by the approximate likelihood is constant across the parameter space, which may not be the case for all approximate likelihoods.

In this paper, we propose sequential Bayesian score calibration which is an extension to the Bayesian score calibration method to provide a more flexible transformation which accounts for bias in the approximate likelihood. Sequential Bayesian score calibration consists of sequentially applying the transformations to make the subsequent transformations more flexible. In this paper, we evaluate our new transformation on a simulate Susceptible-Infected-Recovered (SIR) example. Therefore, we use the Poisson approximate likelihood (PAL) method as described by Whitehouse, Whiteley, and Rimella (2023) as the surrogate likelihood function. Our simulated example has an intractable likelihood and therefore we use the particle method describe by Golightly, Henderson, and Sherlock (2014) to obtain samples from the true posterior, which is computationally very costly.

This paper is organised as follows. Section 2 outlines the methods proposed in this paper as well as relevant background information. Section 3 provides an example of the application of sequential Bayesian score calibration on an SIR model, comparing to the more costly simulation-based approach. In Section 4, we provide a discussion of our findings and direction of future works.

1.1 Statement of Authorship

This project was produced under the supervision of Professor Chris Drovandi and Mr Adam Bretherton. The project expands on the the work previously done by Bon et al. (2023) on Bayesian score calibration. The project applies models and methods produced by Whitehouse, Whiteley, and Rimella (2023). This project also uses code first written by Mr Adam Bretherton.

2 Methods

In this section, we discuss the underlying methods for sequential Bayesian score calibration. Specifically, we provide an overview on approximate Bayesian inference, approximate modelling, and Bayesian score calibration. In order to evaluate realistic systems, Bayesian inference is required. This often requires a prior distribution and a likelihood function to produce a posterior distribution. However, realistic systems often have intractable likelihoods, affecting our capability to obtain samples from the true posterior distribution.

To address the presence of intractable likelihoods, alternative methods are used, such as approximate modelling and approximate Bayesian inference.

2.1 Approximate Bayesian Inference

Approximate Bayesian inference are a class of methods that can be used to estimate parameters in systems with intractable likelihoods. These techniques may be likelihood-free, such as simulation-based inference. Simulation-based inference models simulate the model to generate synthetic data, which is then compared to the observed data (Sisson, Fan, and Beaumont 2018).

In this paper, particle Markov chain Monte Carlo (MCMC) is used as a simulation-based inference method. This method uses a marginal likelihood estimated from the output of a particle filter to perform (simulation-based) inference (Endo, van Leeuwen, and Baguelin 2019).

Unfortunately, simulation-based inference techniques such as particle MCMC can require up to millions of model simulations (Bon et al. 2023). Therefore, approximating the posterior distribution with this technique would be computationally burdensome.

2.2 Approximate Modelling

An alternative to approximate Bayesian inference is approximate modelling. Approximate modelling uses a surrogate likelihood to approximate the intractable likelihood. The tractable surrogate likelihood allows direct evaluation of the system of interest. Approximate modelling method have the benefit of being fast to evaluate as they do not require a simulation from the model. In this paper, we use the PAL method to approximate the intractable likelihood of the SIR model described in Section 4.1. Specifically, the PAL method approximates the number of observed infected individuals at each time step with a Poisson distribution. The PAL method is fast and has no tuning parameters, making it easy to implement (Whitehouse, Whiteley,

and Rimella 2023). However, the PAL method is inherently misspecified and only leads to an approximation of the true posterior distribution.

2.3 Bayesian Score Calibration

The Bayesian score calibration method (Bon et al. 2023) takes advantage of the surrogate likelihood and data generating process to provide fast, accurate inference. Specifically, Bayesian score calibration uses a small number of model simulations to attempt to correct the posterior approximation arising from the use of a surrogate likelihood. The number of model simulations utilised in the Bayesian score method is significantly smaller (hundreds) than the standard simulation-based inference (millions).

Bayesian score calibration uses a small number (e.g. 100) of approximate posterior samples referred to as calibration samples, and the data generating process to simulate a calibration dataset for each calibration sample. Then, Bayesian score calibration generates approximate posterior samples, referred to as approximate samples, for each calibration dataset and uses the calibration samples and associated approximate samples to learn a transformation given by

$$f_0(\theta) = \Sigma \cdot (\theta - \mu_\theta) + \mu_\theta + b, \tag{1}$$

where f_0 is the learnt transformation, Σ is a lower triangular matrix, μ_θ is the approximate posterior mean of the parameters, and b is a bias vector of each parameter.

The parameters of this transformation, Σ and b, are learned with an optimisation of the energy score (Gneiting and Raftery 2007) between calibration sample and associated set of approximate samples.

The location-scale transformation given in Equation 1 provides a single value for b across the parameter space. Therefore, this transform may lack flexibility. For example, if the bias in a system increases linearly, the transformation in Equation 1 will be unable to correctly calibrate. This lack of flexibility might produce poor approximations of the true posterior.

3 Sequential Bayesian Score Calibration

We introduce sequential Bayesian score calibration as a new method to calibrate approximate posteriors where misspecification introduced by the surrogate likelihood changes across the parameter space. This new method extends Bayesian score calibration by sequentially applying

the location-scale transformation described in Equation 1 n times to the approximate posterior. This results in a sequence of learnt transformation parameters, $b_1, \ldots, b_n$ and $\Sigma_1, \ldots, \Sigma_n$. We stop the process once there is no visual change in the recalibration results.

Specifically, we examine the parameters of the transform after each iteration stopping when the diagonal of Σ approaches one and b approaches a zero vector. In our example, we find that after 5 iterations the transformations make no meaningful difference.

4 Motivating Example

For the motivating example of this project, we use an SIR model. When evaluating an approximate posterior we use the PAL method and take 11000 samples, discarding the first 1000 as burn-in, using a random-walk Metropolis-Hastings algorithm (Hastings 1970; Metropolis et al. 1953). We set the number of calibration samples to $M = 100$ and select a new subset at each iteration of our sequential process.

4.1 SIR Model

The model that was investigated in this project was an SIR model, which is similar to the one found in the Whitehouse, Whiteley, and Rimella (2023) BSFLU example. The SIR model is a time series model evaluated at T discrete time steps. In this project, we set $T = 20$, and denote each timestep with $i \in \{0, 1, \ldots, T\}$.

The SIR model in this project has the following three parameters:

$$\theta = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix} \tag{2}$$

where $\theta_1 \in (0, \infty)$ controls the new infections, $\theta_2 \in (0, \infty)$ controls the rate of recoveries, and $\theta_3 \in (0, 1)$ controls the probability of observing a member of the infected population. In this project, the true value of these parameters are set to $\begin{bmatrix} 1.5 & 0.5 & 0.8 \end{bmatrix}$, referred to as the true data generating parameters.

This SIR model has four states, one for susceptible members of a population (S), one for asymptomatic members of a population (A), one for symptomatic members of a population (I), and one for recovered members of the population (R). Each of these states are updated at

each time step by their corresponding formulas

$$\begin{aligned} S_{i+1} &= S_i - \alpha + s \quad A_{i+1} = A_i + \alpha - d \\ I_{i+1} &= I_i - r + d \quad R_{i+1} = R_i + r - s, \end{aligned} \tag{3}$$

where α is the number of newly infected individuals of the population, r is the number of recovered individuals of a population, d is the number of newly recovered individuals of the population, and s is the newly susceptible members of the population. These values are drawn from binomial distributions given by

$$\begin{aligned} \alpha &= binom(S_i, 1 - exp(-\theta_1 \times (A_i + I_i))) \quad r = binom(I_i, 1 - exp(-\theta_2)) \\ d &= binom(A_i, 1) \qquad\qquad\qquad\qquad s = binom(R_i, 0.1). \end{aligned} \tag{4}$$

Finally, we report on the observed number of infected which is governed by

$$binom(I_{i+1}, \theta_3). \tag{5}$$

The output of the SIR model simulated at the true parameter values can be seen in Figure 1. This shows the number of observed infected members of the population follows a fluctuating trend where the infected population increases as the number of immune members of the population is low, and decreases as the immune population increases. This pattern continues as members of the population loses their immunity and becomes susceptible again.

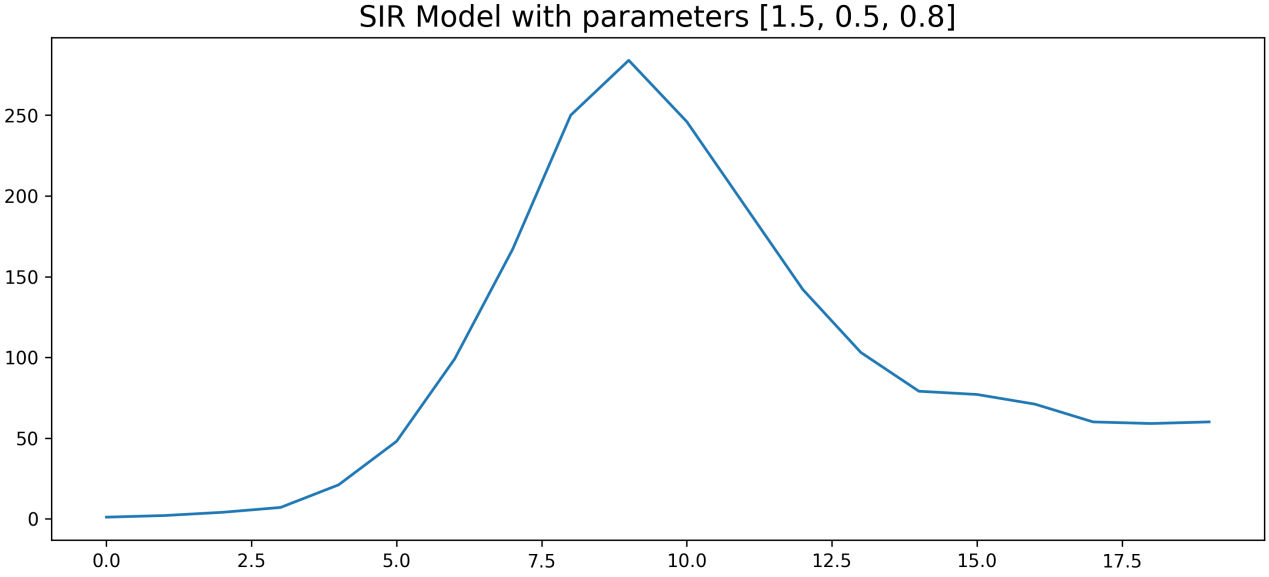


Figure 1: Plot of simulation of the SIR model, where $\theta_1 = 1.5$, $\theta_2 = 0.5$, and $\theta_3 = 0.8$ and $T = 20$.

This SIR model does have an intractable likelihood, however simulation is relatively easy,

so the true posterior distribution can be evaluated with the methods discussed in Section 2.1. To approximate the likelihood we make use of the PAL method discussed in Section 2.2.

4.1.1 Bayesian Inference of SIR Model

First, we consider the posterior from the approximate likelihood, PAL, and from particle MCMC. Figure 2 demonstrates the approximate and true posteriors differ markedly. The true parameter values lie far in the tails of the approximate posterior, demonstrating its inability to recover the true parameter values. However, there was a much greater computational cost when evaluating the true posterior as opposed to the approximate posterior. Particle MCMC took 4 hours to produce a posterior. In contrast, the approximate posterior based on the PAL took less then 20 seconds to run.

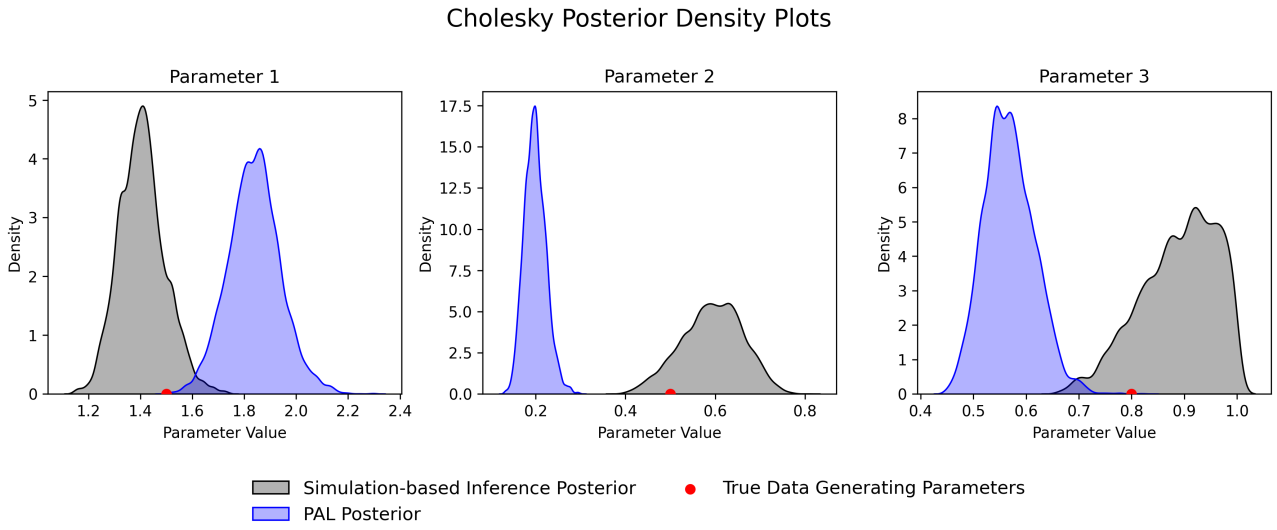


Figure 2: Estimated posteriors from particle MCMC (black) and PAL approximate method (blue) with true data generating parameters shown in red.

4.1.2 Bayesian Score Calibration of SIR Model

Next, we apply Bayesian score calibration. Figure 3 shows the posteriors of this method are closer to the true data generating parameters. However, we see that the true parameter values still remain in the tail of the recalibrated posterior. This indicates that the learnt transformation is not flexible enough to account for the misspecification of the PAL.

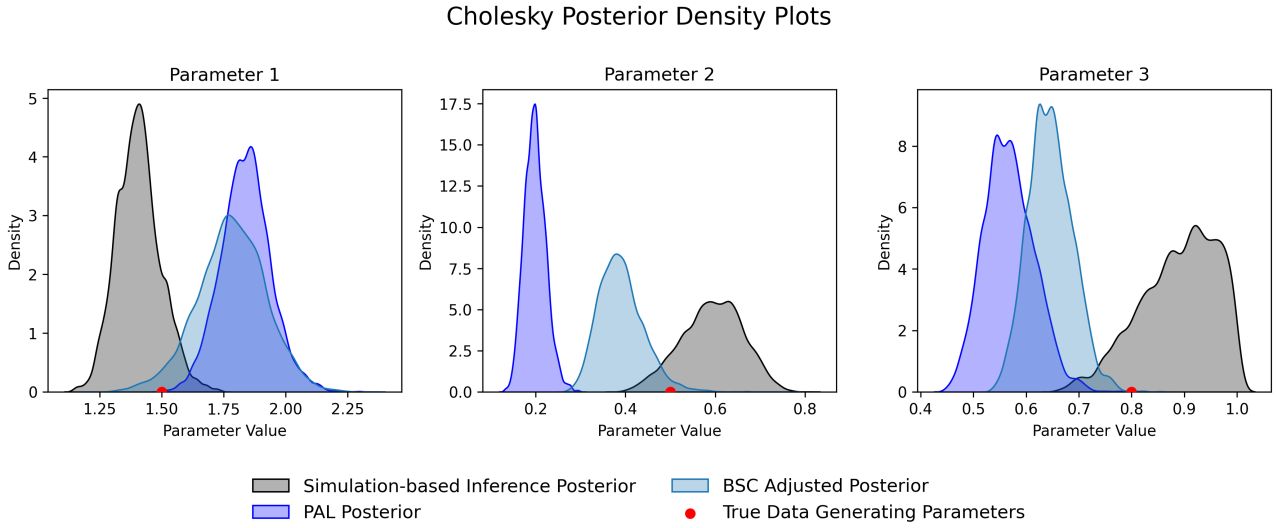


Figure 3: Estimated posteriors of the Bayesian score calibration method (light blue), against the original approximate posterior (dark blue), and the particle MCMC method (black). The true data generating parameters are shown in red.

By examining Figure 4 we can see that the bias in the approximate model changes across the parameter space and we expect that multiple applications of the location-scale transform to be flexible enough to correct the approximate posterior.

4.1.3 Sequential Bayesian Score Calibration of SIR Model

After each application of sequential Bayesian score calibration, the quality of the adjusted approximate posterior was evaluated qualitatively. This was based on the sequential Bayesian score calibrations approximate posteriors probability of the true data generating parameters as well as how this compares to the previous iterations approximate posteriors performance. After each run through of the method, the bias vector and the Σ matrix was also recorded to provide insight on whether the transformation had provided any change. These evaluation techniques were used to determine when further application of sequential Bayesian score calibration would provide any further improvements to the transformation of the approximate posterior.

Initially, the sequential Bayesian score calibration method was applied four times after the initial application of Bayesian score calibration as seen in Figure 5. We decided to stop the process as there was little change in the posterior approximation. This process took about 16 minutes to perform, with each iteration of the sequential method taking less than 4 minutes to run.

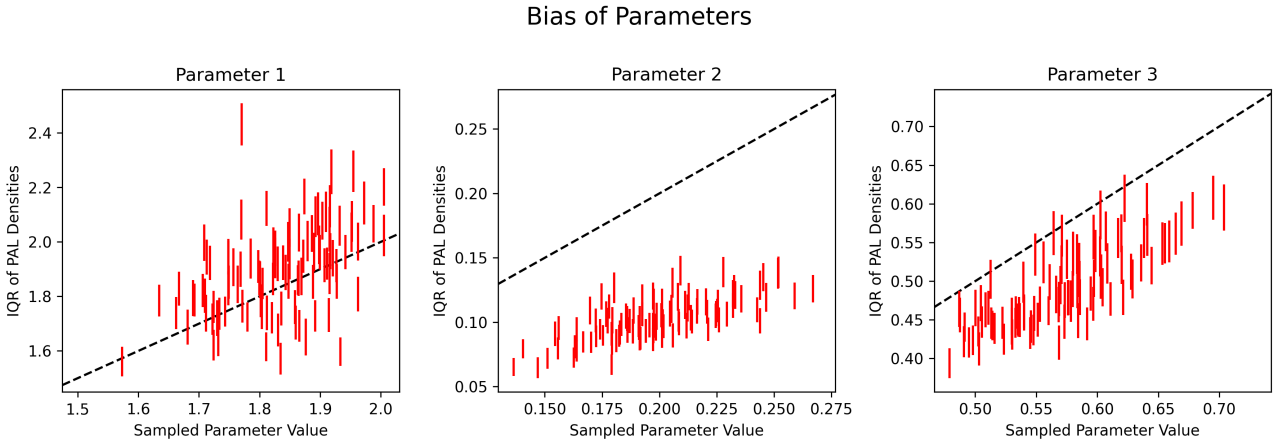


Figure 4: Comparison of the interquartile range of the posterior obtained with the PAL likelihood with the true parameter values for the 100 calibration parameter samples. The black dashed line is a 45 degree line representing an unbiased posterior approximation.

From Figure 5, it can be seen that second transformation of our sequential process significantly improves the original Bayesian score calibration approach, seeing better coverage of the true parameters. The third transformation significantly improves on all parameters. Our fourth transformation centres the posteriors around the true data generating parameters. Finally, our firth transformation saw little improvement to the posteriors.

We decided to stop after five iterations after considering the transformation parameters. At the fifth iteration the transformation parameters where

$$\Sigma = \begin{bmatrix} 0.8886 & 0 & 0 \\ 0.8943 & 0.7617 & 0 \\ 1.0411 & 1.0437 & 1.0205 \end{bmatrix}, b = \begin{bmatrix} 0.0010 & 0.0022 & 0.0350 \end{bmatrix}, \tag{6}$$

We can see that the diagonal of Σ is approaching one and b is approaching the zero vector, indicating that this is a sensible place to stop the algorithm.

5 Discussion and Future Works

Bayesian score calibration is an effective method to sample from complex models with intractable likelihoods. The transformation used by Bayesian score calibration requires that the misspecification introduced by the surrogate likelihood does not change across the parameter

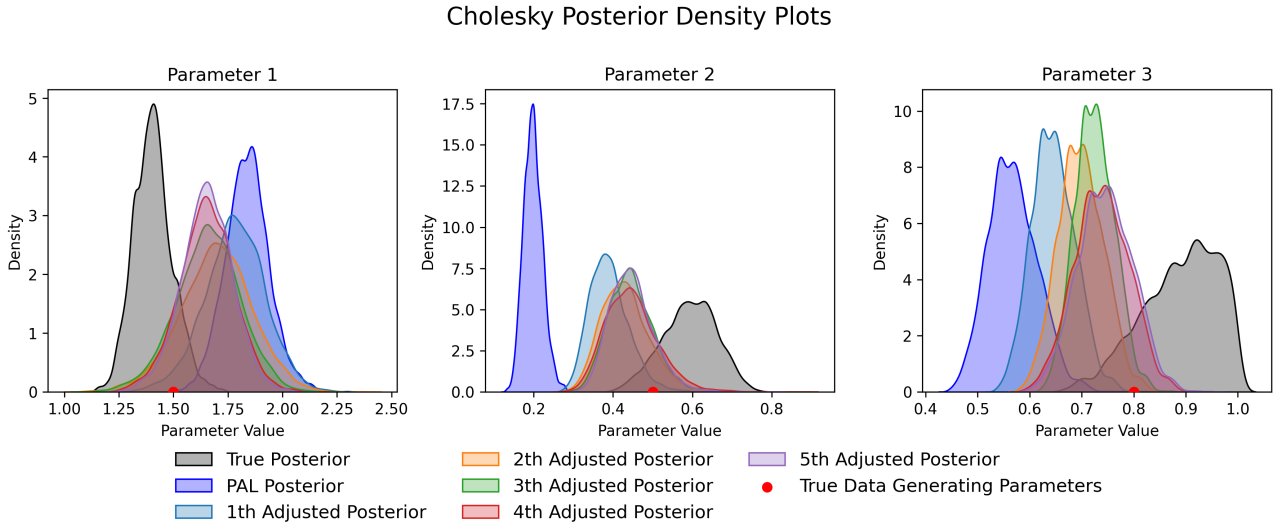


Figure 5: Estimated posteriors for Bayesian score calibration (1st sequential Bayesian score calibration) posterior (light blue), the 2nd sequential Bayesian score calibration (orange), the 3rd sequential Bayesian score calibration (green), the 4th sequential Bayesian score calibration (red), the 5th sequential Bayesian score calibration (purple) and the particle MCMC method (black). The true data generating parameters are shown in red.

space.

In this paper, we have presented sequential Bayesian score calibration as a extension to the Bayesian score calibration method. Specifically, our new sequential transformation methods can correct approximate models with approximation error that changes over the parameter space. Additionally, we retain the computational benefits of the Bayesian score calibration method.

In our motivating example, sequential Bayesian score calibration has successfully produced an adjusted posterior distribution that had greater coverage of the true data generating parameters. Sequential Bayesian score calibration was significantly less computationally burdensome than the simulation-based inference method, with our technique being able to generate accurate posterior samples 15 times faster than particle MCMC. This is a result of the number of model simulations performed, with hundreds of model simulations being performed for the sequential Bayesian score calibration method as opposed to millions required for the simulation-based inference method.

Sequential Bayesian score calibration can be applied to additional complex models to evaluate the effectiveness of the technique. Sequential Bayesian score calibration was unable to recover the true posterior distribution which was right skewed. The current location-scale transformation only accounts for the 1st and 2nd order moments. A more complex transfor-

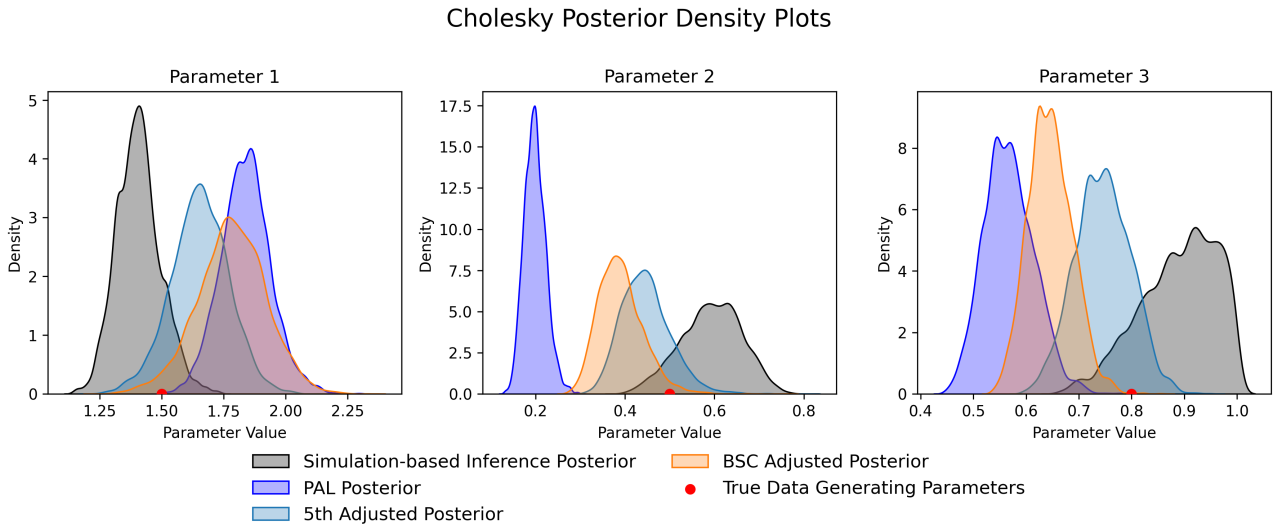


Figure 6: Estimated posteriors for the Bayesian score calibration posterior (orange), the 5th sequential Bayesian score calibration (dark blue) and the particle MCMC method (black). The true data generating parameters are shown in red.

mation with higher order moments may provide better results, such as accounting for the 3rd and 4th order moments.

Finally, we determined the number of sequential transformation by considering the change in transformation parameters, a qualitative measure. Future work could explore more quantitative measures to stop the process.

6 References

Sisson, Scott A, Yanan Fan, and Mark Beaumont (2018). *Handbook of approximate Bayesian computation*. CRC press.

Bon, Joshua J et al. (2023). *Bayesian score calibration for approximate models*. arXiv: 2211.05357 [stat.CO]. URL: https://arxiv.org/abs/2211.05357.

Whitehouse, Michael, Nick Whiteley, and Lorenzo Rimella (July 2023). "Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85.4, pp. 1173–1203. ISSN: 1369-7412. DOI: 10.1093/jrsssb/qkad065. eprint: https://academic.oup.com/jrsssb/article-pdf/85/4/1173/51801291/qkad065.pdf. URL: https://doi.org/10.1093/jrsssb/qkad065.

Golightly, Andrew, Daniel A. Henderson, and Chris Sherlock (May 2014). "Delayed acceptance particle MCMC for exact inference in stochastic kinetic models". In: *Statistics and Computing* 25.5, pp. 1039–1055. ISSN: 1573-1375. DOI: 10.1007/s11222-014-9469-x. URL: http://dx.doi.org/10.1007/s11222-014-9469-x.

Endo, Akira, Edwin van Leeuwen, and Marc Baguelin (2019). "Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers". In: *Epidemics* 29, p. 100363. ISSN: 1755-4365. DOI: https://doi.org/10.1016/j.epidem.2019.100363. URL: https://www.sciencedirect.com/science/article/pii/S1755436519300301.

Gneiting, Tilmann and Adrian E Raftery (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". In: *Journal of the American Statistical Association* 102.477, pp. 359–378. DOI: 10.1198/016214506000001437. eprint: https://doi.org/10.1198/016214506000001437. URL: https://doi.org/10.1198/016214506000001437.

Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, pp. 97–109. ISSN: 00063444, 14643510. URL: http://www.jstor.org/stable/2334940 (visited on 02/28/2025).

Metropolis, Nicholas et al. (June 1953). "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092. ISSN: 0021-9606. DOI: 10.1063/1.1699114. eprint: https://pubs.aip.org/aip/jcp/article-pdf/21/6/1087/18802390/1087_1_online.pdf. URL: https://doi.org/10.1063/1.1699114.